Thesis for The Degree of Doctor of Philosophy

Pushing the boundaries of single molecule microscopy, optical sensing and mass spectrometry through deep learning

Henrik Klein Moberg

Department of Physics
Chalmers University of Technology
Gothenburg, Sweden, 2024

Pushing the boundaries of single molecule microscopy, optical sensing and mass spectrometry through deep learning

Henrik Klein Moberg

ISBN 978-91-8103-118-8
Doktorsavhandlingar vid Chalmers tekniska högskola, Ny serie nr 5576.
ISSN 0346-718X
Technical Report No. 5576

Department of Physics
Chalmers University of Technology
SE-412 96 Göteborg,
Sweden
Phone: +46 (0)31 772 29 63

Cover image: A text-to-image artificial intelligence's interpretation of the phrase "Artificial Intelligence Studies Biological Molecules Through Optical Microscopy". Credit: Midjourney from Midjourney, Inc.

Printed by Chalmers Digitaltryck,
Gothenburg, Sweden 2024.

to my dear friends and family

Pushing the boundaries of single molecule microscopy, optical sensing and mass spectrometry through deep learning

Henrik Klein Moberg

Department of Physics
Chalmers University of Technology

Abstract

Deep learning has transformed sensing and characterization technologies, enabling significant advancements across various scientific domains. This thesis investigates the application of deep learning techniques to enhance the applicability of microscopy, spectrometry, and sensing, particularly under high noise conditions. The central hypothesis of this research is that deep learning can substantially improve the sensitivity and specificity of these technologies, allowing for the detection and analysis of minute signals that were previously obscured by the noise.

Key contributions include the development of novel deep learning methods that enhance nanofluidic scattering microscopy, nanoplasmonic sensing, and mass spectrometry. These methods enable precise quantification of chemical reactions on nanoscale surfaces, detailed detection of cellular structures and molecular interactions, and accurate identification of low-concentration substances amidst strong background signals. Collectively, these advancements push the boundaries of what can be measured and observed at microscopic and molecular scales, offering groundbreaking applications in environmental monitoring and healthcare diagnostics.

Keywords

Deep Learning, Nanofluidic Scattering Microscopy, Nanoplasmonic Sensing, Mass Spectrometry, Cross-modality Transformation, Virtual Staining, Biomolecule Detection, Environmental Monitoring,

List of Publications

This thesis is based on the following publications:

[**Paper I**] Barbora Špačková, **Henrik Klein Moberg**, Joachim Fritzsche, Johan Tenghamn, Gustaf Sjösten, Hana Šípová-Jungová, David Albinsson, Quentin Lubart, Daniel van Leeuwen, Fredrik Westerlund, Daniel Midtvedt, Elin K. Esbjörner, Mikael Käll, Giovanni Volpe, and Christoph Langhammer, *Label-free nanofluidic scattering microscopy of size and mass of single diffusing molecules and nanoparticles*
Nature Methods, vol. 19, no. 6, pp. 751–758, 2022, issn: 1548-7105. doi: 10.1038/s41592- 022- 01491- 6..

[**Paper II**] **Henrik Klein Moberg***, David Tomecek*,Sara Nilsson, Athanasios Theodoridis, Iwan Darmadi, Daniel Midtvedt, Giovanni Volpe, Olof Andersson and Christoph Langhammer, *Neural network enabled nanoplasmonic hydrogen sensors with 100 ppm limit of detection in humid air*
*Nature Communications, vol. 15, no. 1, p. 1208, 2024, issn: 2041-1723. doi: 10.1038/s41467- 024- 45484- 9., *equal author contribution.*

[**Paper III**] **Henrik Klein Moberg**, Ievgen Nedrygailov, David Albinsson, Joachim Fritzsche, Giovanni Volpe, Christoph Langhammer, *Deep-learning enabled mass spectrometry of the reaction product from a single catalyst nanoparticle*
In review, Nature Communications 2024.

[**Paper IV**] Viktor Martvall, **Henrik Klein Moberg**, Athanasios Theodiridis, David Tomecek, Pernilla Tanner, Paul Erhart, Christoph Langhammer, *Accelerating Plasmonic Hydrogen Sensors for Inert Gas Environments by Transformer-Based Deep Learning*
In review, ACS Sensors 2024.

[**Paper V**] **Henrik Klein Moberg***, Benjamin Midtvedt*, Dana Hassan*, Jesús Pineda*, Jesús Domínguez*, *Review: cross modality transforms in biological microscopy enabled by deep learning*
*In review, Advanced Photonics 2024, *equal author contribution.*

[**Paper VI**] **Henrik Klein Moberg**, Bohdan Yeroshenko, David Albinsson, Joachim Fritzsche, Daniel Midtvedt, Barbora Špačková, Giovanni Volpe, Christoph Langhammer, *Computer Vision Enabled Label-free Nanofluidic Scattering Microscopy Weight-and-Size Characterization of Single Biomolecules down to the 6kDa Regime*
In manuscript.

Publications not included in this thesis

The following publications were published during my PhD studies, or are currently in submission/under revision. However, they are not appended to this thesis, due to contents overlapping that of appended publications or contents not related to the thesis.

[**a**] Benjamin Midtvedt, Jesús Pineda, **Henrik Klein Moberg**, Harshith Bachimanchi, Joana Pereira, Carlo Manzo, Giovanni Volpe,
Deep Learning Crash Course
To be published 2025 in No Starch Press.
Available at: https://nostarch.com/deep-learning-crash-course.

[**b**] Laura Waller, Kristina Monakhova, Jason Smith, Wang Hao, Lei Tian, Natasa Sladoje, Christoph Langhammer, **Henrik Klein Moberg**, Barbora Špačková, Aydogan Ozcan, Carlo Manzo, Saga Helgadottir, Aykut Argun, Mite Mijalkov, Daniel Vereb, Yu-wei Chang, Joana Pereira, Daniel Midtvedt, Ida-Maria Sintorn, Benjamin Midtvedt, Jesús Pineda, Sage Daniel, Estibaliz Gómez-de-Marisca, Wei Ouyang, Ebba Josefson, Carolina Wälhby, Lei Tian, Giovanni Volpe, *Roadmap on Deep Learning for Microscopy*
In review in JPhys Photonics.

My Contributions to the Publications

[**Paper I**] I developed the deep learning methods used for analyzing proteins, DNA fragments, extracellular vesicles and lipid nanoparticles, as well as carried out all the aforementioned analysis. I wrote all sections in the paper related to deep learning, as well as some portions of the results section relating to the results of the deep learning analysis.

[**Paper II**] I designed, trained and used the deep learning models, as well as produced all figures containing results from the model and wrote the related portions of the manuscript.

[**Paper III**] I designed, trained and used the deep learning model, performed all experimental measurements, as well as wrote the first draft of the manuscript and produced all figures.

[**Paper IV**] I co-developed the deep learning model, as well as provided all preliminary results and figures for the manuscript.

[**Paper V**] I wrote the section on molecular imaging and produced its corresponding figure.

[**Paper VI**] I developed all theoretical results concerning the limits of the deep learning methods, designed, trained and used the models, as well as wrote the first draft of the manuscript and produced all figures.

Contents

Chapter 1

Introduction

Extracting tiny signals from noise is a general and important challenge in experimental science. In mass spectrometry, this challenge manifests itself as the need to quantify chemical reactions on nanoscopic surface areas, such as single nanoparticles or even single atoms. In microscopy, this manifests as the necessity to discern and quantify fine details at cellular or molecular scales amidst significant forms of visual interference. In sensing, this manifests itself as the need to detect and quantify low concentrations of substances or minimal physical changes in the presence of significant background noise. The distinction between these applications comes from the underlying physical principles giving rise to the measured phenomena and, crucially, in common for all these seemingly disparate fields of experimental science is thus the problem formulation on a high level of abstraction. This means that if one can solve the generic challenge of separating signal from noise, the potential ramifications for all fields of experimental science can hardly be overstated. The goal of this thesis is to approach such a solution using artificial intelligence (AI), because, to borrow the words of Nobel Laureate Dr. Hassabis of Google's DeepMind, "AI is perfect at taking weak, messy signals and making sense of them." [1]

Through the paradigm of deep learning (DL), recent progress in AI has been truly astounding; quickly evolving from struggling to recognize animals in relatively simple images to leading a revolution in autonomous vehicles [2], image-based medical diagnosis [3], material science [4], astronomy [5], weather prediction [6], waste disposal [7], natural language processing [8] and many other fields in the span of just a few years. The keys underpinning the success lie in the fruitful marriage of a few seemingly disparate technological advancements; the development of computationally tractable self-learning neural network-based algorithms [9], the ever-growing availability and quality of digital acquisition devices enabling high-throughput experimentation, and the significant growth of the financially lucrative personal computer market leading to ever-faster parallellizable computing devices. The combination of large amounts of high-quality images, effectively 2-dimensional vector spaces

containing locally dependent features, with convolution-based DL architectures known under the umbrella term of *convolutional neural networks* (CNNs), has been particularly potent. Self-learning CNN-based algorithms learning general short-range dependencies between locally correlated data-points (i.e., pixels in images), has been the workhorse of most recent progress [10]. This generality means that algorithms developed in the context of, e.g., animal classification or facial recognition can often directly be applied to other image-based tasks with relatively minor tweaks, as a result of the generic self-learning nature of DL-based algorithms. This generality has effectively formed a chain reaction within quantitative image-based analytical research, in which lessons learnt in one application of DL-based AI can often directly be applied to every other application.

One field where the potential of DL is becoming increasingly clear is optical microscopy, in which data analysis is often a limiting step because of images which are both highly complex and have low signal-to-noise ratio. Promising proof-of-principle results on super-resolving fluorescence microscopy [11], DL-enhanced holography [12] and virtual chemical staining [13] to name a few examples, suggests that DL is beginning to lead a revolution also in optical microscopy. In life sciences, optical microscopy has proven instrumental for fast and accurate characterization of biomolecule properties in- and outside of complex biofluids. Historically, in the regime of small biomolecules ($\sim$100s of kDa), sensitive imaging techniques such as fluorescence microscopy and interferometric scattering microscopy (iSCAT) fill this characterization role, though they suffer from problems arising from fluorescent labelling [14] and necessity of surface binding, respectively. To remedy this, we have in paper I introduced Nanofluidic Scattering Microscopy (NSM), which relies on the sub-wavelength interference of light scattered from a nanofluidic channel and a nano-object, such as a biomolecule, inside it. However, the combination of generally low SNR resulting in typically "blinking" trajectories, changing imaging conditions, and noise which is complex in both the spatial and temporal dimensions, makes it very difficult to reliably calculate the biomolecules' properties through standard heuristic-based methods alone.

The increasing need for advanced data science within life sciences forms the essential background against which the works of Papers I, V and VI of this thesis are justified; we show that by employing DL-based AI algorithms to analyze biomolecule trajectories measured by our in-house NSM system, we not only manage to analyze complex biological nanoparticles in the form of extracellular vesicles, but also surpass state-of-the-art results on nano-scale biomolecule weight-and-size screening and trajectory tracking. We show that a custom CNN-based architecture trained on simulated biomolecule trajectories atop simulated noise surpasses the performance of standard algorithms in terms of tracking and determining the molecular weight (MW) and hydrodynamic radius (R_s) of biomolecules in the low-kDa regime. Furthermore, this model can accurately track and measure MW and R_s of biomolecules even below the 10-kDa regime, constituting approximately an order of magnitude improvement in limit of detection (LoD) over current state-of-the-art. In this regime, hitherto

elusive species of biomolecules become accessible for label-free study using NSM, including cytokines ($\approx 5 - 25$ kDa) important for cancer research and the protein hormone insulin (≈ 5.6 kDa), potentially opening up entirely new avenues of biological research [15].

In the realm of experimental science, the task of distinguishing minute signals from background noise represents a universally critical challenge. One example is in the field of so-called single particle catalysis, where the study of individual catalytic particles holds the potential to greatly impact our understanding of catalyst structure-activity correlations. The faint signals emanating from these individual particles pose a significant unresolved experimental challenge. Inspired by the advances in DL, in Paper III, we tackle this issue by integrating nanofluidic reactors with a deep learning-based denoising auto-encoder. This enables the detection of reactions occurring on the tiny surfaces of single nanoparticles, with CO oxidation on Pd serving as a case study. The nanofluidic device coupled with this deep learning technique achieves 3 orders of magnitude improvement in LoD of the quadrupole mass spectrometer (QMS) readout over current state-of-the-art in terms of total catalytic surface are measured, and enables new possibilities for analytical chemistry on the surface of single catalytic nanoparticles.

Given the importance of hydrogen as a clean energy source, the development of highly sensitive, selective, and fast-response hydrogen sensors is of paramount interest. Here, the extraction of complex patterns from sensor data, which are indicative of the presence and concentration of hydrogen or variations in humidity, is an important challenge. These patterns might be too subtle or complex for conventional analysis techniques but can be identified by trained neural networks that have learned to correlate specific signal features with environmental conditions or gas concentrations. In Papers II and IV, we demonstrate the potential of AI-based analysis in the field of hydrogen gas sensing, particularly in inert gas environments critical for safety in hydrogen technologies and in high humidity ambient air environments critical for safety sensor applications. Through a novel application of deep learning, specifically tailored transformer architectures, the detection capabilities of plasmonic hydrogen sensors have been significantly enhanced. The developed transformer model accelerates the sensor's response time by up to 40 times in oxygen-free environments and improves the reliability and safety of hydrogen detection by providing accurate predictions of hydrogen concentrations and eliminating the pressure dependence of the sensor's response. The transformer model also enables up to an order of magnitude lower LoD of sensor signal in highly humid environments. Such advancements underscore the potential of AI in overcoming longstanding challenges in sensor technology, promising substantial improvements in safety measures for hydrogen energy applications.

In this thesis, which is partially based on my licentiate thesis [16], I delve deeper into the complexities of DL-based AI. I present how we have employed it to NSM to analyze single biomolecule trajectories and establish a new state-of-the-art in microscopy in certain respects, demonstrate its application in enhancing

the sensitivity and specificity of hydrogen gas sensors, and unlock the potential for analytical chemistry at the nanoscale within single nanoparticle catalysis. These examples are emblematic of a broader paradigm shift across diverse fields of experimental science, where the capacity of AI to parse through and make sense of weak, complex signals is not just an incremental improvement but represents a foundational shift in how we approach and solve problems. The combination of deep learning technology with the intricacies of experimental science has opened new doors to discovery, allowing us to explore phenomena and mechanisms that were previously beyond reach. By harnessing the ability of AI to discern patterns and correlations within complex datasets, we are poised to uncover insights that can propel forward our understanding in fields ranging from life sciences to energy sustainability and beyond. The journey from recognizing animals in images to redefining the limits of detection and analysis across experimental science underscores the transformative impact and impressive progress of AI, reaffirming Dr. Hassabis's vision of its capability to interpret and understand the complex language of the natural world. This thesis stands as a modest testament to the profound and versatile power of deep learning, and its role as a critical tool in the future of scientific discovery and innovation.

1.1 Scope of the Thesis

The primary objective of this thesis is to address the generic challenge of separating signal from noise in experimental science using DL-based AI algorithms. Specifically, we aim to:

1. **Enhance Biomolecule Analysis in NSM**: Employ DL-based AI algorithms to analyze biomolecule trajectories measured by our NSM experimental platform, surpassing state-of-the-art results on nanoscale biomolecule weight-and-size screening and trajectory tracking.

2. **Detect Reaction Products in Single Nanoparticle Catalysis**: Integrate nanofluidic reactors and quadrupole mass spectrometry with deep learning-based denoising auto-encoders to detect chemical reactions occurring on the tiny surfaces of single nanoparticles, with CO oxidation on Pd serving as a case study.

3. **Improve Hydrogen Gas Sensing**: Demonstrate the potential of AI-based analysis in the field of hydrogen gas sensing, particularly in inert gas environments and in humid air environments critical for safety in hydrogen technologies, by enhancing detection capabilities of plasmonic hydrogen sensors using tailored transformer architectures.

By addressing these objectives, we seek to demonstrate the transformative potential of AI in overcoming longstanding challenges in experimental science.

The scope of this thesis focuses on the integration of DL technologies with advanced experimental science methods, particularly within the areas of microscopy, mass spectrometry and nanoplasmonic sensing. The primary objective

is to advance the ability to detect, analyze, and interpret weak or complex signals that are often obscured by noise in experimental data. This endeavor aims to solve the pervasive challenge of signal-noise separation across these fields, leveraging the capabilities of DL to improve the sensitivity, specificity, and efficiency of various analytical and observational techniques.

The thesis presents a series of studies demonstrating the application of DL in enhancing the performance of optical plasmonic hydrogen sensors, particularly under conditions of high humidity or in inert gas environments, which are critical for safety in hydrogen technologies. Additionally, it explores the development of Nanofluidic Scattering Microscopy, which enables the label-free characterization of biomolecules at the nanoscale. Through the application of DL algorithms, the thesis also addresses the challenges in single particle catalysis by facilitating the analysis of reaction products from individual catalyst nanoparticles.

Each chapter of the thesis contributes to a broader understanding of how DL can be applied to overcome limitations in traditional experimental science methodologies. This includes enhancing the limit of detection in sensing applications, improving the resolution and contrast in microscopy, and achieving more accurate and sensitive detection of catalytic reactions at the nanoscale using mass spectrometry. The work underscores the transformative potential of integrating DL with experimental science, opening new avenues for research and application in these critical fields.

This thesis does not cover the hardware development of microscopy or sensing devices, nor does it delve into the theoretical underpinnings of DL algorithms beyond their application in the context of the presented research.

1.2 Thesis Structure

The thesis is organized as follows:

- **Chapter 2 (Paper V)**: Provides a detailed introduction to deep learning, its principles, relevance to signal extraction challenges in experimental science, and applications in modern science.
- **Chapter 3 (Paper I and VI)**: Presents the application of DL-based AI algorithms to NSM for analyzing single biomolecule trajectories, demonstrating improved tracking and measurement capabilities.
- **Chapter 4 (Papers II and IV)**: Explores the enhancement of hydrogen gas sensors using AI-based analysis, detailing how DL models accelerate sensor response time and improve detection accuracy.
- **Chapter 5 (Paper III)**: Examines the integration of nanofluidic reactors with DL-based denoising auto-encoders to detect chemical reactions on single nanoparticles, achieving significant improvements in limit of detection.
- **Chapter 6**: Summarizes appended papers included in this thesis.

- **Chapter 7**: Concludes the thesis with a summary of findings, reflections on the research, and suggestions for future work.

Chapter 2

Deep Learning (DL)

Neural networks are effectively piecewise linear function approximators, capable in principle of approximating any function through finding hyperplane partitions of the input data during training [17]. The depth of a neural network affects how complex these partitions might grow, in essence improving the representational ability of the whole network architecture [17]. Effectively, subsequent layers of neural networks (assuming equal or growing layer size) increase the dimensionality of the latent representation of the input data to the extent where finding some hyperplanar partition to approximate the output values is always possible [17]. A neural network is considered deep if it consists of more than one intermediate ("hidden") layer, and every application of such a deep neural network is considered to be *Deep Learning* (DL). This simple yet powerful idea forms the basis of all neural network-based computation, and the associated field of modern deep learning revolves around finding ways of steering these networks to learn more physically relevant, generalizable and interpretable functions during their training process [18]. The simplest example of this steering is the CNN, the backbone of modern computer vision, which connects neurons together in specific ways to bias them toward learning locally relevant correlations (important for, e.g., images).

Black-box based DL analysis has seen a healthy amount of skepticism in recent years, owing equally to its inherent lack of interpretability and unreliable generalizability beyond its training set due mainly to underspecification [19]. However, many AI industry veterans argue that this skepticism often reaches unfounded levels in many practical applications; many problems common to deep learning in the past are better understood and can be mitigated through rigorous implementation [20]. There are even accounts of non-mission-critical DL-based algorithms reaching high accuracy never being employed in industry [20], mainly due to a mystical distrust of self-learnt algorithms. Thus, it is often difficult to see whether the error lies in the foundations of DL, or in the faltering diffidence of the end-users who should be benefitting the most. To this dilemma, the efficacy of its remedy lies in its simplicity; the mathematical foundations of

DL must be deeper explored and its limitations unequivocally presented. This section is based on presenting these mathematical foundations of (convolutional) neural networks in its proper context; as that of the cornerstone of modern computer vision and the catalyst for the deep learning revolution as a whole.

2.1 The Artificial Neuron

Conceptually, deep learning is inspired by the structure and function of the brain. Specifically, the artificial neural network at the heart of all deep learning is explicitly inspired by our own biological neural networks; the conglomeration of neurons, synapses, chemical and electrical signals which make up our brains. In its simplest interpretation, the biological neuron can be approximated as a computational unit which fires a signal of a certain strength if its input signal exceeds a certain threshold. Specifically, the biological neuron "fires" (activates) with strength $w \times x$ if its input signal overcomes threshold b [21]. Every artificial neural network consists of these individual computational units, known as artificial neurons, with a certain weight w and bias b, and activation function a, such that each neuron computes the simple linear function

$$z = a(wx + b) \tag{2.1}$$

given an electrical signal of strength x which overcomes a treshold (bias) b, for a given weight w and activation function a [21]. The values of the weights w and biases b are determined during training of the network on some dataset, whilst the activation function is fixed and chosen by the engineer dependent on the nature of the dataset in question [21]. This simple function is the basis of all deep learning-based systems, regardless of their ultimate manifestations' complexity or advanced functionality.

This computational model of artificial neural networks was introduced in 1943 by McCulloch and Pitts [22], and the following decades of developing theory and practice of neural networks is a complex sequence of cycles of growing hype, investment, disappointment and stagnation. This story of the highly nonlinear development of deep learning has been told many times [23]–[25] and will not be reiterated in this thesis. Rather, we note the most relevant contribution to the field of deep learning in the context of computer vision, and arguably, the most significant contribution in modern deep learning overall; the work of LeCun et. al. [9] in introducing convolutional neural networks trained with learnable parameters. This would later form the basis for AlexNet [26], which would go on to convincingly win the long-standing image classification challenge ImageNet [27] and spark the deep learning revolution which is still gaining speed to this day. Following this success, several specially-designed CNN architectures have been introduced for different purposes: most notably for the purposes of this thesis are those developed for biomedical image analysis [28], object detection [29] and skip-connection-based networks for enabling deeper networks, i.e. networks with more consecutively connected layers, than

previously possible [30]. To understand the significance of these algorithms in the context of computer vision in broad and microscopy in particular, we need to delve into their precise mathematical formulation.

2.2 Convolutional Neural Networks

As alluded to in the introduction of this chapter, the power of deep learning comes from representation (feature) learning. The reason convolutional neural networks dominate computer vision is because it learns to represent inherent features of datasets better than handcrafted algorithms. The inductive locality prior, weight sharing and stochastic gradient descent through backpropagation [9] condense this computational power into a tractable and ultimately parallelizable form which enables training neural networks in feasible timeframes with practically relevant data [31].

In their basic manifestation, convolutional neural networks consist of a combination of three types of layers; fully connected neural layers, convolutionally connected neural layers, and pooling layers, the details of which are recounted in the sections below.

2.2.1 Fully Connected Neural Layers

The fully connected neural network layer consists of any number of neurons computing the linear transformation in equation 2.1, where every neuron in the layer is connected to every other neuron in the previous layer (see Figure 2.1).

As shown in Figure 2.1, the neurons' activation consists of a weighted sum of the inputs to the neuron plus a bias, where the values of the weights and bias are learnt during training. This value, known as the *logit*, is then passed through an activation function (e.g. a sigmoid) before being passed further to the next layer of neurons (or output if it was the last layer). A sequence of such fully connected layers constitutes the simplest version of a deep neural network, and in turn the basis for the simplest realization of deep learning. Even this simple network should not be underestimated, since mathematical theorems [32] prove that such a fully connected network with only a single sufficiently sized layer can approximate any finite-dimensional function to arbitrary accuracy. However, the lack of any inductive bias within the network architecture makes learning a very data-intensive process, and any trained model is highly prone to both overfitting and underspecification (i.e. manifested as unrobust generalizability). One simple way of introducing inductive bias into our models is by, for instance, including memory of past inputs in the neuron activations (recurrent neural layers) or only considering information from inputs which are spatially localized close together (convolutional layers).

2.2.2 Convolutional Layers

Convolutional layers are a special case of the fully connected layer; we remove connections between neurons such that only the inputs which are close together

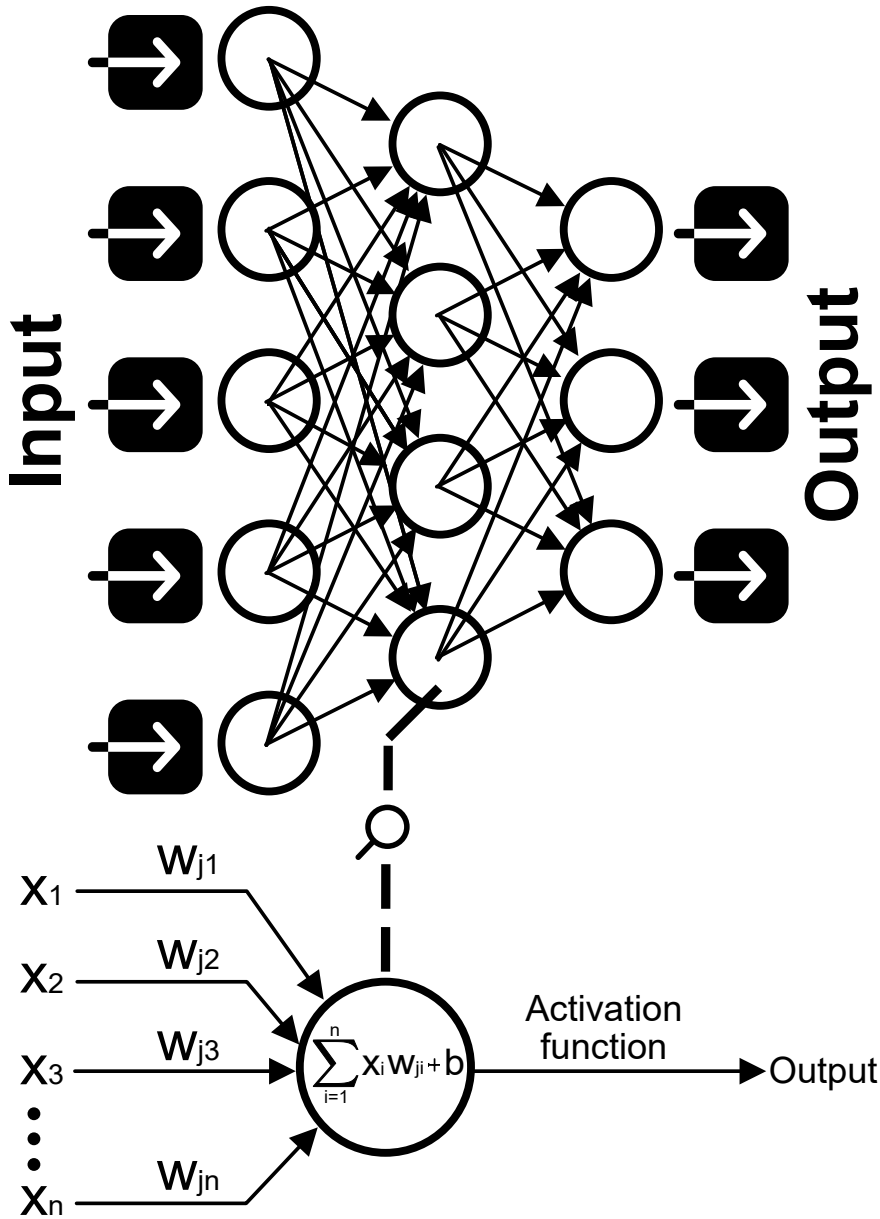

Figure 2.1: *Schematic of a fully connected neural network, in which each point x_i in the input tensor is connected through a weight w_{ji} to every neuron n_j in the next layer, as shown in the schematic on the bottom. A sequence of such layers is denoted as a deep (and fully connected) neural network.*

count toward the activation sum. This has shown to be a particularly powerful approach for images, in which information about the position and classification of objects is usually highly correlated within spatially proximal pixels. In practice, this is easiest accomplished through a convolution as shown in Figure 2.2.

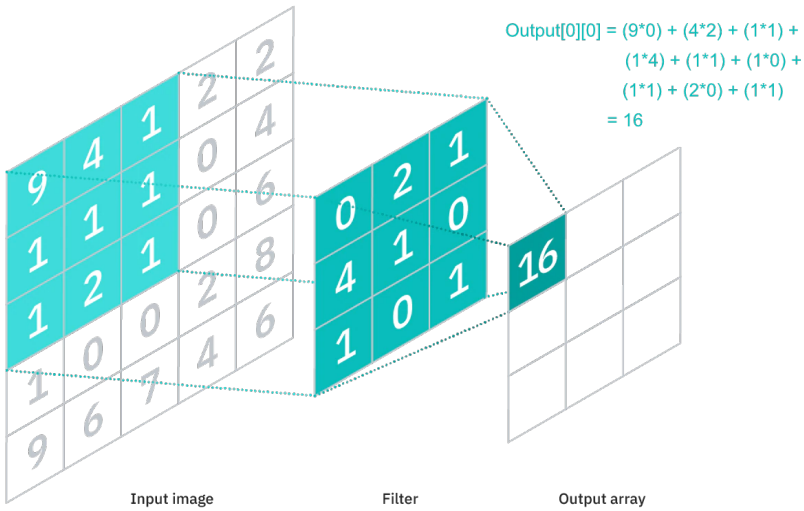

Figure 2.2: *Demonstration of the operation of convolution, wherein a segment of size $[ks_1, ks_2]$ of the input image is convolved with a filter of equal size. The operation consists of a dot product between the pixels of the input and filter, whereafter the filter shifts sideways and repeats the process until the entire image has been convolved. In a neural convolutional layer, the numeric values of the weights of this filter are learnt during training. Figure from [33], reproduced under the Creative Commons license.*

The weighting of each pixel in the convolutional layer is learnt during training, effectively applying the neuron's computation in equation 2.1 in a convolution rather than a linear fully connected fashion. When creating the layer, the engineer defines the kernel size, stride (of the sliding window), and number of filters. Filters work essentially as parallel channels all applying convolutions to the same input but with separately learnt weights, effectively developing separate "feature maps" which may represent different relevant features of the input. Thus, using convolutional layers, a full CNN uses various kernels to convolve the whole input into intermediately sampled feature maps, which are in turn convolved in an equivalent manner. Ultimately, this realization of neural networks enables fully rich feature representations of images whilst prospectively fully preserving locality and richly detailed information. Further, it has been shown that early layers of a CNN tend to learn basic features relevant for image analysis, i.e. shapes, edges and simple structures, whereas later layers learn more abstract features such as vehicle outlines and facial details [34].

2.2.3 Pooling Layers

The purpose of pooling layers is to reduce the dimensionality of data between subsequent convolutional layers, reducing the impact of, as coined by Prof. Bellman, *the curse of dimensionality* [35]. In short, as the dimensionality of data increases, the total volume of the information space increases so quickly that any available dataset becomes sparse, worsening still the problems of underspecification [19]. Hence, the downsampling of dimensionality which pooling layers provide results not only in less computational overhead but also has been shown to improve generalization, latent space featurization and faster training convergence [36].

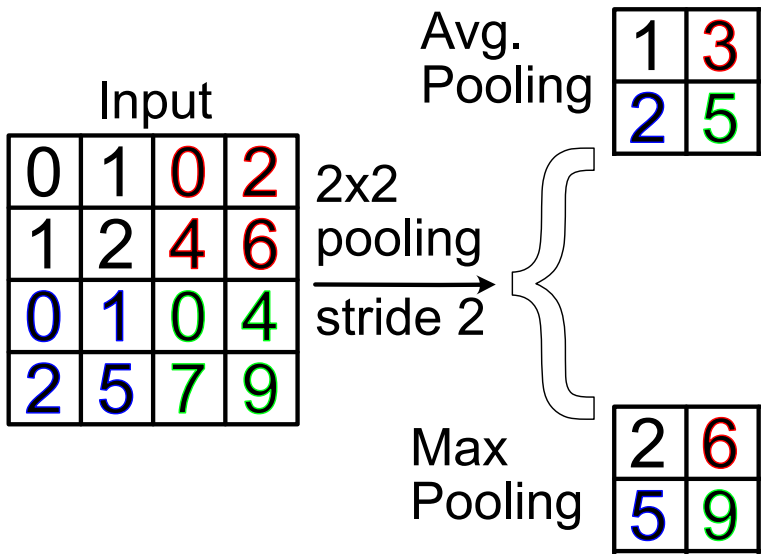

Figure 2.3: *Schematic of the pooling convolution, demonstrated here for average pooling on top and max pooling at the bottom for a 2x2 convolution passing over the input with a stride of 2. In average pooling, the output of each convolution is the average value within the corresponding pooling window. In max pooling, the output is instead the maximal value within said pooling window.*

In Figure 2.3, an example of max- and average pooling layers is shown. Note that though these are the most common pooling layers, and used in most standard implementations of CNNs, there are several others [37]. For instance, stochastic pooling for network regularization [38] and spatial pyramid pooling to remedy certain architectures' strict sizing requirements [39].

2.2.4 Convolutional Architecture

Most convolutional neural networks in use today use some combination and configuration of the convolutional, pooling and fully connected layers within their architecture. To reiterate in concrete terms, each CNN architecture uses these layers to accomplish three features: localized receptive fields (convolutional layers), weight sharing, and spatial subsampling. Localized receptive fields and weight sharing enable the introduction of inductive bias into our models in the form of translational invariance, which promotes the learning of general features relevant for (translationally invariant) imaging data. Each subsequent convolutional layer in a deep convolutional architecture receives inputs from a set of neighbouring units from the previous layer. In practice, this configuration has been shown [10] [34] to effectively allow the correlations learnt in each subsequent layer to grow in complexity; i.e., early layers learn basic abstractions, edges and shapes, whereas later layers learn higher-order representational features. Elementary features are often equally important regardless of where they appear in the image; hence local receptive fields with tied weights are appropriate [40].

In a traditional convolutional architecture, each convolutional layer is followed by a pooling layer to reduce the problems related to the curse of dimensionality [35]. Following several sequences of convolutional and pooling layers, the final layer of highest-level reasoning occurring within the network is usually achieved via fully connected layers. The neurons in fully connected layers are connected to every other neuron in its two neighbouring layers, hence their name, and their output is simply computed as an activation function of a matrix multiplication with a bias offset as also described earlier in this section. This generic architecture effectively enables the definition of a function approximation whose input is a data type with locally relevant correlations (i.e. an image) and whose output is a vector of numbers. By connecting this vector of numbers to a regression (i.e. mean absolute error) or classification (i.e. cross-entropy) [26] loss when compared with labelled data, and then *backpropagating* the numeric value of loss through the network's weights and biases in the opposite direction of the gradient, we arrive at a trainable convolution-based network architecture for generic image-to-vector translation. Note that this output does not necessarily have to be the final output of a full model architecture - it could rather be considered a latent feature vector for further processing through another network type or standard methods [41]. Note that it is also possible to control the nature of the output data type by modifying the output layer of the network architecture - i.e. a convolutional layer instead enables learning an image-to-image translation function approximation. Herein lies one of the greatest strengths of neural network architectures in general: changing the architecture to accommodate different data types, i.e. sequences, graphs, images or tabular data, is relatively straightforward with modern powerful deep learning frameworks like Keras [42] and PyTorch [43].

2.3 Attention Mechanism

The attention mechanism describes a recent new group of layers in neural networks that has attracted a lot of interest in the past few years, especially in sequence tasks [44]. There are a lot of different possible definitions of "attention" in the literature, but the one we will use here is the following: the attention mechanism describes a weighted average of (sequence) elements with the weights computed based on an input query and elements' keys. The goal is to take an average over the features of multiple elements. However, instead of weighting each element equally, we want to weight them depending on their actual values. In other words, we want to dynamically decide on which inputs we want to "attend" more than others. In particular, an attention mechanism usually has four parts we need to specify:

Firstly, the *query*, which is a feature vector that describes what we are looking for in the sequence, i.e. what might we want to pay attention to. Secondly, the *keys*, where we have a key for each input element which is again a feature vector. This feature vector roughly describes what the element is "offering", or when it might be important. The keys should be designed such that we can identify the elements we want to pay attention to based on the query. Thirdly, *values*, where we also have a value vector for each input element. This feature vector is the one we want to average over. Finally, the *score function*, which is used to rate which elements we want to pay attention to. The score function takes the query and a key as input, and outputs the score/attention weight of the query-key pair. It is usually implemented by simple similarity metrics like a dot product, or a small MLP.

Given a sequence of input vectors $X = \{x_1, x_2, \dots, x_n\}$, the attention mechanism first transforms X into three different spaces to obtain queries (Q), keys (K), and values (V) using three weight matrices W^Q, W^K, W^V respectively:

$$Q = XW^Q, \quad K = XW^K, \quad V = XW^V \tag{2.2}$$

The weights of the average are calculated by a softmax over all score function outputs. Hence, we assign those value vectors a higher weight whose corresponding key is most similar to the query. The self-attention mechanism computes the attention scores between each query and all keys to determine how much focus to put on other parts of the input sequence for each element in the sequence:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{2.3}$$

where d_k is the dimensionality of the key vectors, which scales the dot products to prevent them from growing too large. The softmax function is applied to the rows of the resulting matrix, ensuring that the weights sum to 1 and can be interpreted as probabilities. The core of the transformer's attention

mechanism is the scaled dot-product attention. The scaling factor $\sqrt{d_k}$ is crucial as it prevents the dot product from growing too large in magnitude, leading to gradients that are too small for effective learning. To allow the model to jointly attend to information from different representation subspaces at different positions, the transformer uses a mechanism called multi-head attention. This involves running several attention mechanisms in parallel, concatenating their outputs, and projecting the result with another learned linear transformation:

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \ldots, \text{head}_h)W^O \tag{2.4}$$

where each head is defined as:

$$\text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V) \tag{2.5}$$

and W^O is the weight matrix for the output linear transformation.

This mechanism allows the transformer to capture various aspects of the input data and their interdependent correlations, making it extremely powerful for tasks requiring an understanding of general long-range dependencies within data. The ability of the attention mechanism to focus on relevant parts of the input sequence has been a key factor in the success of transformers across a wide range of applications, from language understanding and translation to image recognition and beyond.

2.3.1 Transformers

Moving forward, we will explore the implementation of the multi-head attention mechanism within the Transformer architecture. The transformer, introduced in 2017 by Vaswani et al. [44], is a particularly influential example of an attention-based architecture. Initially conceived for the purpose of machine translation, the Transformer employs an encoder-decoder setup. The encoder processes the input sentence from the source language, creating an attention-informed representation. Conversely, the decoder, mirroring the behavior of a traditional recurrent neural network, uses this encoded data to autoregressively generate the sentence in the target language. Although this configuration is highly effective for Sequence-to-Sequence tasks requiring autoregressive decoding, our discussion will predominantly center on the encoder component. The adoption of purely encoder-driven Transformer models has spurred significant progress in the field of natural language processing [45], [46].

The transformer architecture is again based on the idea of self-attention, where the model learns to weigh the importance of different parts of the input data. This is particularly useful for tasks where the input data is not spatially ordered, such as natural language processing. Transformers revolutionize sequence-to-sequence tasks by employing a self-attention mechanism that allows for parallel processing of sequences and capturing long-range dependencies without the

sequential computation limitations of RNNs and LSTMs. The transformer model consists of an encoder and a decoder, each comprising a stack of identical layers. The architecture's core is the attention mechanism, which, in the context of transformers, is mainly the multi-head self-attention mechanism described previously. Each layer in both the encoder and decoder consists of sub-layers, including multi-head self-attention mechanisms and position-wise fully connected feed-forward networks. Normalization is applied before each sub-layer, and residual connections around each of the sub-layers help facilitate deep network training, as:

$$\text{LayerNorm}(x + \text{Sublayer}(x)) \tag{2.6}$$

where $\text{Sublayer}(x)$ is the function implemented by the sub-layer itself. For the encoder, this includes multi-head attention and feed-forward networks, while the decoder includes an additional multi-head attention layer that attends to the encoder's output. Since transformers do not use recurrence or convolution, positional encodings are added to the input embeddings to provide some information about the position of the tokens in the sequence. This is given by:

$$PE_{(pos,2i)} = \sin(pos/10000^{2i/d_{\text{model}}}) \tag{2.7}$$

$$PE_{(pos,2i+1)} = \cos(pos/10000^{2i/d_{\text{model}}}) \tag{2.8}$$

where pos is the position, i is the dimension, and d_{model} is the dimensionality of the token embeddings. This allows the model to use the relative or absolute position of the tokens in the sequences. The encoder's self-attention layers allow each position in the encoder to attend to all positions in the previous layer of the encoder. Similarly, in the decoder, self-attention layers allow each position to attend to all positions up to and including that position in the decoder. In the decoder, the "encoder-decoder attention" layer allows the decoder to focus on different parts of the input sequence, similar to the attention mechanism in sequence-to-sequence models with RNNs. This is crucial for tasks like translation, where the model needs to focus on the relevant part of the input sequence for each word in the output sequence. In addition to attention mechanisms, each layer in the encoder and decoder contains a fully connected feed-forward network, which is applied to each position separately and identically. This consists of two linear transformations with a ReLU activation in between.

$$\text{FFN}(x) = \max(0, xW_1 + b_1)W_2 + b_2 \tag{2.9}$$

Transformers' ability to process entire sequences simultaneously, combined with their reliance on attention mechanisms, makes them highly efficient and effective for a wide range of sequence modeling tasks. Their architecture facilitates deep network training and allows for significant parallelization, leading to substantial improvements in training time and performance across tasks.

2.3.2 Vision Transformers

Vision Transformers (ViTs) extend the transformer architecture, initially developed for natural language processing, to the domain of computer vision [47]. Unlike traditional CNNs, which as mentioned rely on local receptive fields to extract spatial hierarchies, ViTs apply self-attention mechanisms across image patches to capture long-range dependencies and global context simultaneously. This shift in paradigm has demonstrated competitive performance in image analysis tasks, scaling particularly well with increased computational power and data, challenging the dominance of CNNs in computer vision.

In the previous section, we discussed how transformers are highly effective in handling sequential data due to their ability to dynamically learn relationships between inputs through attention mechanisms. However, a major drawback is that they require n^2 pairwise computations for an input sequence of length n, making it challenging, or even impossible, to process large inputs with current hardware. This issue is especially pronounced in image analysis, where images typically have a size of $n \times n$, leading to a total computational load of $(n^2)^2$. For example, processing a single 256×256 image would necessitate billions of computations using a transformer. The Vision Transformer (ViT) addresses this problem by breaking down the image into smaller patches and processing them sequentially. As illustrated in Figure 2.4, the image is first divided into smaller sections, which are then flattened, encoded with positional information, and passed through a standard transformer model as previously described.

The core idea behind ViTs is to represent an image as a sequence of fixed-size patches, analogous to the tokens in a sentence for natural language processing. Specifically, an image of size $H \times W \times C$ (where H and W are the height and width of the image, and C is the number of color channels) is split into N patches, each of size $P \times P \times C$. Each patch is then flattened into a vector and linearly projected to form a token embedding, similar to how word embeddings are handled in NLP transformers. These embeddings are fed into a standard transformer encoder.

Each input image patch x is projected to a d-dimensional embedding through a learnable linear projection matrix E:

$$z_0 = [x_1 E; x_2 E; \ldots; x_N E] + E_{\text{pos}} \tag{2.10}$$

where z_0 is the input sequence, and E_{pos} is a learnable positional embedding that encodes the location of each patch, preserving spatial information, since transformers inherently lack built-in spatial inductive biases.

The transformer encoder processes this sequence of patch embeddings using multi-head self-attention layers, where the self-attention mechanism operates globally across all patches, allowing the model to capture relationships between distant regions of the image. This capability to model long-range dependencies is a significant advantage of ViTs over CNNs, which are inherently limited by the locality of convolutional operations.

Vision Transformer (ViT)

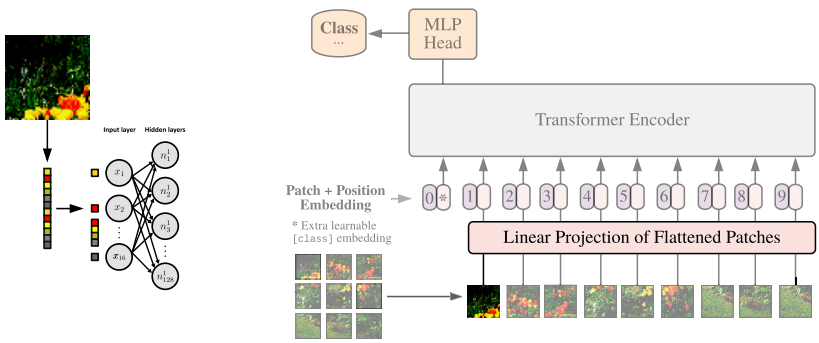

Figure 2.4: Vision transformer. *We transform an input image into a sequence of patches, which are then flattened and combined with a positional embedding layer. The Vision Transformer also includes an additional learnable node (denoted with *) at the 0th index, which gathers information from the rest of the sequence and provides a global summary. This embedded sequence is then fed into a standard transformer encoder, as outlined in the corresponding lecture. The output from the 0th index of the transformer encoder is subsequently passed through a dense MLP head for final classification.*

The output of the transformer encoder, which represents contextualized patch embeddings, is typically passed through a classification head for image classification tasks. A special class token is prepended to the input sequence, and its corresponding output embedding is used for the final classification.

$$\hat{y} = \text{MLP}(z_{\text{class}}^{L}) \tag{2.11}$$

where z_{class}^{L} is the final embedding of the class token after L layers, and MLP is a multi-layer perceptron that outputs the predicted class.

One of the key advantages of ViTs is their scalability. By increasing the model size (number of layers and attention heads) and the amount of training data, ViTs have been shown to outperform traditional CNNs on large-scale image datasets such as ImageNet [27]. Additionally, ViTs benefit from pretraining on massive datasets, enabling strong transfer learning capabilities.

However, ViTs have also been noted to require significantly more data and computational resources for training compared to CNNs. Their performance on smaller datasets is generally inferior without extensive data augmentation or transfer learning. Nevertheless, ViTs represent a powerful alternative to convolutional approaches, particularly in tasks where capturing global context and long-range dependencies is critical.

Effectively, Vision Transformers apply the principles of the transformer architecture to visual tasks by treating image patches as tokens and leveraging global self-attention mechanisms. This approach offers competitive performance in image classification and shows promise in expanding the use of transformers in various computer vision applications beyond natural language processing.

2.4 Training Methods

Now that we have a better idea of the underlying algorithms used in this thesis, it is natural to delve deeper into the specific training methods which are vital to reaching the high accuracy and precision and low LoD of the models used in this thesis. In particular, this section describes transfer learning; the process of transferring learnt knowledge from one network to another, ensemble modelling; the process of including the predictions of several neural networks in a given task, and curriculum learning; the process of successively increasing the difficulty of the task during training to be learnt to facilitate learning in very challenging conditions.

2.4.1 Transfer Learning

Transfer learning is a technique pioneered in the fields of AI and DL that involves transferring knowledge gained while solving one task to another (similar but not identical) task. The underlying idea is that some of the knowledge and skills learned while solving the first problem can be applied directly to the second problem, resulting in improved performance and shorter training time.

It has been shown that a neural network trained on a particular type of dataset is much better at quickly recognizing new instances of similar data, when compared to a completely untrained network [48]. For instance, though a neural network may need thousands of examples of cars to start to accurately recognize cars, such a network can learn much quicker to recognize other vehicles when introduced into its training set compared to an untrained network [49]. This is a consequence of, as mentioned previously, neural networks learning increasingly complex abstractions in deeper layers of the network. Thus, a majority of the learning a network has to do to solve any given task is simply to find a good representation of the data, and transfer learning can thus be achieved by "freezing" the training of the first layers of a neural network and only training the last few layers on the new data.

There are several benefits of this approach, most evidently that it can help reduce training time and data requirements for new tasks. Another benefit is that it can improve generalization performance, meaning that models trained using transfer learning are more likely to achieve good results on unseen data than those trained from scratch specifically for the new task at hand. Additionally, because transfer learning typically relies on pre-trained models which have been already tuned for good performance on a wide range of tasks, it can help avoid overfitting problems which can plague neural networks when they are retrained from scratch for a specific purpose.

Despite its many advantages, there are also some potential drawbacks to consider with transfer learning strategies. One issue may be poor adaptation if there are significant differences between how tasks were originally learned and how they need to be applied later on, and special care should be taken when applying pre-trained models so as to not inadvertently introduce bias into results due to inaccuracies or irregularities in how the old network was transfer-learned into the new.

2.4.2 Ensemble Modeling

Ensemble modelling is a generic term meaning incorporating several models' predictions to improve accuracy, precision and overall generalizability of a final (ensemble) model. This technique has been used in traditional statistical methods for many years [50], but is equally appropriate in (deep) machine learning owing to the typically fast inference time of individual models. In general, models in an ensemble can be trained on different datasets or different permutations of the same datasets, be defined as identical architectures or variations of the same underlying architectures, or even be entirely based on completely separate architectures or algorithms. In principle, an ensemble model can even be any combination of the above alternatives.

There are several advantages of this approach [42], dependent fully on its specific implementation, but in general it's usually used to improve accuracy and generalizability by mitigating the consequences of exploding state spaces [51] and underspecification [19] and bias. In short, models trained on different permutations (or sometimes even the same) of the same dataset tend to learn

different biases in different directions, such that their individual biases all cancel out with a large enough ensemble of models. Another related benefit is that of stability, wherein the variability of an ensemble model across different datasets or with changed parameters is decreased compared to using a single model. Additionally, ensemble modelling has been shown to be able to identify patters than may not be visible to any individual module [52]. In essence, combing information from several different models may reveal otherwise hidden patterns, making them especially useful for complex problems where traditional methods struggle to find accurate solutions.

2.4.3 Curriculum Learning

Another technique of significant impact to this project is curriculum learning [53], wherein we take full advantage of the benefits of transfer learning. Human beings often do not learn anything when approached with an insurmountable challenge, learn only the minimally required information when approached with a trivial problem, and learn optimally only when approached with a problem perfectly tuned to be slightly challenging given their particular skillset [54]. Neural networks, interestingly, have been shown to learn in analogous ways [34]. To take advantage of this, we may implement active learning [55] or curriculum learning schemes, wherein we vary which portions of the datasets are fed into the neural network during training to tune the challenges of learning, for instance by automatically feeding in more examples of data-points in regimes where the network retains low accuracy. As mentioned, neural networks need to train on a large amount of data only to learn optimal representations of the dataset in its initial layers, but once these representations are learnt, transfer-learning it on new data is usually only a question of tuning the specifics of the last few layers' learnt abstractions. To take advantage of this, we might employ a curriculum learning scheme wherein we train the networks with the easiest examples or the ones we have the most data-points of first, and successively increase the difficulty of the dataset as the network's loss decreases.

2.5 Quantitative Deep Learning for Optical Microscopy

Perception-based tasks in which data are collected under conditions which are consistent across training and inference, are where the greatest strengths of deep learning truly manifest [40]. The difficulties of deep learning; its black-box nature, its oft-brittle performance [56] and fickle unpredictability inferring on datasets orthogonal to its training set due to underspecification [19] and high sensitivity to poorly labelled data [57], are mostly avoided in these applications.

In practice, DL-based algorithms often outperform handcrafted algorithms [12], [58], as the heuristics necessary for complex image analysis are often limited by technical expertise, technological limitations (e.g. processing speed for real-time applications) and ultimately limits of human ingenuity and imagination.

Through powerful representation learning, deep learning-based AI systems can learn features more representative and relevant than other approaches [18].

Intuitively, this may be understood as a result of the complexity inherent in nature being difficult to capture in simplified heuristics or theoretical models alone, whereas data-driven DL approaches can learn directly from minimally processed measurements of physical systems only with limited inductive bias. Thus, DL systems trained in a supervised fashion can be interpreted in practice as an opportunity to draw direct statistical correlations between datasets without the potentially limiting intervention of researcher bias, an oft-cited reason for why promising yet unfashionable theories struggle to gain ground in physics [59]. In this section, I describe a few uses of CNNs in practice, and particularly how they have proven useful for quantitative analysis within optical microscopy [60].

2.5.1 Semantic Image Segmentation

The purpose of (semantic) image segmentation is to convert the pixel-wise information in an image to the class of the object in the pixel with a given probability. In classic image segmentation, this often reduces to outputting the probability in each pixel containing any of a pre-defined set of classes, as exemplified in Figure 2.5a.

For purposes of optical microscopy, the benefits of this application quickly becomes clear; given a complex image of several cellular components and different types of biological molecules, we can output a probability in each pixel of that pixel containing a molecule or component of a certain class, as exemplified in Figure 2.6. Thus, the complexity of the image is drastically reduced, enabling biologists to analyze the data in a far faster and more intuitive manner. This is especially relevant in the field of biomedical image diagnosis [62], which is a pioneering field of applied DL, and an example of a human-in-the-loop AI system where human experts and AI work in symbiosis to accelerate experts' work. In particular, this approach forms the backbone of cross-modality transformations in optical microscopy, wherein one aims to employ DL-based algorithms to virtually transform images taken with one modality into another [13]. For instance, the virtual staining [13] of cellular components allows one to bypass the manual chemical staining of patient samples, prospectively saving time when diagnosing patients under the effects of anesthesia. This is further enabled by the property of semantic segmentation architectures being able to predict not only the likelihood of existence of objects of interest in an image, but also the likelihood of said images belonging to a certain class along with other properties of interest.

2.5.2 Object Detection

Object detection is a massively popular application within machine learning, whose popularity and influence only seems grow every year through a wide variety of advances within computer vision. For instance, great strides have

Figure 2.5: *Examples from the Common Objects in Context (COCO) image dataset [61].* ***A*** *shows the result of semantic segmentation, wherein each image is "segmented" using DL such that every pixel of each separate object class is coloured in with a separate colour.* ***B*** *shows the result of object detection, wherein each separate object is localized with a so called "bounding box" with a corresponding probability, and classified with a separate corresponding probability.*

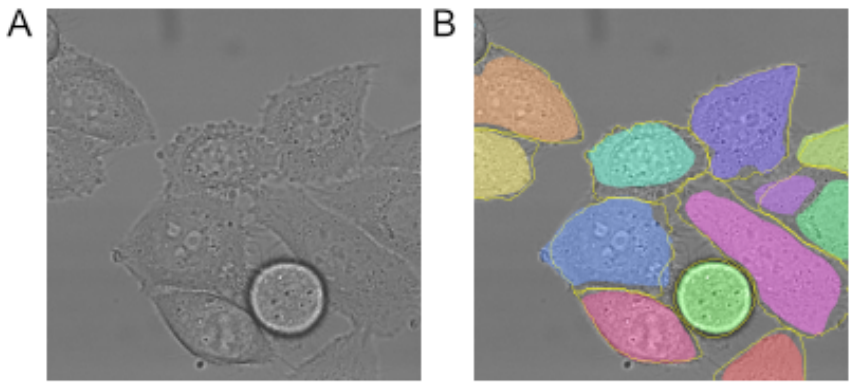

Figure 2.6: *A) Examples of HeLa cell microscopy images. B) Equivalent segmented images, where different types or structures of cells may be identified and coloured in different colours, through a U-net architecture. Figure reproduced with permission from [28].*

been made in facial detection [63], pedestrian detection [64] and autonomous driving [65] in the past few years alone. One of the most popular object detection algorithms leading these recent developments is that of You Only Look Once (YOLO), due to its impressive speed and higher generalizability than many other competing algorithms [66]. These traits are a consequence of the method's namesake, where a single feed-forward pass of an image through the CNN backbone is enough to predict (singular) bounding boxes and class probabilities for each object in said image [66].

The purpose of object detection is to classify, localize and separate individual objects within an image. In object detection applications, the exact classification of each pixel is not important. Rather, each image consists of an unknown number n of objects $o_1, o_2, ...o_n$ pertaining to class c_o from a pre-defined set of classes. A classic object detection algorithm trained in a supervised manner, can never output that an object belongs to a class outside its pre-defined list. Its function is exemplified in Figure 2.5b, where the benefit of object detection becomes clear; a segmentation model could not identify that the two elephants in the bottom left Figure are separate objects. Rather, it would fill in each pixel with a high probability of 'elephant', and contain no quantitative information on the number or position of the separate elephants. Object detection is useful in localizing objects within optical microscopy when pixel-wise segmentation information is unavailable or infeasible to extract [68], as exemplified in Figure 2.7.

2.5.3 Quantitative Property Estimation

It should be noted that whilst object detection and semantic segmentation are objectives which have come to dominate computer vision applications, convolutional neural networks can in principle learn to output any property of interest, as long as the property can be properly mathematically defined and the relevant information and correlations exist in the dataset. In general, these property calculating networks work using the same fully convolutional CNN backbone structure as for semantic segmentation and object detection, with the significant difference only manifesting in the data type which the *neck* and *head* layers of the neural network output. In particular, the last layer of the neural network, the head, is typically a dense layer with continuous outputs connected to a regression-based loss.

One of the most effective and influential CNN architectures for this application is the ResNet [30], whose deep skip connections between non-subsequent layers enabled deeper network architectures than previously possible, and whose subtle yet innovative change to calculate residuals rather than direct output value enabled numerically accurate calculations.

The most important takeaway from this section is that despite the differing applications, the underlying structure of the CNN remains the same. Connections between layers vary - between upsampling and downsampling layers in the U-net and between residually connected layers in the ResNet, but the backbone of convolutional and pooling layers in sequence is consistent.

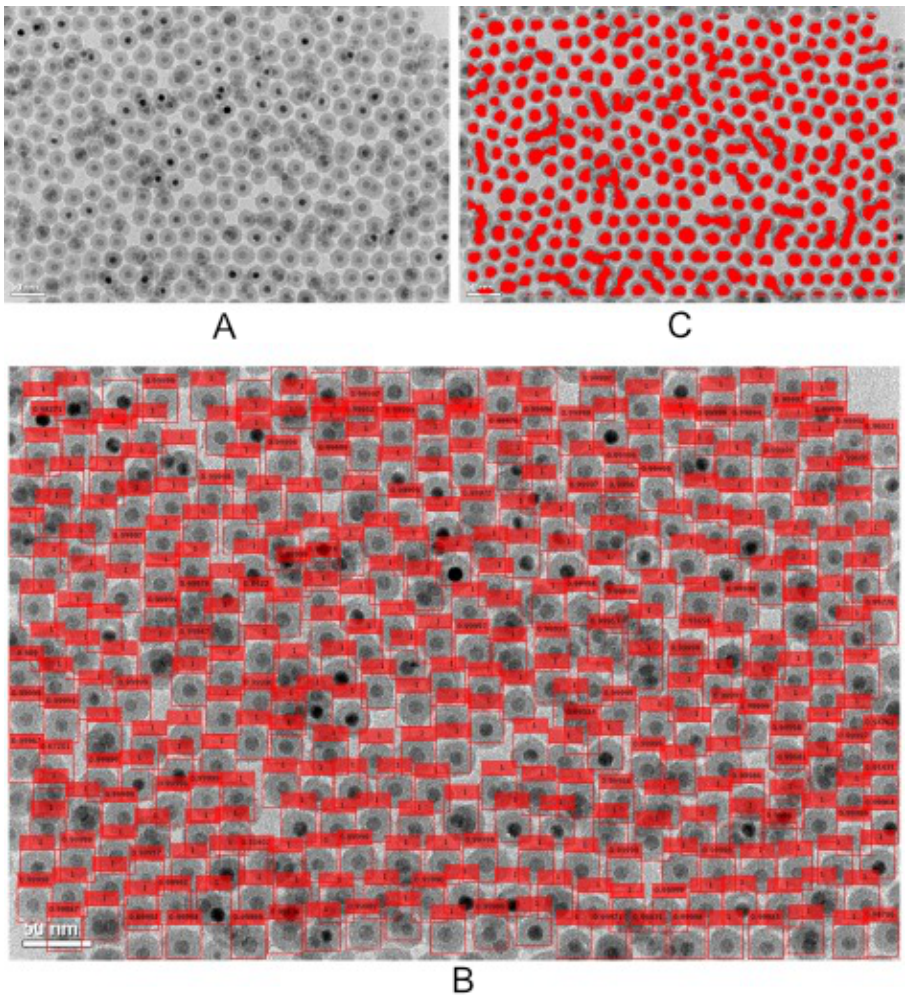

Figure 2.7: *A) Raw unprocessed image, consisting of nano-particle cells. B) The boundaries automatically detected by an object detection network, where each cell is identified with a so-called "bounding box". C) Centers of the detected cells, where the colour of each detection corresponds to the prediction accuracy of the corresponding box. Figure reproduced with permission from [67].*

2.6 Beyond Supervised Learning

So far, the contents of this thesis have been constrained to discussing the particular paradigm of machine learning known as supervised learning, wherein a neural network is used to approximate a function between input x and output y. However, this is likely not where the most significant future strides in deep learning will manifest. Though it is a powerful paradigm, it is significantly limited in the sense that it cannot automatically learn relationships between objects unless specifically pre-defined by their engineer. In particular, they cannot recognize new types of objects not pre-specified within the model definition, or for instance the relationship between words in different languages as is highly relevant for modern natural language processing [69]. Indeed, many of the current issues associated with DL are in practice rather a consequence of supervised learning [70]. Other paradigms of AI, such as unsupervised, self-supervised, generative and reinforcement learning, show much greater robustness beyond their training set and greater potential with smaller datasets (i.e. single-shot [71] or few-shot learning [72]). Therefore, to serve as a backdrop against the future outlook for further development of the work conducted within this thesis, and to introduce the reader to some of the most promising avenues of current AI research, this section is dedicated to describing paradigms beyond the standard of supervised learning.

Most straightforward to describe is unsupervised learning, wherein AI algorithms which do not require any labelled data (i.e. output y) are of definition. Rather, they automatically discover patterns and correlations only in input data x on their own during training. Classes of algorithms which are encompassed by this umbrella are e.g. clustering algorithms such as K-means clustering [73], which automatically cluster input data in a predefined number of clusters through appropriate parameters chosen by the network during training, and auto-encoders which are often used as anomaly detection models [74]. Most famously, some of the statistical analysis underlying the discovery of gravitational waves [75] were based on this approach: in short, an auto-encoder is trained to perfectly recreate the noise present in the LIGO interferometer device used to measure gravitational wave signals. The portion of LIGO device output which cannot be perfectly recreated by a such trained model is therefore either some entirely new source of noise (which thus can be corrected for with a new model) or quantitatively accurate gravitational wave signal down to a very low SNR [76].

Another promising paradigm is that of self-supervised learning, wherein we exploit the inherent symmetries in certain datasets to effectively teach a neural network-based algorithm to learn certain useful transformations within unlabeled datasets. For instance, the most influential use of self-supervised learning today is in that of training massive so-called attention-based transformer models, effectively neural networks particularly tuned to learn relationships between input tokens, on natural language processing. In short, one scrapes a large dataset (e.g. the entirety of Wikipedia), randomly removes singular words with some defined frequency, and trains the transformer model on filling in the

missing words. Such models, when trained on billions of data-points and containing billions of parameters, have been shown to be able to output human-like text in form of essays, blog posts and even scientific articles [77], [78]. Another example is the Low-shot deep Symmetric Tracking And Regression (LodeSTAR) [71] algorithm, which essentially learns to detect particles in microscopy images by training on a single crop of a single image of a single particle. It achieves this by applying several (randomly modified) affine transformations to said single image, and training a convolutional neural network to always recognize to identify the same particle despite all possible affine transformations and corruption schemes. Thus, the CNN learns robust features which allow it to always track the profile of the particle type in different imaging conditions and self-consistent transformations. Further, the natural continuation of the supervised object detection algorithms presented in this thesis is self-supervised vision transformers such as DINO [79], which can learn to identify and separate objects into an arbitrary number of unseen and undefined classes, thus providing a solution to one of the biggest issues with supervised object detection algorithms.

Reinforcement learning (RL) is another paradigm which has seen significant medial attention in recent years, most notably for its use in developing AI-based agents which can learn to play games to superhuman levels, notably beating grand-master Lee Sedol in Go [80] and eventually going on to beat grandmasters in inordinately complex video games such as Starcraft II [81] and Dota 2 [82]. Its use in real-world problems has of yet been underwhelming, owing to the difficult "sim2real" [83] gap associated with bringing algorithms trained on purely simulated data to work in the real world, as well as the difficulties associated with training RL-agents in the real world [84]. However, significant strides have also been made in this regard, including an agent which develops new floorplan designs for computer chips at orders of magnitude higher speeds than human engineers [85] and a fully autonomous robot which plans, executes and analyzes experiments on its own [86].

Finally, and of particularly important note for the field of optical microscopy and biomedical imaging, generative learning is a paradigm which enables AI to generate new datasets given some original training data set. This may at first stand in great contrast to the aforementioned paradigms where the goal is to find correlations within and between datasets and optimize for a loss(reward) function, but note that the function to optimize may well be defined as that of how to generate unique data indiscriminate from some original dataset. Said generative learning algorithms are able to create quantitatively accurate representations for complex objects and patterns in data, which can then be used for tasks such as image synthesis [87], text generation [88], and cell microscopy image generation [89] in combination with e.g. supervised, unsupervised or self-supervised learning. This essentially solves the issue of requiring exceptionally large datasets to train certain models, if only one can train a generative model to generate effectively infinite synthetic data.

Generative models are typically trained in a Generative Adversarial Network

(GAN) environment [90]; a system composed of two neural networks competing against each other in a zero-sum game. The first network, called the generator, tries to generate fake data that is indistinguishable from real data. The second network, called the discriminator, tries to distinguish between real and fake data through a binary value. The generator gets better at creating realistic data as it learns from feedback given by the discriminator.

2.7 Cross-modality Transforms in Biological Microscopy enabled by Deep Learning

As reviewed in Paper V of this thesis, one promising emerging topic in artificial intelligence in the context of microscopy for biological (and non-biological) applications is that of cross-modality transformations enabled by deep learning. In short, this entails the process of training a neural network architecture to transform images taken with one microscope modality to instead contain information as if it were taken by another modality. One example of this is histological staining, wherein we circumvent the time-consuming process of manually chemically staining tissues by transforming e.g. brightfield images of tissues to their chemically stained counterparts through generative learning (see Figure 2.8).

Similarly, another virtual staining cross-modality transform is that of using generative learning to fluorescently stain cellular and sub-cellular structures. In pharmaceutical drug screening and digital cytology, quantitative analysis of cell structures is essential: changes in the morphological properties of cell structures are effective readouts into the physiological state of a cell culture and its response under drug exposure [91], [92]. However, chemically staining samples with fluorescent dyes to make them available for study entails invasive and sometimes toxic procedures, potentially affecting cell health and behavior [93]. Additionally, phototoxicity and photobleaching, occurring during the acquisition of fluorescence images, limit the time scales available for live cell imaging [94].

To remedy these challenges, recent works have proposed the use of DL, and particularly generative learning, as an alternative to mitigate the inherent problems associated with conventional chemical staining [95]. In these works, chemical-staining and fluorescence microscopy are replaced with a virtual staining network that generates virtual fluorescence-stained images from unlabelled samples. The input to the staining network must encode insightful contrast of the various cell structures, providing the network with sufficient information to learn the transformation to the desired high-contrast, high-specificity fluorescently stained samples. Virtual staining of cells and cell structures has been achieved from various imaging modalities, including phase-contrast imaging [96]–[98], quantitative phase imaging [99], and holographic microscopy [100]. Moreover, some work [13], [94], [101] suggests that bright-field images contain the information necessary to reproduce different stainings, despite the detail of these images being largely limited by diffraction. Another way of

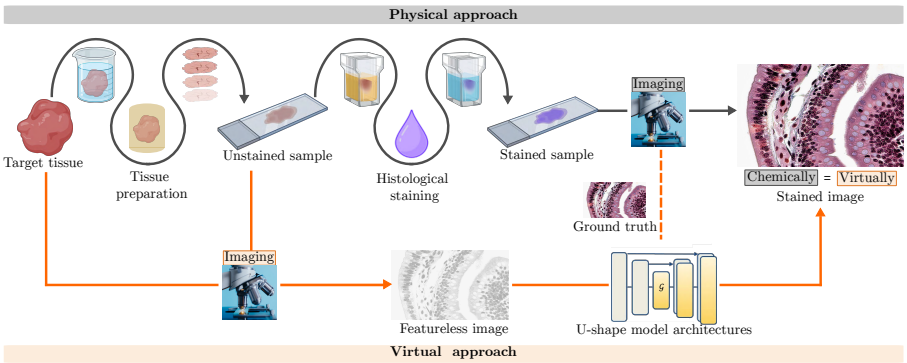

Figure 2.8: *Comparison of manual versus virtual methods for acquiring stained images. In the manual process, the specimen needs to be prepared, stained, and then photographed. Preparing the tissue is a meticulous procedure that may include steps like fixation, embedding, and sectioning of the original specimen. Likewise, preparing an unstained specimen for histological staining entails processes such as permeabilization, application of chemical dyes, rinsing, counterstaining, and fine-tuning the entire protocol until it is ready for photography. On the other hand, virtual staining techniques have simplified these processes, applicable to both tissue preparation and histological staining. With the virtual method, the specimen stays unchanged or unstained, and a virtual staining algorithm converts it into a stained image. The outcome is comparable for both techniques. Images that have been physically stained can serve as a reference for training the model, or as input instead of an unfeatured image, particularly when transforming between different stains is desired.*

overcoming said diffraction limit is by so-called *super resolution* techniques, which are another example of cross-modality transforms recently enabled by DL, as exemplified in Figure 2.9.

Many microscope techniques have been developed to overcome the diffraction limit, including single-molecule localization microscopy (SMLM) methods such as stochastic optical reconstruction microscopy (STORM) [104], photo-activated localization microscopy (PALM) [105] and fluorescence photoactivation localization microscopy [106]. Yet other methods of transcending the standard resolutions of microscopes exist, as mentioned, in the complex numerical estimations of point spread (transfer) functions (PSF) which seek to estimate the diffraction behaviour, in illumination pattern engineering methods which decrease the PSF size [107], as well as specialized flurophores [108]. These approaches come with their own limitations in terms of the complex and multivariate dependency on imaging conditions which make solving the diffraction integrals of the PSF very difficult for practically relevant systems [109], and increased costs associated with aforementioned flurophores [14].

In recent years, another promising paradigm of super resolution has emerged as a consequence of the astounding growth and success of DL-based computer vision algorithms. Allegorically to the cross-modality transforms above, the DL approach to super-resolution revolves around training neural networks to transform one imaging modality (regular-resolution images) to another (super-resolved images). Some of these approaches are through generative learning, i.e. effectively learning the complex interpolation function between regular- and super-resolved images, or through direct supervised learning e.g. by estimating the positions of underlying diffraction-limited emitters. The details on how these networks are trained in practice vary widely between applications, as elaborated upon further in Paper V. Of course, one would be remiss in this context not to mention that another way of effectively transcending the diffraction limit and studying individual molecules is to use powerful scattering-based microscopy methods like iSCAT or NSM.

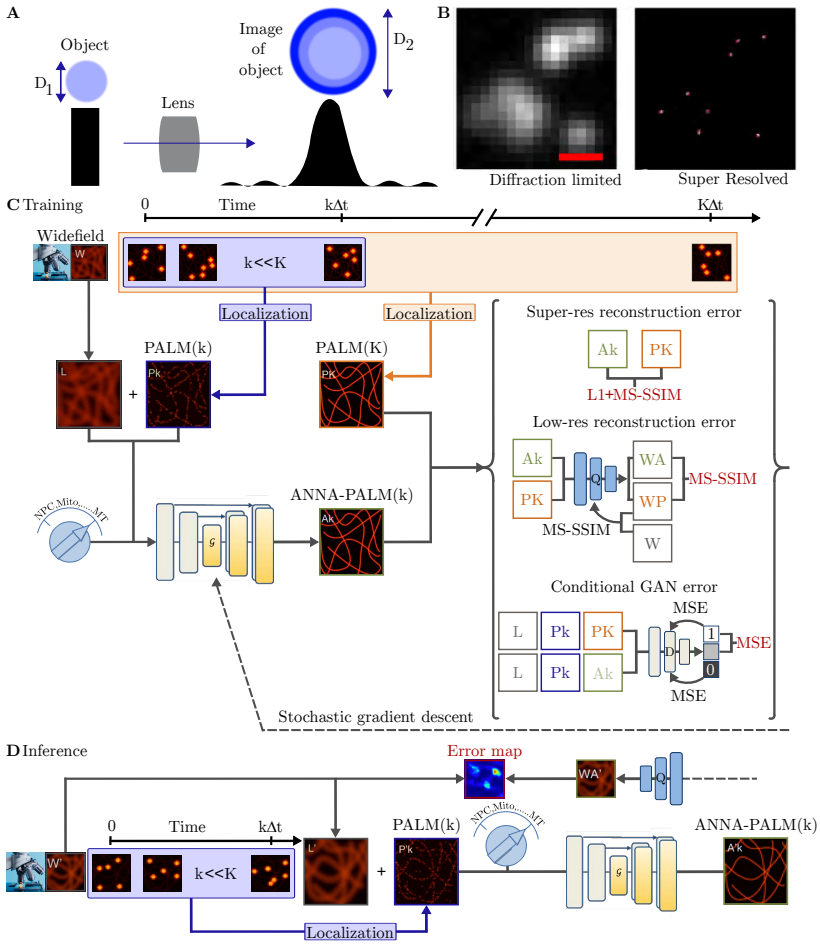

Figure 2.9: *A) Illustration of the PSF resultant of imaging objects below the diffraction limit of an optical system. B) Low-resolution image of simulated emitters and ground truth emitter locations. Image reproduced with permission from [102]. C,D) ANNA-PALM consists of two stages: firstly, acquisition of training images using standard localization microscopy (PALM) followed by neural network (ANN) training, and secondly reconstruction of super-resolution views and low-resolution error maps from new sparse PALM and/or widefield images at inference time. C) Training images are obtained by acquiring long sequences of diffraction-limited single-molecule images with standard PALM. Thereafter, standard localization microscopy algorithms are used to generate super-resolution images. Different types of structures can be distinguished, for example, nuclear pore complexes (NPC), mitochondria (Mito), or microtubules (MT). A GAN is trained by using the sparse PALM images as inputs and the corresponding dense PALM image as target output with the three error functions in the gray-bordered boxes. D) At inference time, a short sequence of low-resolution single-molecule images is acquired. Standard localization algorithms generate a sparse (undersampled) PALM image. This sparse image, the upscaled widefield image and switch setting are fed as inputs to the trained generator, which outputs a super-resolved ANNA-PALM image (A'k). Images C, and D reproduced with permission from [103].*

Chapter 3

DL for Nanofluidic Scattering Microscopy

This chapter is dedicated to introducing the key concepts required for understanding the results produced in Papers I, V and VI. It is not meant as an exhaustive review (see e.g. [110]); rather, its main purpose is to provide an intuitive understanding behind the most important tools of optical microscopy and deep learning which lie at the heart of this work. Specifically, the first section of this chapter introduces the basics of optical microscopy for molecular imaging, the method of nano-channel-based optical microscopy introduced in Paper I, important features of a select few biological molecules studied in this thesis, as well as touching upon the growing need for advanced data science within biotechnology and life science applications. The second section of this chapter introduces deep learning, specifically in the context of computer vision in general and for biological applications in particular.

3.1 Optical Microscopy for Molecular Imaging

In general, an optical microscope is a system which leads light in the visible spectral range through a sequence of lenses to magnify images of small objects. Though these systems have existed for a long time, arguably invented in the 1500s in their simplest implementation [111], it took many hundred years for the technology to become powerful enough to enable imaging of systems on a molecular scale. This may be understood as follows; the shortest physical distance between two points in an image, also known as the *spatial resolution*, is the single most important feature in optical microscopy [112]. The principal limitations impinging the potential of the spatial resolution of optical microscopy are imposed by an irreducible phenomenon; the physics of *diffraction*. Effectively, even in the absence of any imperfections in lens quality or optical component misalignment, the resolution of a microscope is ultimately proportional to the wavelength of scattered or transmitted light

being measured, and inversely proportional to the so-called numerical aperture (NA) of its objective. This happens since light from an emitting light source in a sample generates a fundamentally limited diffraction pattern known as a point spread function (PSF) when passing to the image plane through an object. By our definition above, the minimal distance between two points in a sample which can be resolved thereby becomes limited by the distance at which two overlapping PSFs can be resolved. This limitation may be transcended for instance through super-resolution imaging techniques, i.e. techniques which transcend the diffraction limit, of which DL-based CV methods already show great potential as outlined in Paper V and section 2.7.

Indeed, optical characterization methods have proven instrumental for fast and accurate measurement of biomolecule properties in- and outside of complex biofluids [113]. Historically, for studies which require single-biomolecule resolution, powerful microscopy techniques, such as fluorescence microscopy and iSCAT, fill this role [114]. However, fluorescence microscopy requires the molecules to be labelled by a clearly visible fluorophore or nanoparticle, arguably altering their properties and behaviours [14], and iSCAT requires binding them to a surface prior to measurement, thereby losing valuable information regarding their movement behaviour and leaving a large portion of them undetected. To remedy this, we have in Paper I introduced NSM, which relies on the sub-wavelength interference of light scattered from a nanofluidic channel and a nano-object, such as a biomolecule, inside it. Thus, we enable label-free mass and size screening of freely diffusing single-biomolecule trajectories, effectively constituting a leapfrogging step in the field of optical molecular imaging and characterization. The related field of optical microscopy in the context of molecular imaging is irrefutably immensely rich in history and complexity, and the full breadth of its nuances is not recounted in this chapter. Rather, it is given in its proper context for the purposes of the work accomplished in this thesis; in the form of our in-house developed NSM technique.

3.1.1 Nanofluidic Scattering Microscopy

NSM works by a supercontinuum laser irradiating visible light onto a nano-sized channel embedded in a silicon dioxide matrix within a nanofluidic chip containing single biomolecules freely diffusing in solution, as shown in Figure 3.1a. In this setup, as shown in Figure 3.1b-c, the biomolecule and nanochannel scatter light coherently into the collection optics of a microscope operated in dark-field mode, resulting in an improved optical contrast of the imaged biomolecules by several orders of magnitude. The total scattering intensity recorded from a region of a nanochannel of length $L = \frac{3\pi}{2k}$ can be approximated as (see section 1 in Supplementary Information of Paper I)

$$I_{\mathrm{t}} \approx I_{\mathrm{c}} + I_{\mathrm{m}} - \sqrt{2 I_{\mathrm{c}} I_{\mathrm{m}}}. \tag{3.1}$$

Here, I_{t} is the total scattered intensity from the nanochannel, I_{c} is scattered intensity from the channel and I_{m} is scattered intensity from the biomolecule

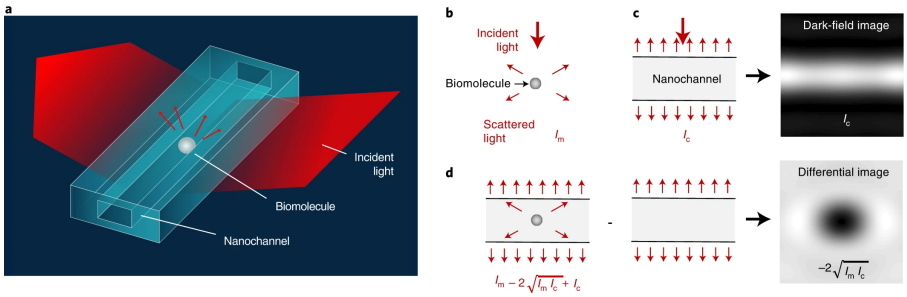

Figure 3.1: ***a****, Artist's rendition of the experimental configuration where visible light irradiates a nanochannel with a biomolecule inside, and where the light scattered from the system is collected in dark-field configuration.* ***b****, Schematic of light scattered by a single biomolecule.* ***c****, Schematic of light scattered by a nanochannel and the corresponding dark-field image.* ***d****, Schematic of light scattered by a nanochannel with a single biomolecule inside, and the corresponding differential dark-field image obtained by subtracting an image of the empty nanochannel from the image of the nanochannel with the biomolecule inside. Figure reproduced with permission from [115].*

(see [115] for further details). A video is then recorded of the total scattered intensity with a camera with high enough frame-rate to also be able to measure the biomolecules' diffusion along the nanochannel. By subtracting an image of the empty nanochannel from an image with a biomolecule inside, as depicted in Figure 3.1d, we acquire a differential image containing only the inference term $\sqrt{2I_c I_m}$ in equation 3.1 (I_m is vanishingly small since scattering from sub- wavelength objects scales as volume squared and the nano-channel's volume is much larger than the molecular volume). This inference term can be several orders of magnitude larger than the scattering intensity produced by the biomolecule alone (I_m) outside the nanochannel, which in essence constitutes the key feature of NSM and enables the direct imaging of diffusing biomolecules and other nano-objects inside a nanochannel. From videos of a sequence of differential images acquired in this way, so-called kymographs, the molecular weight (MW) and hydrodynamic radius (R_s) of each individual molecule can be measured.

Determining the MW of a biomolecule from its trajectory is made possible since the integrated optical contrast (iOC) of said trajectory is linearly dependent on the polarizability α_m of the biomolecule [116], where iOC is defined as

$$iOC = \int_{x=0}^{L} \frac{I_t(x) - I_c(x)}{I_c(x)} dx \tag{3.2}$$

for a channel of length L. This polarizability α_m is in turn proportional to $a \times MW$, where $a \approx 0.46$ Å^3/Da [117]. It is also inversely proportional to the cross-sectional area A of the nanochannel and proportional to $\bar{n} = \left(1.5n_{H_2O}^2 + 0.5n_{SiO_2}^2\right) / \left(n_{H_2O}^2 - n_{SiO_2}^2\right)$ where n_{H_2O} and n_{SiO_2} are the

refractive indices of water and silicon dioxide, respectively [115]. These two factors are constant and easily measured before an experiment, meaning that the MW of individual biomolecules measured in our system can be determined as

$$MW = iOC\frac{A}{\bar{n}a}. \tag{3.3}$$

Determining the R_s of a biomolecule from its trajectory is made possible since its diffusivity is correlated with its size. In particular, by approximating the biomolecule as a solid neutrally charged sphere, its R_s can be estimated using the Stokes-Einstein equation with a correction term K for hindrance effects arising from the diffusion of small objects in restricted volumes [118]. R_s is estimated as

$$R_s = K\frac{k_B T}{6\pi\eta D}, \tag{3.4}$$

where k_B is Boltzmann's constant, T is temperature, and η is the viscosity of the liquid in the nanochannel. By looking at a differential image of the nanochannel over time, even exceedingly minuscule biomolecules can be detected and analyzed, and by utilizing the fact that a biomolecule's optical contrast and diffusivity is correlated with its weight and size through its polarizability and the Stokes-Einstein equation respectively, the largest limiting factor for our analysis of smaller biomolecules quickly becomes the analysis of faint signals in overwhelming noise. To overcome these challenges, we employ DL-based approaches as described Chapter 2.

3.1.2 Nanofluidic Chip

The bases of NSM are the nano-fabricated chips containing the nano-channels within which the biomolecules are analyzed. A schematic of the nanofluidic chips used for our experiments is shown in Figure 3.2, which consists of a series of nanochannels with specific cross-sectional dimensions chosen to accomodate certain ranges of molecule size. The channels are connected to macroscopic in- and outlets through two microchannels, which are pressurized to 2 bar to facilitate sample transport. During measurement, this flow is usually turned off to let the molecules freely diffuse inside the nanochannels. In Paper I, three channel types were used: channel I with cross-sectional area $A_I = 100 \times 27$ nm^2 to study mainly proteins, channel II with $A_{II} = 110 \times 72$ nm^2 to study mainly DNA sequences, and channel III with $A_{III} = 200 \times 225$ nm^2 to study larger biological nanoparticles such as extracellular vesicles (EVs) which are further explored in the next section. In Paper VI, two channel types were used: one with size $A_{IV} = 30 \times 63$ nm^2 to measure molecules in the single-kDa regime, and one with size $A_V = 122 \times 97$ nm^2 to measure DNA sequences.

3.1.3 Biological Molecules and Nanoparticles

Biological molecules are ubiquitously present in all living organisms, existing at multiple different size scales and regulating essentially all biological functions

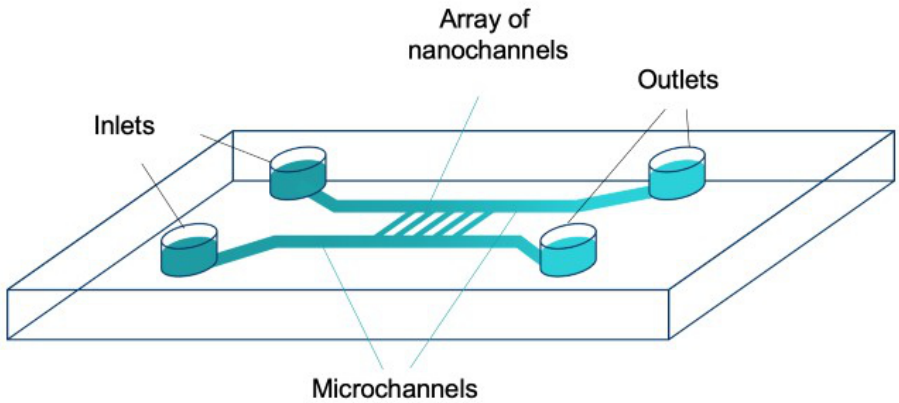

Figure 3.2: *Schematic of the nanofluidic chip. Two inlets connect to corresponding outlets through a series of nanochannels with different cross-sectional dimensions. A laser irradiates the nanochannel, wherein the imaged molecule always remains in focus, as a camera measures the light scattered from said nanochannel whilst biomolecules either freely diffuse or are pushed through with an induced convective flow. Figure reproduced with permission from [115].*

in both animal and plant bodies. Among the most important molecules are the proteins; a type of molecule comprised of amino acids, which influence cell structure, enzyme activity, molecule transport among many other functions [119]. In Paper I, we see evidence both of monomeric proteins, which consist of single molecules made up of one peptide chain, and polymeric proteins which are composed of either two peptide chains (dimeric proteins) or three peptide chains (trimeric proteins) [120]. Polymeric proteins consist of multiple copies (polymers) of a single protein molecule [120]. Studying these proteins and their polymeric aggregates are crucial for safe drug development and furthering our understanding of biological systems in general [119]. In Paper 1, we characterise the dimeric protein thyroglobulin, an important biomarker for thyroid cancer treatment [121], ferritin important for iron regulation [122], Bovine Serum Albumin (BSA) which regulates many biological processes [123], as well as Alcohol Dehydrogenase (ADH) which serves to break down alcohols that otherwise are toxic [124].

DNA strands, similarly, are considered polymeric since they contain multiple strands of a molecule bound together with hydrogen bonds between complementary nucleotides [120]. The study of DNA and the corresponding genes they encode is the source of several separate fields of biological study, playing a key role in everything from molecular [120] to evolutionary biology [125], as well as genomic DNA analysis [126] and genomic sequencing [127].

Extracellular Vesicles (EVs) are small lipid vesicles that are secreted from cells into the extracellular environment and can be found in all body fluids, including blood, saliva, urine and breast milk [128]. They play an important role in cell-

to-cell communication and have been implicated in a variety of diseases, such as Alzheimer's and Parkinson's and certain types of cancer [128]. EVs contain proteins, lipids and genetic material that can be used to identify the cell of origin. This makes them valuable tools for studying cellular function and disease pathology, i.e., by analyzing the RNA content of EVs it is possible to determine which genes are being expressed by a particular cell type. This information can then be used to develop new therapies for aforementioned diseases and cancers. Therefore, EVs have potential as diagnostic tools, i.e., if one could detect specific types of EVs in a patient's body fluid, then this would indicate that they're in a particular disease state, which would potentially allow doctors to diagnose several diseases earlier and begin treatment sooner than is currently possible [128]. Thus, EVs effectively act as mediators of physiological intercellular communication and have key functions in the pathobiology of many diseases [129]. However, their specific properties (surface markers, molecular cargo, weight, size, etc.) are difficult to measure precisely with existing metholodogies, such as dynamic light scattering [130] or nanoparticle tracking analysis [131], thereby necessitating new methods for their precise characterization. In Paper I, we show that NSM is one such promising method for characterization of EVs, exosomes and lipid nanoparticles in general.

3.1.4 Growing Need for Advanced Data Science in Life Sciences

The scale of modern life science has grown equally impressive and imposing in recent decades, as constantly improving automatized digital acquisition devices along with researchers' rapidly increasing knowledge of the molecular and genetic underpinnings of diseases is making the field inordinately complex. This presents a major challenge for traditional bioinformatics methods that rely on hand-crafted algorithms optimized for a specific dataset size or type. Advanced data science techniques are needed to handle this flood of genomic information effectively and extract insights from it. The availability of experimental data is often abundant, which transforms the challenge of analysis to "digging out some key piece of information from that morass of data", to borrow the words of Sir Roger Penrose [59]. To keep pace with this rapid evolution, advanced data science methods grow evermore central and crucial to the modern scientist's toolbox. In particular, there is a growing need for automatized algorithms which can quickly and robustly make sense of large and diverse datasets, as well as faint signals hidden in noise which is difficult or even infeasible to mitigate experimentally.

Fortunately, recent years have seen the advent of powerful data-science-focused languages such as R and Python, as well as AI platforms such as Tensorflow [132] and PyTorch [43] which are advancing at unprecedented speeds thanks to their open-source nature. As these tools continuously grow to be more democratized and easily available, they will effectively enable all scientists to perform sophisticasted statistical analyses and (deep) machine learning on their large and complex datasets. In addition, they provide a variety of libraries that

can be used for specific applications, such as gene expression analysis [133] or drug discovery [134].

In addition to handling large datasets efficiently, advanced data science methods can help identify patterns in complex biological systems [135]. For example, network inference algorithms can be used to identify modules within protein networks or transcriptional regulatory networks [136], or as famously featured in "Method of the year" in both Science and Nature and underpinning the 2024 Nobel Prize in Chemistry; to predict the 3-dimensional structure of proteins using limited information [137]. Such patterns may reveal new insights into how diseases develop and could lead to the development of new drugs or treatments.

Looking further ahead, the ability to optically monitor particles in the sub-10 kDa regime without labels is a tantalizing prospect, as it may enable a quick and cost-efficient analysis of subtle binding events, such as protein-DNA bindings [138] or nanovesicle corona formations [139] [140]. Beyond the fields of life science and biotechnology, it may potentially also enable the analysis of currently invisible contaminator particles in air [141] and catalytic reactions of single particles [142], suggesting significant research potential beyond the immediate consequences such an improvement in detection limits may bring for the field of biology. The study of small proteins is a field ready to emerge [15]; its abscence a symptom of lacking available methods for their characterization, rather than any lack of scientific interest. Studies have shown that sub-10 kDa molecules are ubiquitously present in many biological systems, particularly in membranes of cells, and may even regulate diverse processes, such as spore formation, cell division, enzymatic activities and signal transduction [15]. Further, there remain many questions regarding the fundamental nature of small protein interaction which are yet to be answered [15].

3.2 Deep Learning for Biomolecule Characterization in NSM

This chapter is dedicated to introducing and describing the implementation of the deep learning-based characterization methods used in Paper I and II. Specifically, the first section of this chapter first gives an overview of the entire DL-based analysis pipeline as it is implemented in Paper I, and thereafter delves deeper into the technical details of each separate method within. The second section describes some of the methods of training the networks which were crucial to reaching their final level of performance. Finally, the chapter ends with a description of how the biomolecule trajectories used as training data were simulated.

To track and analyze single biomolecules within nanochannels, which are the core of the NSM method, we implemented a DL analysis pipeline which takes advantage of all the paradigms of CNN-based CV. Image segmentation is used to identify the pixels in an image (kymograph) which contain the scattered light from the biomolecules within the channel. In fact, using a specialized network

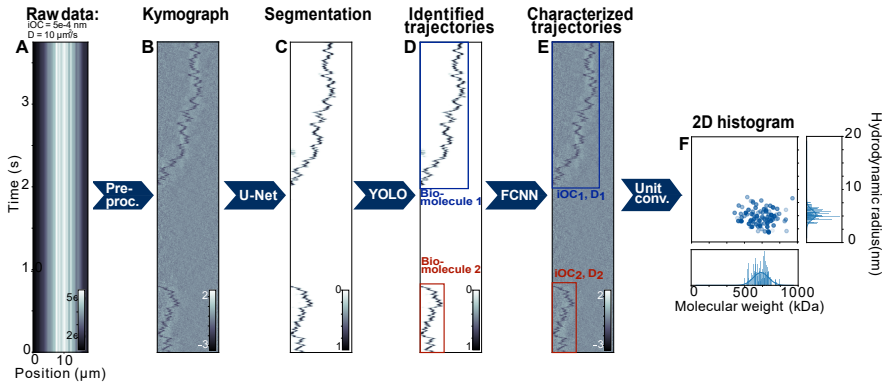

Figure 3.3: *Machine learning (ML) analysis workflow. (A) Raw image data simulated for a single biomolecule with optical contrast $iOC = 0.5$ nm and diffusivity $D = 10\ \mu m^2/s$. (B) Preprocessed kymograph with removed background. (C) Segmented image where the particle positions are detected using a U-net. (D) Single biomolecule trajectory identification using the YOLOv3 neural network. (E) Property calculation (iOC and D) using a custom fully connected neural network (FCNN). (F) Unit conversion from iOC to molecular weight MW and from D to hydrodynamic radius R_s plotted in a 2D histogram for illustrational purposes, for a simulated example molecule. Figure reproduced from [115].*

architecture developed for biomedical image segmentation known as a U-net [28], each output pixel corresponds to the *probability* of said pixel containing scattered light from a particle. Further, object detection is used to identify the separate trajectories and segment the kymographs into sections consisting only of single trajectories. Finally, a fully convolutional neural network is used to calculate the properties of interest - in principle any biomolecule property whose quantitative values can be derived from its trajectory - for each individual trajectory in turn. This pipeline is illustrated in Figure 3.3. It is comprised of five parts:

3.2.1 Pre-processing into Kymographs

The raw image data (Figure 3.3A) was pre-processed to transform it into kymographs (Figure 3.3B), by the following process. First, the intensity of the raw CMOS camera image data was normalized according to

$$\bar{I}(x,t) = \frac{I(x,t) - \langle I(x,t)\rangle}{\langle I(x,t)\rangle} \tag{3.5}$$

where $I(x,t)$ is the intensity at position x and time-frame t, and $\langle I(x,t)\rangle$ represents the time average of said intensity. Second, a low-pass-filtered version of $\bar{I}(x,t)$ was calculated by using two normalized sliding windows of sizes 200×1 and 1×200, and subtracted from $\bar{I}(x,t)$. Finally, to obtain the kymographs

used to calculate iOC, the resulting image was down-sampled by a factor of 4 in the length dimension through mean pooling. Instead, to obtain the kymographs used to calculate D, the image was normalized by its standard deviation before being down-sampled by a factor of 4 in the length dimension through mean pooling.

3.2.2 Image Segmentation

The purpose of image segmentation, as described in 2.5.1, is to transform the value of each pixel within an image to another property which is more conducive to the purposes of one's analysis. For the purposes of particle tracking, we want to transform our kymographs, consisting of the (differential) intensity of scattered light in each pixel, to "segmented" images which contain the probability in each pixel containing scattered light from a particle. This transformation is accomplished through implementation of a particular CNN architecture, whose function is described in section 2.2 and architecture summarized in Figure 3.4, trained in a GAN environment as described in section 2.6. Specifically, the GAN consists of two networks: a generator U-net which generates segmented images from input kymographs (pre-processed as described above), and a discriminator ConvNet which outputs a binary value whether the generated segmented image corresponds to the true trajectories or not. By modifying the architecture of this ConvNet, different behaviours of the U-net can be achieved. For instance, a relatively strong architecture which considers multiple scales of the input image, including filters which convolve the entire size of the input kymographs, is useful to force the U-net to always draw fully connected trajectories (broken trajectories are always unphysical, unless they are broken at the borders). Conversely, a relatively weak architecture, whose filters consider only local features of the image, can be used to force the U-net to only output trajectories in regions of the image where the local SNR is high. Generally speaking, the former approach effectively results in "ghost trajectories" in kymographs with heavy noise, as the generator U-net will always draw fully connected trajectories even in regions where the signal is low enough that it should be impossible. In these localized regions, the network tends to "guess" the likely trajectory path. The latter approach usually results in broken trajectories, since the generator will prefer to draw trajectories only where the signal is sufficiently high. The ConvNet used in this thesis was comprised of 5 convolutional layers of filter size 4, effectively corresponding to a compromise between the two aforementioned approaches.

The U-net was trained on ∼300,000 simulated kymographs (generated as described in section 3.3 and pre-processed as described in section 3.2.1) for which the corresponding ground-truth single biomolecule trajectory is known (using the ADAM optimizer [144]) with particle trajectories in the ranges $10^{-5} \leq iOC \leq 3 \cdot 10^{-3}$ μm, $1 \leq D \leq 100$ μm^2/s. The input during training are simulated images (kymographs) of size 128×2048 with a random number of trajectories and output segmented images of equivalent size. The model is also train-validated every 120 epochs ($\cong 30,000$ simulated kymographs) against 150 simulated kymographs (of size 128×512, 128×1024, 128×2048) with

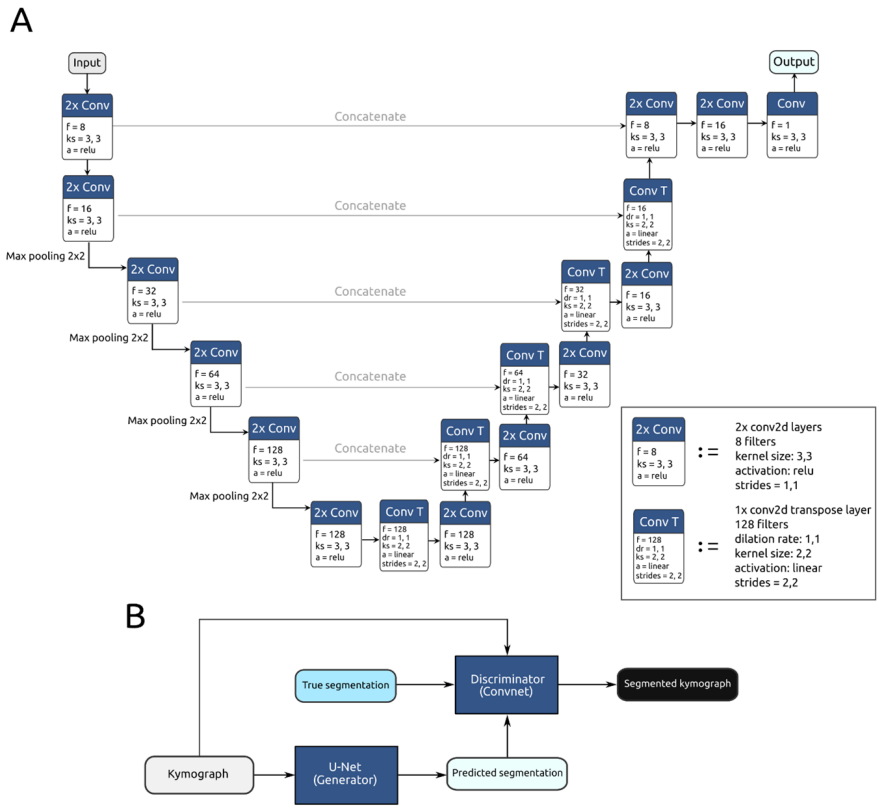

Figure 3.4: *The U-net[28] consists of a series of contraction convolutional layers, a bottleneck, and a series of expansion convolutional layers, as well as a series of skip connections between corresponding contraction and expansion convolutional layers to ensure that information learnt during contraction is not lost at the bottleneck. Each 2x Conv box represents a convolutional block corresponding to 2 convolutional layers in sequence, and each Conv T box represents a single convolutional transpose layer as exemplified in the legend. Here, f is the number of filters in each block, ks is the kernel size, a is the activation function, and dr is the dilation rate. The network is visualized with Netron [143]. (B) Block diagram of the GAN training environment, where kymographs are fed both to a U-net generator network, which predicts a corresponding segmentation, and a Convnet (convolutional network) discriminator network, which takes both original kymographs and true segmentations as input to determine whether the predicted segmentation is correct. A basic Convnet consisting of 5 convolutional layers of size 4x4 and stride 1 connected to a single dense layer is used as discriminator network. Figure reproduced with permission from [115].*

experimentally measured channel noise, using an 80-20 train-validation split.

3.2.3 Object Detection to Identify Trajectories

The next stage of analysis is object detection, to accomplish which we implemented the YOLOv3 [29] (Figure 3.5) algorithm in PyTorch [146]. The purpose of this stage of analysis is to identify and separate individual biomolecule trajectories in a segmented kymograph output from the U-net described in section 3.2.3, which only contains information of the probability in each pixel that it contains the scattered light from a biomolecule. If a kymograph cannot be separated in this manner (for instance as a result of two biomolecule trajectories strongly overlapping each other), the YOLO algorithm instead counts the number of inseparable single biomolecule trajectories in the given kymograph, outputs this value, and disregards the inseparable trajectories as the errors in scattering resultant of particle interactions is not taken into account in any calculations. Concretely, the YOLOv3 algorithm fits a minimally-dimensioned rectangular bounding box around each separate single biomolecule trajectory, and if the minimal bounding boxes of two single molecule trajectories overlap more than the set threshold $b_t = 60\%$ (roughly corresponding to the point where they are more entwined than not), their bounding boxes are merged to a minimal rectangular bounding box around both of them that is labelled as containing two biomolecules. This process runs recursively until all boxes are either combined or separated.

The YOLO algorithm is usually built upon a pretrained backbone, such as a ResNET [147], GoogleNET [148] or in the case of this implementation, a DarkNET [149]. In particular, we use the implementation known as DarkNet-53, its namesake arising as a result of its architecture consisting of 53 convolutional layers. This backbone is then attached to its *neck*, whose purpose is to form features from the base convolutional neural network backbone. The YOLOv3 algorithm uses something akin to a Feature Pyramid Network (FPN) as neck, adding a few convolutional layers in sequence which connect into an eponymous yolo (detection) layer, which outputs a tensor containing bounding boxes, the probability of an object being inside each bounding box and the probability of the object being of a certain class. The feature map from the second prior layer (i.e. two layers back from the detection layer) is upsampled and subsequently merged with a much earlier layer in the network through concatenation, allowing the network to learn semantic information from the upsampled feature map and still retain information of finer details from the early feature map. At this step, a few more convolutional layers are added and another yolo layer makes new predictions, but at a different scale (proportional to the upsampling factor). Through yet another concatenation with an early feature map and upsampling, this process is repeated a third and final time into the third and last yolo layer, which benefits from computations from the backbone DarkNet network and all previous yolo layer computations. Thus, the YOLO algorithm makes detections at three different scales, where each scale of detections make use of information from prior layers detecting at larger scales, effectively resulting in quick and highly accurate detections of generic objects in images.

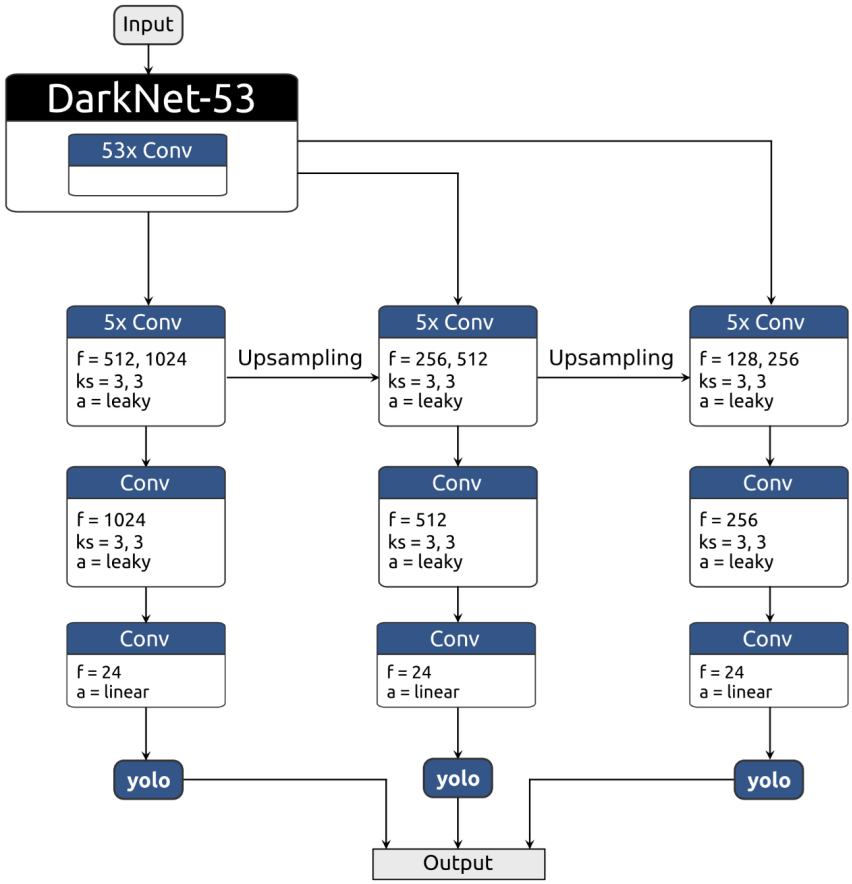

Figure 3.5: *The YOLO algorithm is built upon a DarkNet-53 backbone [145], which consists of a neural network architecture comprised of 53 convolutional layers. This backbone feeds into three necks, which consist of a sequence of 5 convolutional layers, which in turn feed into corresponding heads, which output bounding boxes at three different scales through yolo detection layers [29]. Each neck is up-sampled and fed into its neighbouring neck, to make sure the last yolo detection layer benefits from information collected at every previous stage. Here, Conv represents a single convolutional layer, f is the number of filters in each block, ks is the kernel size, and a is the activation function. The network is visualized with Netron [143]. Figure reproduced with permission from [115].*

The YOLOv3 algorithm was trained on simulated kymographs, generated in the same way as for the U-net (described in section 3.3), but where the input is a perfectly segmented kymograph and the outputs are minimally sized bounding boxes around each separate trajectory. Trajectories are considered separate in the training data if they are more than $T/16$ frames apart, where T is the total amount of frames in the kymograph. Images are continuously generated during training, such that each subsequent image is entirely unique and the risk of overtraining is eliminated. The algorithm was trained using approximately 2 million such unique segmented images, where each image was generated with a constant diffusion in the range $1 \leq D \leq 100$ μm^2/s (the optical contrast is not relevant here because we employ perfectly segmented simulated images) with the standard YOLO error function [145] using the ADAM optimizer [144] with a learning rate of 0.001. The input during training are simulated images (segmented kymographs) of size 128×8192, down-sampled to 128×128 to improve performance, and the output is a list of YOLO-labels containing class, position, probability of occurrence and probability of class of each trajectory in the input image.

3.2.4 Characterizing Trajectories with a Residually Connected Convolutional Neural Network

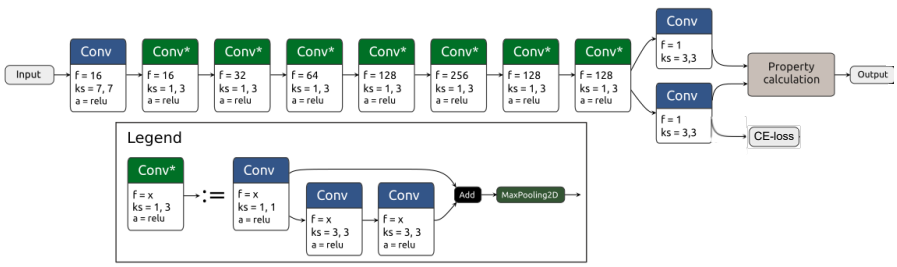

Figure 3.6: *The custom FCNN architecture is comprised of Conv blocks, which are singular convolutional layers, and a sequence of Conv* blocks, which in turn consist of a sequence of convolutional layers followed by a max pooling layer and a skip connection, as exemplified in the legend. This sequence culminates into a custom property layer that calculates iOC and D by the methods described in the text below. Here, f is the number of filters in each block, ks is the kernel size, and a is the activation function. The network is visualized with Netron[143]. Figure reproduced with permission from [115].*

Finally, to calculate intensity *iOC* and diffusivity D for each single biomolecule trajectory, we use a custom neural network architecture, which consists of a FCNN connected with residual blocks. The architecture itself is a sequence of convolutional layers followed by a max pooling layer, including short skip connections to preserve information and mitigate the problems of vanishing gradient as outlined in section 2.2. The full *body* of the architecture consists of seven of these connected blocks, culminating in a computational *head* module,

as exemplified in Figure 3.6. The *body* blocks are the neural networks which bear the brunt of the computation, whereas the *head* controls the nature of the data output by the network. In concrete terms, the *head* of the neural network returns three outputs: (1) The raw output O_{matrix} from the last convolutional layer of the neural network. (2) A mask, which is a down-sampled representation of the original kymograph normalized by its pixel sum and passed through a sigmoid activation function. This effectively gives us a weighted matrix, where each matrix element is the probability that the corresponding down-sampled region in the kymograph contains a biomolecule trajectory. (3) The calculated property for each kymograph, obtained by multiplying O_{matrix} by the mask to produce a value of the property for each pixel in the down-sampled kymograph weighted by the probability of said pixel containing a trajectory, and finally summing over the entire down-sampled image to obtain a single mean value of the property for the entire kymograph. Note here that when calculating D or iOC, the network architecture is identical - the only difference is which property is included in the labelled data and therefore included in the loss which the network's parameters are trained to minimize.

The FCNN networks were trained using the ADAM optimizer [144] with a learning rate of 0.0001 on approximately $300,000$ simulated kymographs in the same range as the U-net and train-validated equivalently to the U-net. The input during training of the FCNNs are simulated images (kymographs) of size $128x2048$ with a single particle trajectory, and the output is a single value of either iOC or D of said trajectory, as well as the mask which as mentioned is a learnt downsampled representation of the original kymograph.

To train the intensity- and diffusivity-calculating FCNN models, a curriculum learning scheme with intermittent checkpoints to be used for later ensemble modelling prediction was employed. Specifically, the intensity-calculating model was initially trained only on a narrow range of high iOC trajectories, representing the highest SNR and in principle easiest case for the model to begin learning correlations, and then slowly curriculum-learned down to the lowest range of relevant iOC values. In each narrow range of iOC values, checkpoint models which are more accurate in that particular narrow range of values are saved separately. Upon model inference, an initial prediction is made with a model trained on the entire range of iOC values with the scheme described above, and then a second model trained on the narrower range of values makes a second prediction on the same trajectory to achieve higher accuracy. The process is equivalent for diffusivity, with the difference being that the range of D values being trained on increases rather than decreases during curriculum learning.

We note here that this ensemble modelling technique adds yet another barrier against false signal detection, since if the epistemic uncertainty inherent in the data is high, which may be the case for very short trajectories or regions of unusually high noise, the two predictions will differ significantly and hence the prediction should be discarded. Of course, as is common in ensemble modelling [150], one can implement an even larger and finer-grained ensemble for even

higher prediction accuracy and accurate uncertainty estimation.

3.2.5 Characterizing Trajectories with a Hierarchical Vision Transformer

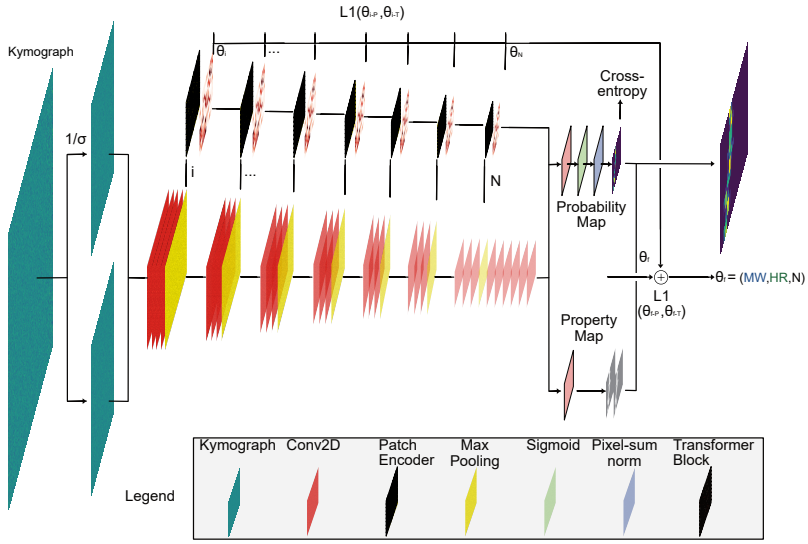

Figure 3.7: Hierarchical Vision Transformer (HViT) Architecture for Single Biomolecule Characterization by NSM. *The HViT model processes kymographs of light scattering intensity to predict key single biomolecule properties, such as MW and (R_s). Initially, Conv2D layers extract spatial features, followed by max pooling for downsampling. The Patch Encoder then encodes these feature maps into smaller patches, while retaining crucial spatial information. Each patch is passed through Transformer Blocks, where multi-head attention mechanisms selectively focus on different regions of the input, capturing long-range dependencies in the kymograph. The outputs of the Transformer Blocks are processed to generate two key outputs: a Probability Map (indicating the likelihood of particle trajectories at different locations) and a Property Map (predicting specific biomolecule properties). The model utilizes cross-entropy loss for classification tasks and L1 loss to minimize the error between predicted and true molecular properties. The estimation of each parameter (θ_i) representing molecular weight, hydrodynamic radius, and number of trajectory points is refined through iterative backpropagation. Finally, the model weighs the Property Map with the Probability Map to output a final estimation of the biomolecule properties, weighted by the probability of said biomolecule existing in any given region.*

To lower the LoD of NSM to the few kDa MW and the few nm R_s regime,

respectively, we develop in Paper VI a new type of deep learning architecture that leverages the Vision Transformer concept [47], and that we term Hierarchical Vision Transformer (HViT). This model builds on recent advancements in attention mechanisms and transformer-based architectures to effectively capture the multi-scale nature of the kymograph data produced by NSM. Specifically, it is designed to estimate the desired properties of molecules - or more generally put "particles" - diffusing inside a nanochannel (such as iOC or MW and D or R_s) without explicitly needing to recreate said particles' pixel-wise trajectories when the particles travel through the nanofluidic channel. Technically speaking this has the distinct conceptual advantage that a global attention mechanism can be leveraged to focus on relevant features of the kymographs, which reduces the impact of local noise and enhances predicted data interpretability. This is an important advance because in data regimes with extremely low SNR at hand here, accurately reconstructing the pixel-wise trajectories of particles becomes practically impossible due to the overwhelming noise that obscures individual particle movements. Specifically, recreating pixel-wise trajectories in noisy conditions often exacerbates signal degradation and increases the computational burden without directly improving property prediction accuracy in the end [151]. To overcome this problem, in HViT we instead reformulate it by using probability maps (rather than pixel-wise trajectories) that encode the likelihood of particle presence at a specific position in the channel at a specific point in time. This allows the model to focus on global feature extraction and particle property prediction by aggregating information from across the entire kymograph to identify consistent patterns indicative of particle presence, even in the presence of noise. By employing multi-head attention mechanisms, the model selectively attends to the most informative regions of the kymograph, dynamically weighting contributions from different parts of the input data to emphasize features that correlate with properties like iOC (and thus MW) and D (and thus R_s). This hierarchical approach enables the HViT to effectively distill the kymograph into a concise representation that captures the essential characteristics of particle diffusion and light scattering, enhancing the model's sensitivity and accuracy for estimating particle properties of interest even when individual trajectories are not clearly visible due to low SNR.

As a further key aspect of the HViT solution, we note that deep learning algorithms generally tend to learn low-frequency modes more quickly and stably than high-frequency modes, a phenomenon known as spectral bias [152]. Since a particle trajectory in a diffraction-limited NSM kymograph essentially constitutes a low frequency signal embedded or hidden in high frequency noise, it is critical to circumvent this spectral bias. In the HViT architecture this is enabled by leveraging its hierarchical nature that forces it to look at different frequencies, i.e., different regions of downsampling, to capture trends in frequencies other deep learning algorithms would ignore. Explicitly, this means that the HViT architecture prioritizes learning the iOC/MW and D/R_s of a particle from a broader context, rather than localized trajectory details, enhancing sensitivity to the very subtle light scattering signals from very small particles. This shift in focus enables the improved detection limits

we demonstrate below, compared to conventional convolutional approaches that often suffer from overfitting to noisy, pixel-wise variations.

The HViT architecture begins with the input processing stage, where the kymograph—a two-dimensional representation of scattered light intensity off a nanochannel over time—is standardized by normalizing it with respect to its temporal standard deviation (Figure 3.7). This pre-processing step ensures that the input data are consistent and thus facilitates the subsequent feature extraction process. The kymograph then undergoes an initial convolution through a layer with a kernel size of 7, utilizing leaky ReLU activation to preserve nuanced variations in the data. This is followed by a series of convolutional blocks, each comprising multiple convolutional layers that progressively increase the number of filters. These blocks serve to downsample the kymograph while simultaneously capturing essential features across different temporal and spatial scales.

A critical innovation of this architecture lies in its hierarchical multi-scale approach. As the model progresses through each convolutional block, the spatial dimensions of the feature maps are reduced via max-pooling operations. At each scale, patches are extracted from the feature maps and encoded using a custom PatchEncoder layer. These encoded patches, now representing condensed and contextually enriched portions of the original kymograph, are fed into a series of transformer blocks. Each transformer block integrates a multi-head attention mechanism, which allows the model to focus selectively on different regions of the patches, and a multi-layer perceptron (MLP), which refines these focused representations. The attention mechanisms within the transformer blocks are pivotal, enabling the model to capture long-range dependencies both in space and time (e.g. that a particle trajectory detected in a region of low noise is likely to continue in a temporally and spatially adjacent region of higher noise), and subtle variations in the kymograph that are critical for accurately estimating properties of interest, such as D and iOC of a particle inside the nanochannel in this present work. At each scale of downsampling, the model generates predictions for a particles' R_s and MW properties through dense layers that process the outputs of the transformer blocks. These predictions are complemented by probability maps, which quantify the likelihood that specific regions of the kymograph contain particle trajectories. These maps are then resized to match the original kymograph dimensions, ensuring that the model's focus aligns spatially with the input data.

The outputs from the different scales of downsampling are subsequently concatenated to form a comprehensive representation of the experimentally obtained kymograph, i.e. that all information contained in the experimental kymograph is condensed into a minimally dimensioned representation, such that all information about the parameters that describe the particle in the channel is pooled together across all scales. This representation undergoes further processing through additional dense layers, culminating in final predictions for the entire kymograph. In this final stage, a probability map and property predictions are generated through a series of convolutional and dense layers. The final

predictions of particle properties MW, R_s as well as the number of frames N are derived using a weighted average, where the weights are set by the probability maps, ensuring that regions with higher probabilities of containing particle trajectories contribute more significantly to the final output.

3.3 Simulations of Biomolecule Trajectories

As mentioned previously, all models are trained using supervised learning, i.e., by training each network to approximate a function correlating input x to label (output) y. To accomplish this, not only do we need to show the networks examples of kymographs containing particle trajectories, but also provide it with reliable ground truth regarding the position and properties of each individual trajectory. This can be done through manual labelling, but we will quickly run into issues regarding the limited datasets which can be procured by manual labour and problems related to human-induced biases. It is also true that such a trained network can never transcend the abilities of the human which provides the ground truth labels. Thus, we turn instead to the possibilities regarding generation of synthetic data, understanding the prospects of which necessitates a deep review of the underlying physics of the NSM method.

Firstly, let us remind ourselves that NSM works by imaging nanofluidic channels within an optically transparent matrix by dark-field light-scattering microscopy, as shown in Figure 3.1. The channels exceed the field of the view lengthwise, and extend between tens to hundreds of nanometres in cross-sectional dimensions depending on the species to be investigated. As a result of this arrangement, the imaged biomolecules are localized within the microscope focal plane during the entire measurement and the optical contrast of the imaged biomolecules is enhanced by several orders of magnitude.

For small biomolecules, such as the proteins and DNA strands imaged in Paper I and VI, their scattering in our system is effectively that of a diffraction-limited spot. To simulate their generated optical response in our system, we thus simulate trajectories with four main varying properties; integrated optical contrast iOC, diffusivity D, gaussian width s and velocity v. At time zero, a position x_0 is randomly chosen along the nanochannel, and the position x_i of the molecule at frame i is generated as Brownian motion with $x_i = x_{i-1} + v\Delta t + \mathcal{N}(0,1)\sqrt{2D\Delta t}$, where a random value $\mathcal{N}(0,1)$ is drawn from a normal distribution with mean 0 and standard deviation 1. The optical response of the biomolecule was then simulated along this path as a gaussian of width s, positions $x_0 ... x_k$ for length of trajectory k and magnitude iOC.

The generated response was then combined with simulated background noise (Figure 3.8B) as $I = I^0 \cdot I^r$, where I^0 is the response of the empty channel and I^r the response of the biomolecule, and kymographs were created according to the procedure described in section 3.2.1. This simulated noise is generated as following. We begin by assuming that the noise in the system is dominated by two sources; shot noise from the channel and imperfections & dirt in the

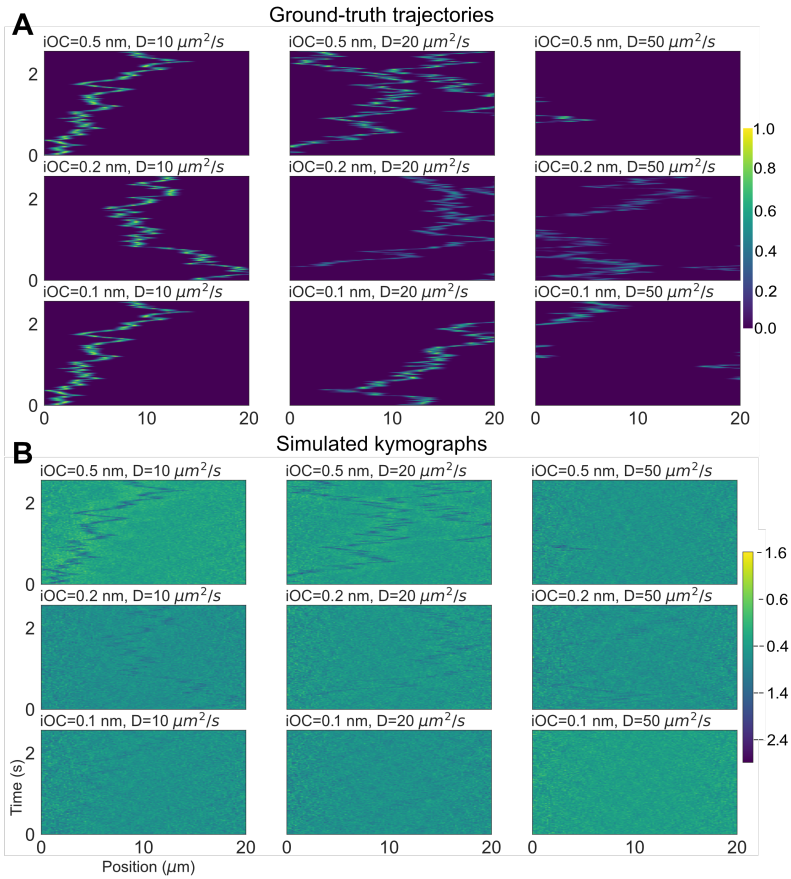

Figure 3.8: ***A:*** *Examples of the simulated response of biomolecules defined by different combinations of optical contrast iOC and diffusivity D.* ***B:*** *Examples of simulated kymographs, acquired by combining the responses in* ***A*** *with simulated noise.*

channel, and the vibration of the channels. The shot noise is generated as an inevitable consequence of quantum-level fluctuations in the number of photons interacting with the sensors of the camera over short periods of time. Shot noise, as a sequence of independently occurring events of photon interaction with a constant mean rate, can be modelled through a Poisson distribution, which in turn can be modelled by a Gaussian distribution in the limit of large N. The second source of noise is the vibrations of the channel (due to movement of the entire setup), modelled as a harmonic vibration in the centroid of the Gaussian distribution. Thus, for each frame t, the background is first initialized as

$$b_0 = e^{(-(x-(x_0+x_\lambda))^2/\lambda^2)} \cdot (1 + \mathcal{N}(0,\,1) * b) \cdot (1 + A\mathcal{N}(0,\,1)) \tag{3.6}$$

where x denotes the position along the channel, x_0, λ, A are numerical parameters randomly drawn from a normal and uniform distribution, respectively, $*$ denotes the convolution operation, and where

$$x_\lambda = (2\mathcal{N}(0,\,1) + sin((\pi - 0.05)t)) \cdot d_x,$$
$$b = \frac{e^{(-(x)^2/C^2)} + x_\lambda}{\sum(e^{(-(x)^2/C^2)} + x_\lambda)},$$

where d_x, C are numerical parameters randomly drawn from uniform distributions. Here, in equation 3.6, the first factor is the aforementioned (partially sinusoidally variant and partially randomly sampled) Gaussian. The second factor corresponds to noise resultant of a perfect channel (1) plus noise resultant of dirt and imperfections in the channel ($\mathcal{N}(0,\,1) * b$), which can be modelled by Gaussian noise with a large correlation length. We achieve this by a convolution with a generic (randomly sampled) point spread function (PSF) of the microscope and the (randomly sampled) position and size of said dirt. The third factor corresponds to the aforementioned shot noise, and is effectively a randomly sampled scalar.

Finally, the total background noise b_f is calculated as

$$b_f = b_0 \cdot (1 + b_A\mathcal{N}(0,\,1)) + 0.4b_A\mathcal{N}(0,\,1) \tag{3.7}$$

where a new normal variable is drawn for each position x along the channel, and b_A is a numerical parameter drawn from a uniform distribution correlated with the total quantitative noise level in the system.

The value of the numerical parameters are chosen such that they roughly correspond to the quantitative levels of noise seen in experimentally measured noise, within a numerical span large enough such that any quantitative variations of the noise are still appropriately accounted for. In principle, so long as the true values of the experimentally measured noise are encapsulated within the range of the simulated noise, we can expect the neural network to properly interpolate and correct for any such variation of noise. However, note that this approach is sensitive to sources of noise which are not included in the simulations. Specifically, if a qualitatively different source of noise excluded in the simulations is

present in a measurement at inference time (i.e. spurious vibrations as a result of human movement, or device malfunction), it will be ambiguous to the neural network whether this portion of the data should be considered signal or noise. This problem can, to an extent, be mitigated by transfer-learning the trained network on a few examples of simulated trajectories within experimentally measured noise before each new set of measurements, effectively constituting a re-calibration step of the network. Alternatively, it is always possible to include new sources of noise into the simulations. However, note that this increases the state-space of possibilities which the network needs to learn and thereby increases the overall training time and increases the requirements for representational ability within the choice of network architecture.

Regardless, this method demonstrates a powerful benefit of using deep learning systems: we do not have to perfectly describe or understand the noise in order to correct for it. In general, so long as the noise is consistent (i.e., no significantly different qualitatively new sources of noise are expected to exist at inference time), and we know the generic representation of the dominating sources of noise, we can expect deep learning systems to properly interpolate learnt correlations to experimentally relevant data.

3.3.1 Simulated Biomolecules in Measured Noise

To estimate the accuracy and precision of the algorithm, we test it on a combination of simulated biomolecules of $iOC = [0.1, 0.2, 0.5, 1, 2]$ nm and $D = [10, 20, 50]$ μm^2/s on top of experimentally measured noise. The trajectories of biomolecules were found using the segmentation & object detection algorithm described in sections 3.2.2 and 3.2.3, from which iOC and D were calculated using the FCNN described in 3.2.4. Such determined values of iOC and D of all detected trajectories are shown as a scatter plot in Figure 3.9. The mean values of iOC and D in Figure 3.9 correspond to the mean calculated value of iOC and D for each separate simulated trajectory. Thus, for a perfect algorithm, we would expect the circular points (mean calculated value) to perfectly overlap the crosses (ground truth), with error bars of minimal length (see [115] for theoretical limits in precision). The values are compared with that found by a standard heuristic-based algorithm (SA) in Figure 3.9B. Clearly, the values identified by the pipeline correspond to the ground truth very well even within experimentally measured conditions, and particularly outperform the standard analysis at the lowest iOC, thereby corroborating the used data analysis pipeline as a whole.

3.4 Theoretical Limits

In this section, we explore the theoretical limits of the NSM system. In the first section, I derive the fundamental limits of estimating iOC/MW and D/R_s via the Cramér-Rao Lower Bound (CRLB), and in the second section I compare these calculated limits to the output of the HViT in a simulated system in ideal conditions.

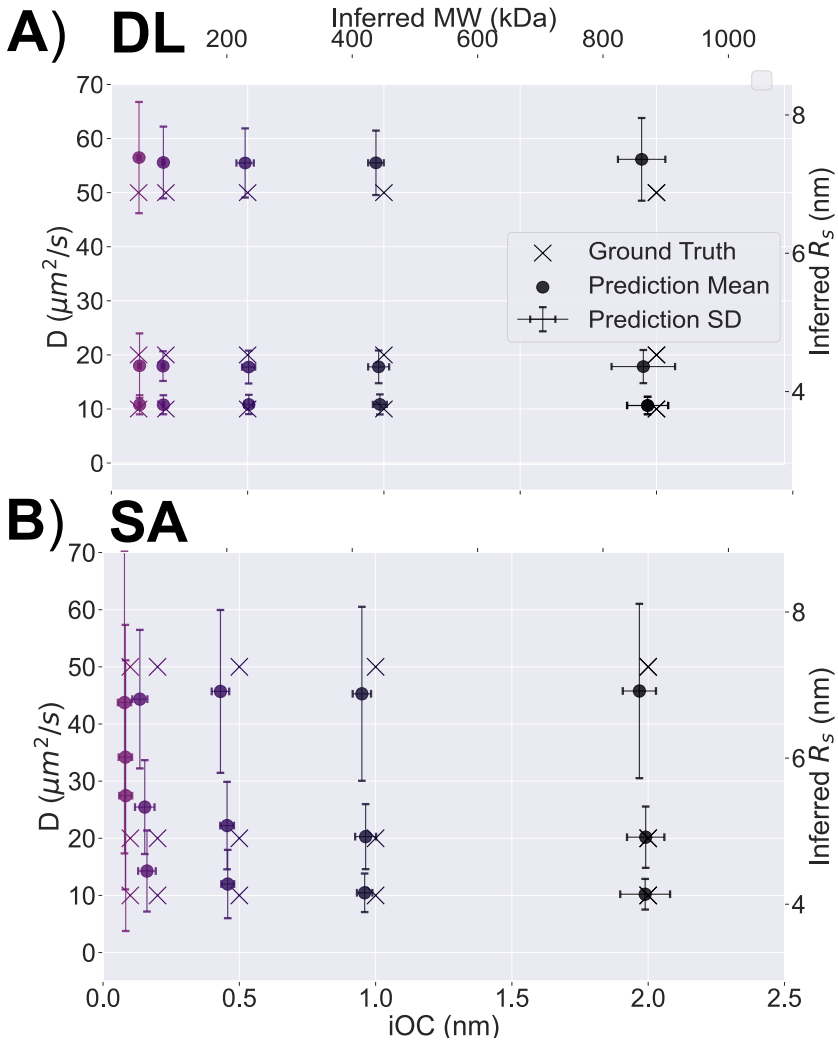

Figure 3.9: *Analysis of the artificially generated data set comprised of experimentally recorded background signal and simulated response of biomolecules defined by a combination of values for* $iOC = [0.1, 0.2, 0.5, 1, 2]$ *nm, and* $D = [10, 20, 50]\mu\mathrm{m}^2/\mathrm{s}$. ***A:*** *Scatter plot of mean value of iOC and D. Error bars correspond to the resolution in iOC and D defined as the standard deviation of the calculated values for iOC and D for each separate* D, iOC *permutation.* ***B****: Same as for DL, but for the standard analysis algorithm (SA) as described in [115].*

3.4.1 Cramér-Rao Lower Bound

The CRLB provides a theoretical lower bound on the variance of any unbiased estimator of a parameter, indicating the best precision achievable under a given statistical model [153], [154]. It is a fundamental concept in estimation theory and is widely used in statistical signal processing [155]. The CRLB is directly tied to the Fisher information, denoted as $I(\theta)$, which measures the amount of information that an observable random variable conveys about an unknown parameter θ. The mathematical expression for the CRLB is given by:

$$\text{Var}(\hat{\theta}) \geq \frac{1}{I(\theta)}$$

where $\text{Var}(\hat{\theta})$ represents the variance of the estimator $\hat{\theta}$. This inequality suggests that the variance of any unbiased estimator cannot be smaller than the inverse of the Fisher information.

It serves as a benchmark for assessing the efficiency of estimators. An estimator that achieves this lower bound is considered efficient, as it has the smallest possible variance among all unbiased estimators for that parameter. The amount of data influences the Fisher information—generally, an increase in data leads to higher Fisher information, which implies a tighter bound and hence, a reduced potential estimation error.

Define the Statistical Model

We begin by specifying the probability density function (pdf) or probability mass function (pmf) of our data, including the scale parameter σ that we aim to estimate. The choice of model depends on the nature of our data (e.g., normal distribution, exponential distribution).

Example: Consider a normal distribution with mean zero and unknown scale (standard deviation) σ:

$$f(x;\sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x^2}{2\sigma^2}\right)$$

This model is appropriate for many natural phenomena due to the Central Limit Theorem [156].

Find the Likelihood Function

For a sample of independent and identically distributed observations $X = (X_1, X_2, \ldots, X_n)$, we construct the likelihood function by taking the product of individual pdfs:

$$L(\sigma; X) = \prod_{i=1}^{n} f(x_i;\sigma)$$

Example:

$$L(\sigma; X) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left(-\frac{\sum_{i=1}^n x_i^2}{2\sigma^2}\right)$$

Compute the Log-Likelihood Function

Next, we take the natural logarithm of the likelihood function to simplify differentiation:

$$\ell(\sigma; X) = \ln L(\sigma; X)$$

Example:

$$\ell(\sigma; X) = -n \ln \sigma - \frac{\sum_{i=1}^n x_i^2}{2\sigma^2} + \text{constant terms}$$

Calculate the First Derivative (Score Function)

We differentiate the log-likelihood function with respect to σ to obtain the score function:

$$\frac{\partial \ell(\sigma; X)}{\partial \sigma} = -\frac{n}{\sigma} + \frac{\sum_{i=1}^n x_i^2}{\sigma^3}$$

Compute the Second Derivative

We now differentiate the score function with respect to σ to get the observed information:

$$\frac{\partial^2 \ell(\sigma; X)}{\partial \sigma^2} = \frac{n}{\sigma^2} - \frac{3\sum_{i=1}^n x_i^2}{\sigma^4}$$

Calculate the Fisher Information $I(\sigma)$

The Fisher Information quantifies the amount of information that our observable data carries about the unknown parameter σ:

$$I(\sigma) = -\mathrm{E}\left[\frac{\partial^2 \ell(\sigma; X)}{\partial \sigma^2}\right]$$

Given that $X_i \sim N(0, \sigma^2)$ and $\mathrm{E}[X_i^2] = \sigma^2$, we have:

$$I(\sigma) = -\left(\frac{n}{\sigma^2} - \frac{3n\sigma^2}{\sigma^4}\right) = \frac{2n}{\sigma^2}$$

This result is consistent with standard statistical texts [157].

Compute the Cramér-Rao Lower Bound

Finally, the CRLB provides the minimum variance bound for any unbiased estimator $\hat{\sigma}$ of σ:

$$\mathrm{Var}(\hat{\sigma}) \geq \frac{1}{I(\sigma)} = \frac{\sigma^2}{2n}$$

This implies that no unbiased estimator of σ can have a variance lower than $\frac{\sigma^2}{2n}$, as established in estimation theory [155].

3.4.2 Cramér-Rao Lower Bound for NSM

In NSM, the intensity of scattered light from biomolecules diffusing in a nanofluidic channel provides insights into molecular weight and hydrodynamic radius. For NSM, the observed intensity I_t of the scattered light from a biomolecule within a nanochannel can be modeled as:

$$I_t = cI_0L|\alpha_t|^2\frac{k^3}{4}$$

where:

- $\alpha_t = \alpha_c + \dfrac{\alpha_m}{L^2}$ represents the total polarizability, composed of the polarizability of the nanochannel α_c and the biomolecule α_m.
- α_m, α_c: Polarizabilities of the biomolecule and the nanochannel.
- I_0: Incident light intensity.
- k: Wavenumber of the light.
- L: Length of the illuminated part of the nanochannel.
- c: Collection efficiency.

Assumptions for the Likelihood Function

Assume the measured intensity I_t is subject to Gaussian noise. Thus, the probability density function of observing I_t given the parameters α_m is modeled as:

$$f(I_t;\alpha_m) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(I_t-\mu(\alpha_m))^2}{2\sigma^2}\right)$$

where:

$$\mu(\alpha_m) = cI_0L\left(\alpha_c + \frac{\alpha_m}{L^2}\right)^2\frac{k^3}{4}$$

and σ is the standard deviation of the measurement noise.

The log-likelihood function $\log f(I_t;\alpha_m)$ is:

$$\log f(I_t;\alpha_m) = -\frac{1}{2}\log(2\pi\sigma^2) - \frac{(I_t-\mu(\alpha_m))^2}{2\sigma^2}$$

Derivation of the Fisher Information

The Fisher Information $I(\alpha_m)$ is defined as the negative expected value of the second derivative of the log-likelihood with respect to the parameter α_m. In the Gaussian noise model with known variance, the Fisher Information simplifies to:

$$I(\alpha_m) = \frac{1}{\sigma^2} \left(\frac{\partial \mu(\alpha_m)}{\partial \alpha_m} \right)^2$$

Compute the first derivative of $\mu(\alpha_m)$ with respect to α_m:

$$\frac{\partial \mu(\alpha_m)}{\partial \alpha_m} = cI_0 L \cdot 2 \left(\alpha_c + \frac{\alpha_m}{L^2} \right) \cdot \left(\frac{1}{L^2} \right) \cdot \frac{k^3}{4}$$

Simplify:

$$\frac{\partial \mu(\alpha_m)}{\partial \alpha_m} = \frac{cI_0 k^3}{2L} \left(\alpha_c + \frac{\alpha_m}{L^2} \right)$$

Note that $\alpha_t = \alpha_c + \dfrac{\alpha_m}{L^2}$, so:

$$\frac{\partial \mu(\alpha_m)}{\partial \alpha_m} = \frac{cI_0 k^3}{2L} \alpha_t$$

Therefore, the Fisher Information is:

$$I(\alpha_m) = \frac{1}{\sigma^2} \left(\frac{cI_0 k^3}{2L} \alpha_t \right)^2$$

Cramér-Rao Lower Bound (CRLB)

The CRLB provides a lower bound on the variance of any unbiased estimator of α_m:

$$\mathrm{Var}(\hat{\alpha}_m) \geq \frac{1}{I(\alpha_m)} = \frac{\sigma^2}{\left(\dfrac{cI_0 k^3}{2L} \alpha_t \right)^2}$$

Simplify:

$$\mathrm{Var}(\hat{\alpha}_m) \geq \sigma^2 \left(\frac{2L}{cI_0 k^3 \alpha_t} \right)^2$$

This expression shows that the variance of the estimator depends on the true value of α_t, which includes the parameter α_m we aim to estimate.

Model for Multiple Measurements

For N independent measurements, the total Fisher Information is:

$$I_{\text{total}}(\alpha_m) = N \times I(\alpha_m)$$

Thus, the CRLB for estimating α_m from N measurements is:

$$\mathrm{Var}(\hat{\alpha}_m) \geq \frac{1}{I_{\text{total}}(\alpha_m)} = \frac{\sigma^2}{N} \left(\frac{2L}{cI_0 k^3 \alpha_t} \right)^2$$

CRLB for Diffusivity in NSM

The relationship between the diffusivity D and the hydrodynamic radius R_s is given by the Stokes-Einstein equation:

$$D = \frac{k_B T}{6\pi\eta R_s}$$

Considering the displacement x of a diffusing particle over time t, the probability density function is:

$$f(x; D) = \frac{1}{\sqrt{4\pi D t}} \exp\left(-\frac{x^2}{4Dt}\right)$$

The log-likelihood function is:

$$\log f(x; D) = -\frac{1}{2}\log(4\pi D t) - \frac{x^2}{4Dt}$$

Compute the first derivative with respect to D:

$$\frac{\partial}{\partial D}\log f = -\frac{1}{2D} + \frac{x^2}{4D^2 t}$$

Compute the second derivative:

$$\frac{\partial^2}{\partial D^2}\log f = \frac{1}{2D^2} - \frac{x^2}{2D^3 t}$$

Calculate the Fisher Information by taking the negative expected value of the second derivative. Since $E[x^2] = 2Dt$, we have:

$$I(D) = -E\left[\frac{\partial^2}{\partial D^2}\log f\right] = \frac{1}{2D^2}$$

Thus, the CRLB for a single measurement is:

$$\mathrm{Var}(\hat{D}) \geq \frac{1}{I(D)} = 2D^2$$

For N independent measurements:

$$I_{\text{total}}(D) = N \times I(D) = \frac{N}{2D^2}$$

So the CRLB becomes:

$$\mathrm{Var}(\hat{D}) \geq \frac{1}{I_{\text{total}}(D)} = \frac{2D^2}{N}$$

Final CRLB Expressions

$$\boxed{\mathrm{Var}(\hat{\alpha}_m) \geq \frac{\sigma^2}{N}\left(\frac{2L}{cI_0 k^3 \alpha_t}\right)^2}$$

$$\boxed{\mathrm{Var}(\hat{D}) \geq \frac{2D^2}{N}}$$

3.4.3 Including Localization Error in CRLB

So far, we have assumed a theoretically optimal but practically impossible localization error of 0 for a molecule in an NSM experiment. If we include the theoretical limits of localization error, for N independent displacement measurements over a total observation time $T = n\Delta t$, the CRLB is derived as follows.

Let σ_L be the standard deviation of the localization error. The measured position $x_m(t)$ at time t is:

$$x_m(t) = x(t) + \epsilon$$

where $x(t)$ is the true position, and ϵ is a zero-mean Gaussian random variable with variance σ_L^2.

The observed displacement Δx_m between two time points separated by Δt is:

$$\Delta x_m = x_m(t + \Delta t) - x_m(t) = \Delta x + \epsilon_2 - \epsilon_1$$

where $\Delta x = x(t + \Delta t) - x(t)$ is the true displacement due to diffusion.

The variance of the observed displacement Δx_m is:

$$\mathrm{Var}(\Delta x_m) = 2D\Delta t + 2\sigma_L^2$$

Likelihood Function and Fisher Information

Assuming that the observed displacements Δx_m are independent and normally distributed, the likelihood function for observing a displacement Δx_m given diffusivity D is:

$$f(\Delta x_m; D) = \frac{1}{\sqrt{4\pi(D\Delta t + \sigma_L^2)}} \exp\left(-\frac{(\Delta x_m)^2}{4(D\Delta t + \sigma_L^2)}\right)$$

The log-likelihood function is:

$$\log f(\Delta x_m; D) = -\frac{1}{2}\log\left(4\pi(D\Delta t + \sigma_L^2)\right) - \frac{(\Delta x_m)^2}{4(D\Delta t + \sigma_L^2)}$$

The first derivative with respect to D is:

$$\frac{\partial}{\partial D}\log f = -\frac{\Delta t}{2(D\Delta t + \sigma_L^2)} + \frac{(\Delta x_m)^2 \Delta t}{4(D\Delta t + \sigma_L^2)^2}$$

The second derivative is:

$$\frac{\partial^2}{\partial D^2}\log f = \frac{\Delta t^2}{2(D\Delta t + \sigma_L^2)^2} - \frac{(\Delta x_m)^2 \Delta t^2}{2(D\Delta t + \sigma_L^2)^3}$$

The Fisher information $I(D)$ from a single displacement measurement is the negative expected value of the second derivative:

$$I(D) = -E\left[\frac{\partial^2}{\partial D^2}\log f\right]$$

Since $E\left[(\Delta x_m)^2\right] = 2(D\Delta t + \sigma_L^2)$, we have:

$$I(D) = -\left(\frac{\Delta t^2}{2(D\Delta t + \sigma_L^2)^2} - \frac{2(D\Delta t + \sigma_L^2)\Delta t^2}{2(D\Delta t + \sigma_L^2)^3}\right) = \frac{\Delta t^2}{2(D\Delta t + \sigma_L^2)^2}$$

For n independent measurements:

$$I_{\text{total}}(D) = n \times I(D) = \frac{n\Delta t^2}{2(D\Delta t + \sigma_L^2)^2}$$

Cramér-Rao Lower Bound

The CRLB for diffusivity D, considering the localization error, is:

$$\text{Var}(\hat{D}) \geq \frac{1}{I_{\text{total}}(D)} = \frac{2(D\Delta t + \sigma_L^2)^2}{n\Delta t^2}$$

Simplifying:

$$\text{Var}(\hat{D}) \geq \frac{2}{n}\left(\frac{D\Delta t + \sigma_L^2}{\Delta t}\right)^2$$

The localization precision σ_L is related to the measurement noise σ and the experimental parameters:

$$\sigma_L = \frac{\sigma}{\sqrt{cI_0L|\alpha_t|^2\frac{k^3}{4}}}$$

Substituting σ_L into the CRLB expression:

$$\sigma_L^2 = \left(\frac{\sigma}{\sqrt{cI_0L|\alpha_t|^2\frac{k^3}{4}}}\right)^2 = \frac{4\sigma^2}{cI_0L|\alpha_t|^2k^3}$$

Thus, the CRLB becomes:

$$\text{Var}(\hat{D}) \geq \frac{2}{n}\left(\frac{D\Delta t + \frac{4\sigma^2}{cI_0L|\alpha_t|^2k^3}}{\Delta t}\right)^2$$

Simplify the numerator inside the parentheses:

$$D\Delta t + \frac{4\sigma^2}{cI_0L|\alpha_t|^2k^3} = \Delta t\left(D + \frac{4\sigma^2}{cI_0L|\alpha_t|^2k^3\Delta t}\right)$$

Therefore, the CRLB simplifies to:

$$\text{Var}(\hat{D}) \geq \frac{2}{n}\left(D + \frac{4\sigma^2}{cI_0L|\alpha_t|^2k^3\Delta t}\right)^2$$

Final CRLB Expression for Diffusivity in NSM

The modified CRLB for diffusivity D in NSM, incorporating the localization precision, is:

$$\boxed{\mathrm{Var}(\hat{D}) \geq \frac{2}{n}\left(D + \frac{4\sigma^2}{cI_0L|\alpha_t|^2k^3\Delta t}\right)^2}$$

The term $\dfrac{4\sigma^2}{cI_0L|\alpha_t|^2k^3\Delta t}$ represents the influence of localization error on the variance of $\hat{D}$. Reducing measurement noise σ or increasing the detected signal improves localization precision, thus reducing this term. Increasing Δt reduces the impact of localization error relative to diffusion, enhancing estimation precision. The variance decreases inversely with the number of measurements, here the number of acquired frames with a molecule inside the nanochannel, emphasizing the benefit of collecting more data.

In the absence of localization error ($\sigma_L^2 = 0$), the CRLB reduces to:

$$\mathrm{Var}(\hat{D}) \geq \frac{2D^2}{n}$$

which matches the standard result derived without considering localization error.

3.4.4 Comparing CRLB to Ideal Model Precision

To illustrate the predictive capabilities of the HViT model in terms of MW and R_s, we trained it on simulated kymographs and corresponding probability maps of particles with MW in the range 1 - 30 kDa and R_s in the range 1 - 2.7 nm. We furthermore categorise the obtained data into three groups according the simulated trajectory length, N, according to $N = 100$, $N = 2000$ and $N = 10000$ measurements, respectively. These three categories serve the purpose of investigating the significant impact of N on the models' particle property prediction accuracy, as it mediates the amount of data available to the model, which is critical for reducing noise. Corresponding scatter plots of the predicted versus true MW and R_s for the three different trajectory length categories reveal significantly improving accuracy and precision for increasing trajectory lengths with $N = 10000$ delivering almost perfect agreement between true and predicted MW (Figure 3.10A-C), as well as true and predicted R_s (Figure 3.10 D-F). This alignment indicates that by simplifying the learning problem through the use of probability maps, the HViT model effectively utilizes the increasing amount of available data to produce more reliable estimates of molecular properties, remarkably even for simulated particles with MW as low as 5 kDa in a nanochannel with cross-sectional diameter $A_V = 30 \times 63$ nm^2.

To further analyze and fundamentally understand these results, it is instructive to plot the relative mean error ("accuracy") and relative standard deviation ("precision") of the predicted MW and R_s values as functions of the number of

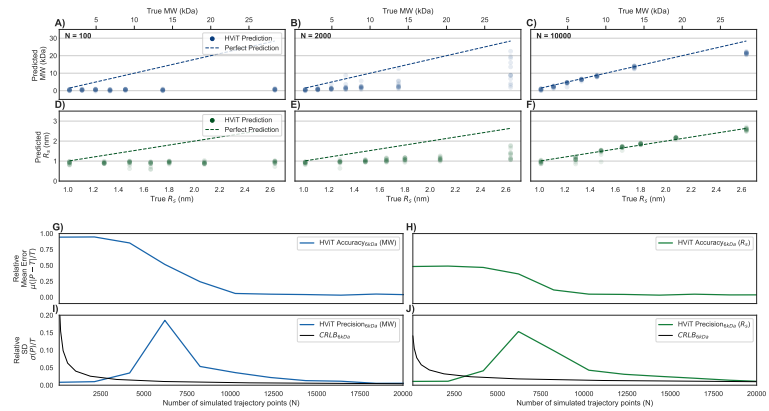

Figure 3.10: *(A,B,C): Scatter plots of predicted MW vs. true MW, based on kymographs with increasing trajectory lengths (N = [100, 2000, 10000]). Each dot represents a simulated biomolecule, and the blue dotted line represents the ideal prediction line (perfect MW match). (D,E,F): Scatter plots of predicted hydrodynamic radius (HR) vs. true HR for the same trajectory lengths as above, with each green dot corresponding to a simulated molecule. (G,H,I,J): Relative mean error (μ) and relative standard deviation (σ) of predicted MW and HR, showing the accuracy (top row) and precision (bottom row) as a function of the number of measurements. Blue and green lines represent the ideal model accuracy and precision for MW and HR, respectively, while black lines indicate the Cramér-Rao Lower Bound (CRLB) for the 6 kDa molecule, illustrating the theoretical limit of prediction.*

measurements N (Figure 3.10G-J). We chose to specifically do so for a particle with $MW = 6$ kDa to emulate our second experiment discussed below using the peptide hormone Insulin with $MW = 5.8$ kDa. Clearly, the accuracy in both MW (Figure 3.10G) and R_s (Figure 3.10H) predictions increases dramatically for increasing N and approaches relative mean error values of 3.8% (MW) and 3.7% (R_s), respectively, for $N > 10000$. Interestingly, for the precision of the predictions of MW (Figure 3.10I) and R_s Figure 3.10J), we find very low values of the relative standard deviation for short trajectories ($N < 2500$), a maximum at $N \approx 6200$, and asymptotically decreasing values for $N > 6200$.

This at first somewhat surprising result can be understood as a consequence of the model being a biased estimator which tends to underfit the data at intermediate trajectory lengths. This bias occurs because the model is designed to interpret kymographs with small N as highly likely to be noise, resulting in a systematic tendency to predict MWs close to zero and R_s close to 1 nm, in these cases, which for small true MW and R_s values indeed is a quite "good" estimate. As N increases, the network has more data at its disposal, allowing it to increasingly distinguish between actual signal due to the presence of a particle and noise. Nevertheless, initially in this regime, the precision of the model prediction decreases because the network is no longer dismissing as much data as noise. Hence, it starts to process kymographs that, for smaller N, would have been classified as noise and thus assigned MWs close to zero or R_s values close to 1 nm. Consequently, this introduces transiently larger variance into the predictions, which is reflected in the observed transiently reduced precision (Figure 3.10 I, J). However, with even further increasing N, the model continues to refine its predictions, as the additional measurements provide clearer molecular signals, and both accuracy and precision improve. By the time the trajectory length reaches $N = 10000$, the HViT model is thus able to accurately and consistently predict MW and R_s of the 6 kDa particle, thereby effectively balancing the trade-off between filtering out noise and accurately capturing signals in the kymograph that stem from the presence of the particle.

To put the model prediction precision as function of N of both MW and R_s for the 6 kDa particle into perspective, we have calculated the corresponding theoretical limits set by the Cramér-Rao Lower Bound (CRLB) of an NSM measurement. It becomes clear that for large N, the model performs near the theoretical limit set by the CRLB (Figure 3.10I,J). It also shows how the model apparently makes predictions with higher precision than the CRLB in the regime where N is small, due to the aforementioned systematic tendency to predict MWs close to zero and R_s close to 1 nm when the signal becomes indistinguishable from noise. Taken all together, the CRLB analysis illustrates how the HViT model dynamically adjusts its performance based on the amount of data available (expressed as trajectory length, N, which in the experiments translates to the number of frames included in the measured kymograph), and how it thereby eventually overcomes initial biases that enables it to make predictions that align closely with the true particle properties, even for a particle as small as 6 kDa. This in turn corroborates that the probabilistic HViT approach

not only improves the particle detection capability in low SNR conditions but also pushes both the accuracy and precision of the made predictions closer to the fundamental limits of the NSM technique. Simultaneously it highlights the importance of significant trajectory lengths when operating in this regime. Experimentally, this constitutes a challenge since the intrinsically higher D of small particles or molecules reduces the time they spent in the field of view of the microscope, which at a given frame rate of the imaging camera results in smaller N for smaller particles. Fortunately, however, if necessary this can be compensated by, e.g., statistical bootstrapping [158], simply by chance given long measurements, or potentially by electrostatic entrapment of molecules [159].

3.5 HViT- enabled NSM of 5.8 kDA Insulin

To further push the LoD of NSM by HViT data analysis, we nanofabricated a fluidic chip with the to-date smallest nanochannels demonstrated for NSM, which feature cross-sectional dimensions $A_V = 30 \times 63$ nm^2, to intrinsically boost the iOC of small analyte molecules. We then applied this system to investigate the peptide hormone Insulin in PBS buffer. Insulin has a nominal MW = 5.8 kDa and $R_s \approx 1.5$ nm [160]. As expected for this ultrasmall MW regime, the scattering intensity kymographs measured for both Insulin in PBS-buffer and pure PBS-buffer control do not resolve any trajectories or other significant features due to very small iOC. Hence, they look essentially identical Figure (3.11A,D). Inputting the kymographs to the HViT model produces again probability maps that highlight regions where the network has detected Insulin with high confidence (Figure 3.11B). Corresponding probability maps obtained from kymographs measured from channels only filled with PBS buffer exhibit only very low probabilities, confirming the absence of Insulin molecules in these control samples (Figure 3.11E). Subsequently, based on these probability maps, the HViT model predicts the MW and R_s values for each measured kymograph for the Insulin and PBS-control samples, first without applying any probability threshold (Figure 3.11C). Clearly, identified molecules in the Insulin sample cluster around a common median value of MW = 6.4 +/- 13.1 kDa and R_s= 1.55 +/- 1.1 nm, which is similar to the nominal MW = 5.8 kDa and R_s = 1.5 nm.

Nevertheless, we also notice a significant spread in the individual predicted values, as indicted by the error bars in Figure 3.11C. This spread is the consequence of variability in the scattering signal intensity due to the inherent noise in the low SNR conditions at hand. To this end, we also note that the level of transparency of the individual data points in the scatter plot reflects N for each corresponding kymograph, with more opaque points corresponding to larger N values and, consequently, more reliable MW and R_s predictions, as discussed above. For the pure PBS-buffer control measurements the model predicts two clusters of particles, one at lower MW and R_s than the Insulin sample and one at significantly higher values. This is the consequence of residual noise in the system being misinterpreted as ultra low-molecular-weight

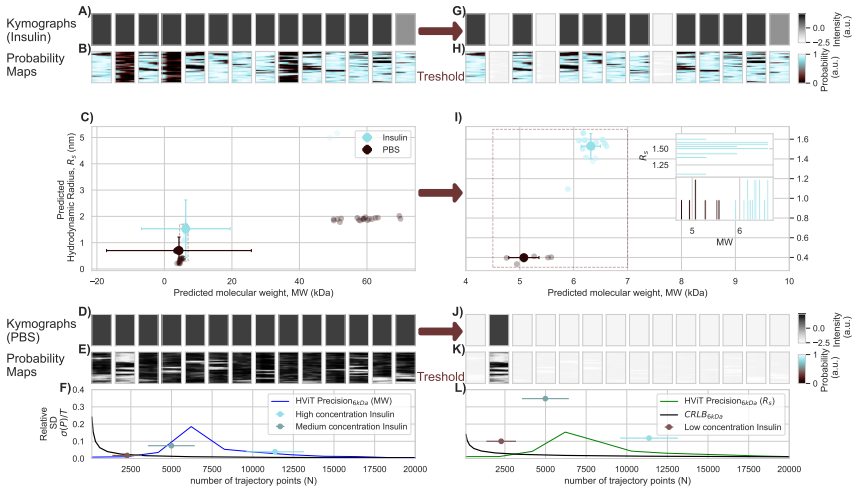

Figure 3.11: Insulin Analysis in Nanofluidic Channel (30x63 nm). *The experimental results of analyzing Insulin at three different concentrations, alongside the control PBS-buffer, are shown. (A) Kymographs for Insulin measurements, demonstrating the interference scattering signal captured by NSM and representing the input to the network. (B) Corresponding probability maps, indicating regions of Insulin detection across the different concentrations by the network. (C) Predicted hydrodynamic radius (HR) versus molecular weight (MW) for Insulin (turquoise) and PBS (black) measurements, where each data point corresponds to a kymograph and the error bar corresponds to the standard deviation of MW (HR) across all measured kymographs. The transparency of each data point is correlated with the estimated number of measurements within the respective kymograph. (F-J) Same as (A-E), but after tresholding low-likelihood kymographs. (K,L) The theoretical limits for MW and HR prediction precision alongside the experimental Insulin measurements. The Insulin measurements at different concentrations are superimposed on the theoretical results, at x-values corresponding to the estimated estimated number of measurements N in each kymograph.*

particles in the lower cluster, and as larger objects in the higher cluster when essentially no real detectable objects are in the system. Notably, in particular the data points in the higher cluster are also very faint, which means they were extracted from kymographs with small N and thus assigns them an inherent low reliability.

As the next step, we applied a threshold to the probability maps to be included in subsequent MW and R_s predictions also in this case (Figure 3.11G-K). As expected, this thresholding step significantly improves the precision and accuracy, as evidenced by only probability maps with significantly concentrated high probabilities being included in the analysis (Figure 3.11H) and consequently significantly narrowed predicted MW and R_s distributions (Figure 3.11I). This is the consequence of the increased SNR in the selected probability maps, which effectively filters out low-confidence detection of molecules, and renders the identification of true Insulin-related signals more consistent, and thus the MW and R_s predictions more accurate and precise. This is reflected by a much tighter distribution of the points in the scatter plot (Figure 3.11I).

As a second aspect, we also see that the thresholding effectively eliminates the low confidence points from the PBS control data (Figure 3.11J, K) and thereby also enhances the distinction between Insulin and PBS measurements as it generates a clearer separation between the two data point clusters (Figure 3.11I). As the final aspect we discuss the absolute values of predicted median Insulin MW and R_s values after thresholding, i.e. $MW = 6.3$ kDa +/- 0.18 kDa and $R_s = 1.53$ nm +/- 0.13 nm (Figure 3.11I), which slightly deviates from the nominal value of 5.8 kDa. As the main reason for this slight discrepancy we attribute to deviations in the actual nanochannel dimensions from the nominal value, which affect the conversion from iOC to MW. Similarly, the mean predicted R_s is 1.53 nm, which slightly deviates from the nominal value of ca. 1.5 nm.

To further explore the limits of the HViT model in characterizing biomolecules at low concentrations, we extended our experiments to include Insulin samples with lower concentrations, emulating a scenario with highly varying total numbers of measurements, N, across different experimental conditions. In these experiments, we systematically varied the Insulin concentrations to create data sets with a wide range of N values, where N represents the total number of measurements summed over all kymographs collected for each condition. This approach allows us to assess the model's performance in situations where the SNR can vary significantly due to the reduced number of molecules present in the nanochannel.

We used these data sets to generate the data points displayed in Figure 3.11F and Figure 3.11L, which depict the relative standard deviation (precision) of the MW and the relative standard deviation of R_s as functions of the number of measurements, N. For each concentration, we calculated the precision by comparing the spread in the predicted values of MW and R_s against their respective means, normalized by the total number of measurements in the kymographs. This analysis enabled us to quantify how the variability in the

data decreases as the number of measurements increases, reflecting the model's improved precision with larger sample sizes.

To determine the ideal model precision for MW and R_s, we employed theoretical limits derived from the CRLB. These limits represent the minimum possible variance in the parameter estimates given the noise characteristics of the system, setting a benchmark for the best achievable prediction precision. We computed the CRLB for each scenario based on the known properties of the nanochannel and the Insulin molecule, taking into account factors such as the scattering signal intensity, the background noise level, and the temporal resolution of the measurements. The ideal model precision curves shown in Figure 3.11F and Figure 3.11L were generated by applying these CRLB-derived limits to the experimental conditions, providing a reference against which to compare the experimental Insulin data.

In panels Figure 3.11F and Figure 3.11L, the Insulin measurements at different concentrations are superimposed on the theoretical precision limits for molecular weight and hydrodynamic radius predictions. These limits are derived from the CRLB, representing the best achievable prediction precision given the noise in the system. Each Insulin concentration is positioned on the x-axis according to the estimated number of measurements in the corresponding kymographs. Remarkably, the experimental Insulin data closely follow the ideal model's performance, approaching the theoretical limit for both MW and HR prediction precision. This strong alignment between experimental and theoretical results reinforces the robustness of the NSM technique, combined with deep learning, in characterizing molecules in the sub-10 kDa regime if the number of measurements are high enough.

The results presented in Figure 3.11 showcase the capacity of this advanced NSM method to detect and accurately measure Insulin molecules in real time, in a label-free manner. The close correspondence between the experimental data and theoretical performance underscores the efficacy of this approach in pushing the boundaries of biomolecule detection and characterization in nanofluidic systems.

Chapter 4

DL for Nanoplasmonic Hydrogen Sensing

In this chapter, we go through the fundamentals of localized surface plasmon resonance, as well as how we employed transformers to enable and accelerate hydrogen gas sensing in challenging sensing environments. Note that the phenomenon of (nano)plasmonic resonance is well documented [161]–[165] and will not be recounted fully in this thesis, but rather only to give context in sufficient detail to understand the solutions developed in this thesis.

4.1 Plasmon Resonance

Metals, when downsized to dimensions smaller than the wavelength of light, exhibit a dramatic shift in their optical characteristics. This leads to the appearance of vivid colors in materials that are otherwise typically gray and shiny, a phenomenon originating from Localized Surface Plasmon Resonances (LSPR). LSPR occurs due to the resonant collective oscillations of the electrons that conduct electricity within these metals. The intriguing aspect of these resonances is their sensitivity to various properties of the nanoparticles, including the type of material, their size, shape, and the permittivity of their surrounding environment. This section aims to explore the underlying physics that dictate the optical behaviors of metallic nanoparticles, and to discuss their significance in the context of nanoplasmonic hydrogen sensing.

4.1.1 Plasmonic Nanoparticles

Electrons in atoms or molecules occupy fixed energy levels, and shifting an electron between these levels requires a precise amount of energy. This energy may correspond to the absorption of a photon that bridges the energy gap and thus induces an electronic transition. Similarly, in semiconductors, electron excitation from the valence to the conduction band occurs across the band gap.

When this absorption falls within the visible spectrum, it alters the material's color by absorbing specific photon wavelengths. However, metals lack a band gap, which means they don't require a set amount of energy to excite an electron. In the presence of light, a metal's free electrons quickly reposition to "neutralize" the light's electric field. If they manage to "counteract" the incoming light effectively, the metal reflects it. Yet, at frequencies beyond the metal's plasma frequency, electrons cannot adjust rapidly enough, preventing the reflection of the electromagnetic wave. Furthermore, metals can undergo interband transitions similar to semiconductors if their band structure permits electron excitation from high-density states to the Fermi level. This mechanism is responsible for the color seen in metals like gold, where an interband transition at 2.5eV(496 nm) absorbs blue light, giving gold its yellow hue. Metals with interband transition energy below visible light (< 1.65eV or > 750 nm) can absorb photons across the spectrum, resulting in a muted, less reflective appearance, as observed with, e.g., palladium. The interaction between metals and electromagnetic fields is quantifiable through their complex dielectric function.

A famous and historically relevant example of plasmonic metal nanoparticles are Faraday's studies of how the size and shape of gold nanoparticles affect their observed colour, as illustrated in Figure 4.1. The important takeaway from these results is that simply enlarging a gold nanoparticle does not necessarily change its optical properties in any meaningful way. However, adding an anisotropy to the nanoparticle, such as growing nanorods in this case, the observed properties may change considerably. This tends to be true in general for metal nanoparticles, but not necessarily for nanoparticles of other materials.

The mean free path of free electrons in gold particles is ≈ 50 nm, so interactions on nanoparticles smaller than this are only expected to occur on the surface [166]. Results of scattering and absorption spectroscopy has shown that standing resonance can occur for the electrons in the conduction band of a given nanoparticle, if the wavelength of the incident light is comparable to the nanoparticle's size [166]. This occurs because the wave front of incident light polarizes the electron density to the surface, thereby causing the electron density to oscillate with the frequency of incident light in a standing oscillation [166]. The observed colour is a consequence of these collective oscillations of electrons within the conduction band of the nanoparticle, also known as surface plasmon oscillations, as is discussed further in section 4.1.2.

4.1.2 Localized Surface Plasmon Resonance (LSPR)

Illustrated in Figure 4.2a is a nanoparticle smaller than the wavelength of an incident time-varying electric field. This field causes the electrons within the nanoparticle to shift relative to the atomic nuclei, creating a charge separation and, as a result, a restorative force aiming to reposition the electrons. This dynamic results in the nanoparticle acting as a damped harmonic oscillator, with its response characteristics—such as spring constant and damping—being influenced by the particle's shape, size, and material composition. Hence, when

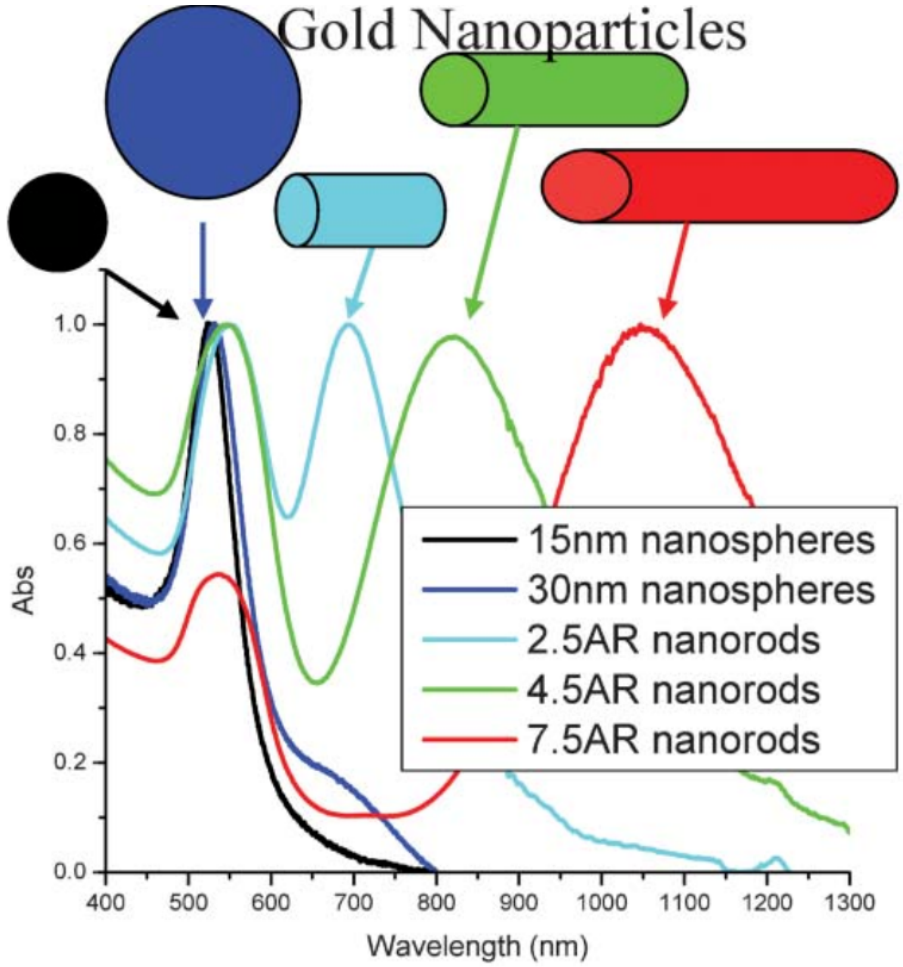

Figure 4.1: *Absorption of electromagnetic radiation within gold nanoparticles as a function of wavelength, shown here for five different samples of gold nanoparticles. Note that differently sized gold nanospheres have similar optical properties, whereas differently sized gold nanorods have very different properties. Absorption spectra doesn't change significantly with the size of nanospheres while for nanorods as the aspect ratio (AR) increases corresponding spectra moves towards higher wavelength. Reproduced from [166] with permission from The Royal Society of Chemistry.*

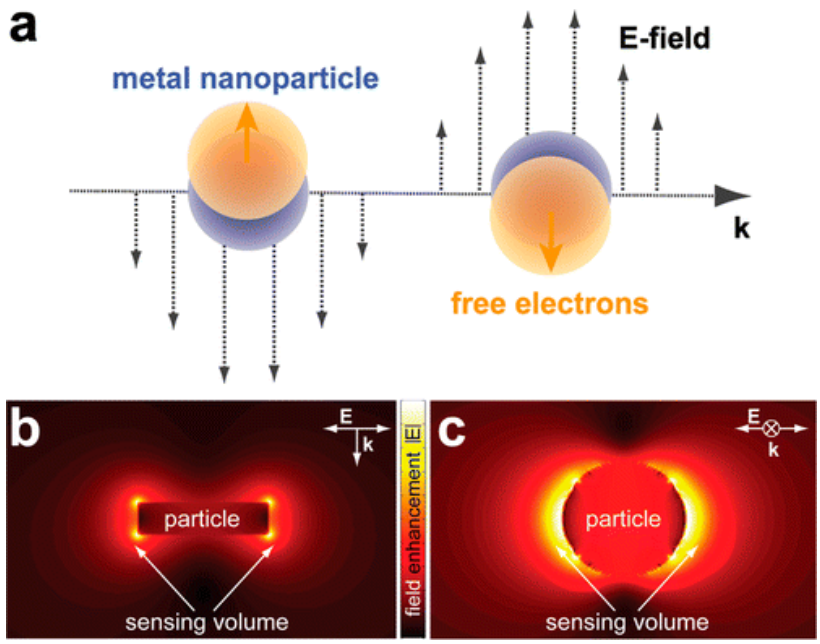

Figure 4.2: Top: *The polarization of the surface of metal nanoparticles can start to oscillate upon interaction with near-visible light, leading to the phenomenon of LSPR.* ***Bottom:*** *Side and top view of plasmonic gold nanoparticle illuminated at the LSPR wavelength. The locally enhanced dipolar field is illustrated schematically by the intensity of light, as calculated by the FDTD method. This enhanced field region effectively acts as a nanosized sensing volume, where local changes in poralisation can be detected. Reprinted with permission from [167]. Copyright 2012 American Chemical Society.*

the nanoparticle resonates at the frequency of incoming electromagnetic radiation, significant absorption and scattering occur, leading to visible coloration if the resonance frequency falls within the visible spectrum.

These enhanced field regions, as shown in Figure 4.2b-c, serve as probes for the dielectric environment surrounding the nanoparticle, making them effective for optical nanoscale sensing in various applications, including biosensing, chemical sensing, and gas sensing [162], [167]–[170].

To initially describe the LSPR phenomenon in a qualitative manner, one can consider a simplified model of a metallic sphere whose radius is significantly smaller than the wavelength of light. This model allows for the use of Gustav Mie's analytical solutions to determine the electromagnetic response of the sphere [171]. For such a small sphere, the electrostatic approximation is valid, indicating that the electric field does not vary spatially across the sphere at any given moment. Thus, when exposed to an electric field, E_0, an induced dipole moment is described by:

$$P = \varepsilon_d \alpha E_0 \tag{4.1}$$

Here, ε_d represents the dielectric function of the medium surrounding the sphere, and α denotes the sphere's polarizability, given by:

$$\alpha(\omega) = 4\pi r^3 \frac{\varepsilon_m - \varepsilon_d}{\varepsilon_m + 2\varepsilon_d} \tag{4.2}$$

where r is the sphere's radius, and ε_m is the metal's frequency-dependent dielectric function. The efficacy of the interaction between the particle and light is quantifiable as a cross-section, with specific calculations for scattering, absorption, and extinction cross-sections available for a perfect sphere [172].

$$C_{\text{absorption}} = k, \text{Im}(\alpha) = 4\pi k r^3 \text{Im}\left(\frac{\varepsilon_m - \varepsilon_d}{\varepsilon_m + 2\varepsilon_d}\right) \tag{4.3}$$

$$C_{\text{scattering}} = \frac{k^4}{6\pi}|\alpha|^2 = \frac{8\pi}{3} k^4 r^6 \left|\frac{\varepsilon_m - \varepsilon_d}{\varepsilon_m + 2\varepsilon_d}\right|^2 \tag{4.4}$$

$$C_{\text{extinction}} = C_{\text{absorption}} + C_{\text{scattering}} \tag{4.5}$$

Here, $k = \frac{2\pi}{\lambda}$ denotes the wavenumber associated with the incoming light. From these expressions, it's evident that enhancing the polarizability maximizes both absorption and scattering effects. Moreover, we observe that the scattering effect is proportional to r^6, contrasting with absorption, which is proportional to r^3. This implies that larger particles tend to scatter light more, while smaller particles are more inclined to absorb light. Critical to both phenomena is the condition when the denominator approaches zero, defined as:

$$\varepsilon_m(\omega) = -2\varepsilon_d \tag{4.6}$$

which establishes the resonance condition for localized surface plasmon resonance (LSPR).

It's noteworthy that the optical attributes of various metals differ significantly, leading to variations in ε_m. A simplified approach to model these metals is through the Drude model, which considers electrons as freely mobile, thereby neglecting potential interband transitions. This model is encapsulated by [172]:

$$\varepsilon_m(\omega) = 1 - \frac{\omega_p^2}{\omega^2 + i\Gamma\omega} \tag{4.7}$$

where ω_p signifies the plasma frequency, and Γ represents a damping factor accounting for energy dissipation mechanisms within the material. The plasma frequency is given by:

$$\omega_p = \sqrt{\frac{Ne^2}{\varepsilon_0 m_e}} \tag{4.8}$$

with N, e, and m_e denoting the density of free electrons, the electron charge, and mass, respectively, and ε_0 is the vacuum permittivity.

Assuming minimal damping effects and integrating eq. 4.7 into eq. 4.6 to solve for ω, we obtain the LSPR frequency as:

$$\omega_{LSPR} = \frac{\omega_p}{\sqrt{1 + 2\varepsilon_d}}$$

or, when expressed in terms of wavelength:

$$\lambda_{LSPR} = \lambda_p\sqrt{2\varepsilon_d + 1}$$

where λ_p is the plasma wavelength, defined as $\lambda_p = \frac{c}{\omega_p}$ with c being the speed of light. Although this model simplifies the concept, it effectively captures the essence of LSPR, including its sensitivity to the refractive index of the surrounding medium, which induces a spectral redshift as the index increases.

However, in addition to the sensitivity of λ_{LSPR} to the dielectric constant of the surrounding medium ε_d, the resonance wavelength is also sensitive to changes in the metal's plasma frequency ω_p, or equivalently, the plasma wavelength λ_p. This sensitivity arises because λ_{LSPR} is directly proportional to λ_p, as seen in the expression:

$$\lambda_{\mathrm{LSPR}} = \lambda_p\sqrt{2\varepsilon_d + 1} \tag{4.9}$$

The plasma frequency ω_p is a fundamental property of the metal and is defined by the free electron density N:

$$\omega_p = \sqrt{\frac{Ne^2}{\varepsilon_0 m_e}} \tag{4.10}$$

where e is the elementary charge, ε_0 is the vacuum permittivity, and m_e is the electron mass.

In hydride-forming metals such as palladium (Pd), the absorption of hydrogen atoms leads to the formation of palladium hydride (PdHx), which alters the free electron density N within the metal lattice. As hydrogen atoms are incorporated into the metal lattice, they interact with the conduction electrons, effectively modifying the electron density and the electronic structure of the metal. This change in N directly impacts the plasma frequency ωp, and consequently, the plasma wavelength λ_p.

The alteration in λ_p results in a shift of the LSPR wavelength λ_{LSPR}, which can be expressed as:

$$\Delta\lambda_{\mathrm{LSPR}} = \Delta\lambda_p\sqrt{2\varepsilon_d + 1} \tag{4.11}$$

where $\Delta\lambda_p$ represents the change in plasma wavelength due to hydrogen absorption.

Therefore, the sensitivity of the LSPR to changes in ω_p (or λ_p) is the critical mechanism exploited in hydrogen sensing using plasmonic nanoparticles. Unlike the sensitivity to the surrounding dielectric environment (changes in ε_d), which is the basis for many plasmonic sensors detecting changes external to the nanoparticle, hydrogen sensing leverages intrinsic changes within the nanoparticle's material properties caused by hydrogen absorption.

This mechanism allows for the direct optical detection of hydrogen gas through measurable shifts in the LSPR wavelength. By monitoring $\Delta\lambda_{\mathrm{LSPR}}$, the amount of hydrogen absorbed by the nanoparticles can be quantitatively determined, enabling sensitive and specific hydrogen gas sensing.

Despite assuming negligible damping for simplicity, real nanoparticles experience damping that significantly influences their optical properties, with the broadening of the resonance peak in the frequency domain indicating a reduced LSPR lifetime. Damping arises from both radiative (scattering) and non-radiative processes, such as electron-hole excitations within the metal.

While the aforementioned description aids in conceptual understanding, actual nanostructures deviate from ideal spheres, and their dielectric functions are more complex than what the Drude model suggests. Addressing realistic materials and shapes necessitates alternative approaches such as electrodynamic simulations, and the precise shape, size and structure of nanoparticles have profound impacts on the nature of LSPR behaviour.

4.1.3 Nanoparticle Size, Shape and Structure Impact on LSPR

Gold (Au) and silver (Ag) are the cornerstone materials in plasmonics, primarily due to their ability to sustain strong localized surface plasmon resonance (LSPR) bands in the visible light spectrum. Their electronic structures allow for efficient plasmon excitation, and their chemical stability makes them ideal for long-lasting plasmonic devices. Specifically, gold is chemically inert and does not easily oxidize, while silver, though more efficient in terms of plasmonic properties, can suffer from oxidation, requiring protective coatings. However, the scope of materials utilized in plasmonics extends beyond these two, incorporating copper (Cu), platinum (Pt), Pd, nickel (Ni), and aluminum (Al), as well as noble metal alloys [173], [174] and nonmetallic entities like nitrides [175]. Each material's unique electronic structure influences its plasmonic properties by altering the dielectric function.

The size of nanoparticles plays a crucial role in determining their optical properties. As nanoparticle size increases, the interaction with the incident electric field becomes non-uniform due to the larger spatial extent of the particle compared to the wavelength of light. This non-uniformity introduces retardation effects, where different parts of the nanoparticle experience different phases of the electromagnetic field. This phase shift between the incident and induced dipolar fields within the nanoparticle causes a redshift in the resonance frequency because the collective electron oscillations are delayed. Moreover, as Eq. 4.2 suggests, the scattering intensity scales with particle volume, leading to enhanced radiative damping. This effect results in larger nanoparticles having a shorter resonance lifetime and a broader resonance peak due to the increased damping. The non-uniform electric field across the particle's surface further complicates the plasmonic response, particularly for larger particles, where multiple scattering events and phase differences occur across the nanoparticle. Non-radiative damping, primarily through absorption, also contributes to further broadening of the resonance peak, especially in materials with significant absorption or interband transitions in the visible spectrum, like Pt or Pd, leading to very broad resonance peaks. [176].

For complex-shaped nanoparticles, while analytical extensions of Mie's theory for spheres to spheroids exist [177], real-world nanostructures often have more intricate forms and may be placed on substrates with their own complex surfaces. To accurately simulate the optical properties of such entities, reliance on numerical solutions to Maxwell's equations is necessary [178]. Techniques including finite difference time domain (FDTD) [179], discrete dipole approximation (DDA) [177], finite element method (FEM), boundary element method (BEM) [180], and the modified long wavelength approximation (MLWA)[177], [178] enable predictions on how single nanoparticles or assemblies will interact with electromagnetic fields, providing a thorough understanding of their plasmonic behavior.

4.1.4 Nanoplasmonic In Situ Spectroscopy

Applying the LSPR method described above, it is possible to measure the LSPR excitations within a single or an array of Au nanoparticles with near-visible light to study the physical changes within other nanoparticles, such as catalysts, in the vicinity. The optical characteristics of metal nanoparticles, important in plasmonics, are influenced by their dielectric properties, size, and the environment around them, which can lead to observable changes in their optical spectra [181]–[183]. This principle underpins plasmonic sensing, a technique that leverages absorption and scattering spectra from LSPR-active structures to detect minute variations in via change in ω_p or around via change in ϵ_d the nanoparticles. This approach was initially demonstrated two decades ago with gold nanoparticles used to sense the binding of antigens to ligands on their surfaces [184]. An example of how optical changes in nanoparticles are detected, such as through extinction spectroscopy, is shown in Figure 4.3. The figure illustrates how environmental changes around the particles alter their LSPR resonance condition via changing ϵ_d. Key parameters like the resonance wavelength (λ_{peak}), peak intensity (Ext), and the full-width-at-half-maximum (FWHM) help quantify these changes, providing insights into the transient and steady-state dynamics of the system [185], [186].

Plasmonic sensors find applications across various domains, predominantly in biological sensing [162], [187], [188], but also in detecting chemical and gas changes [189], [190], as well as monitoring catalytic reactions [167], [168]. Their utility in solid-state chemical reaction sensing is particularly noted for its ambient pressure and high-temperature applicability, employing low-power visible light for non-invasive measurements. Relevant examples include tracking the phase transition in Pd and Mg from metal to hydride [190]–[194] and the oxidation of Cu nanoparticles [195], [196].

The sensitivity of LSPR-based refractive index change sensing is confined to a small volume around the nanoparticle due to the rapid field decay, with the enhancement region shaped by the particle's resonance and incoming field polarization, as well as its geometry, where sharp edges typically greatly enhance local field intensities. This specificity allows for detecting changes down to the scale of single molecules [170], [197].

Practically, an induced shift in plasmonic resonance wavelength as a result of tiny changes within a sensing volume of plasmonic sensing particles defined by their enhanced field region underlies the basis of the data analysis of the experiments in this thesis. This peak shift can be extremely subtle, potentially only fractions of nanometres, but this sensitivity has been shown to be possible to achieve through function fits of the data [186]. However, to fully realise the potential of such plasmonic sensing, we require a method of analysis that is sensitive to such extraordinarily subtle changes in the optical spectra and robust to, e.g., optical noise.

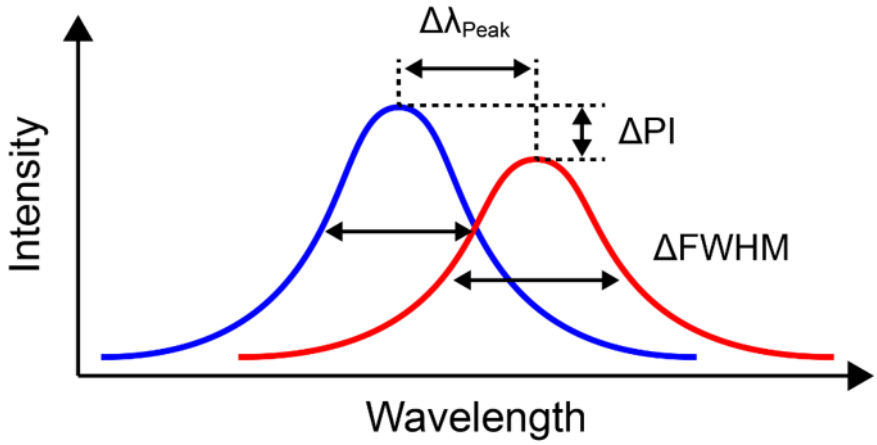

Figure 4.3: *Idealised signal readout of a generic nanoplasmonic in situ spectroscopy experiment, in which physical surface changes on the sample material or changes of the bulk sample material itself induce a change in the peak position of the resonance wavelength. Additional observables, such as the difference in particle intensity (ΔPI) and full-width at half-maximum (Δ FWHM) may provide additional information or provide a better signal-to-noise ratio.*

4.1.5 Nanoplasmonic Hydrogen Gas Sensing

In hydrogen gas sensing, specifically using nanoplasmonic sensors, one of the key mechanisms does not arise from localized field enhancements, as often seen in surface plasmon resonance techniques, but rather from physical changes within the particles themselves due to hydrogen absorption, i.e., a change of ω_p. When hydrogen molecules dissociate and are absorbed into hydride-forming metal nanoparticles such as Pd, several phenomena occur that contribute to the sensing mechanism. Firstly, hydrogen absorption leads to a change in the particle's dielectric function. As hydrogen atoms enter the metal lattice, the electronic properties of the particle, and thus ω_p, are altered. This change is observable in the nanoparticle's LSPR characteristics. The LSPR frequency, which is highly sensitive to the dielectric environment and electronic structure of the nanoparticle, shifts as hydrogen is absorbed. This shift in the LSPR peak can be measured optically and is directly related to the concentration of hydrogen absorbed by the nanoparticle. The magnitude of this spectral shift is proportional to the amount of hydrogen absorbed, making it a reliable indicator of hydrogen concentration [198].

Secondly, the volume expansion of the nanoparticles due to hydrogen absorption also plays a role in the sensing mechanism. Hydrogen absorption into the metal lattice causes the lattice to expand, leading to a slight increase in the volume of the nanoparticle. This volumetric change, although minor, contributes to the overall optical response because it influences the scattering and absorption

properties of the particle. The degree of this expansion depends on the type of metal and its ability to form hydrides. For example, Pd nanoparticles exhibit notable volume expansion when forming palladium hydride, which further shifts the LSPR peak, enhancing the sensitivity of the sensor in detecting hydrogen gas [199]. Thus, in hydrogen sensing, the primary mechanism relies on changes in the particle itself, both through alterations in its dielectric function due to hydrogen absorption and through minor contributions from volume expansion. These combined effects lead to detectable optical shifts, which are critical for the real-time and precise detection of hydrogen concentrations in a range of environments.

4.2 Neural Networks for Nanoplasmonic Sensing of Hydrogen Gas

To reiterate, neural networks are efficient piecewise linear function approximators, and are theoretically capable of approximating any function by delineating hyperplane partitions within the input space. This principle underpins the neural network design adopted in our research, where we endeavour to develop architectures with appropriate representational ability to discover general function approximations between measured LSPR extinction spectra and surface state of nanoparticles across time. Here, we focus on the application of hydrogen gas sensing, aiming to significantly advance the field by offering precise, real-time detection capabilities essential for both safety and efficiency in hydrogen energy technologies.

4.2.1 Nanoplasmonic Hydrogen Sensing

Efforts are underway to curb greenhouse gas emissions through significant investments in hydrogen gas technologies. These technologies encompass a wide array of applications including H_2-powered transport modes, refueling infrastructures, and energy production methods. As H_2 is flammable in air, all of these applications present heightened risks of H_2-related accidents due to their operation in confined or public spaces. The imperative for robust H_2 safety sensors is underscored by these potential risks, necessitating devices capable of precise H_2 detection in air [200], [201]. Moreover, H_2 sensors play a crucial role in monitoring processes across various applications, from domestic heating to fuel cell efficiency. Today, the market offers a variety of H_2 sensors utilizing principles such as resistive, electrochemical, and catalytic transduction [198].

Optical nanoplasmonic sensors, leveraging LSPR in Pd nanoparticles or similar nanostructures, have emerged for purposes of H_2 detection [198], [202]. These sensors operate on the principle of H_2 dissociation and absorption at the nanostructure surface and in the bulk, respectively, inducing reversible changes in optical properties visible as spectral shifts $\Delta\lambda_{peak}$ in the LSPR peak [198], [199]. The linear relationship between spectral shift and H_2 absorption levels, along with the inherent selectivity of hydride-forming metal-based sensors

against, e.g., CO_2 and hydrocarbons, positions nanoplasmonic sensors as highly effective [203]–[205]. Addressing the need for improved H_2 sensing capabilities, performance targets have been set, notably by the US Department of Energy (DoE), aiming for sensors to detect hydrogen concentrations rapidly (within 1 second) in a range from 0.1% to 4% under ambient conditions (DoE-targets [206], [207]). While some studies have showcased sensors meeting these response times, these tests were conducted under ideal conditions, significantly removed from the complexity of real-world applications where sensor performance can be degraded by impurities, oxygen and water [208]–[211]. In this thesis, the focus is on using deep learning to enable detection limits in the low ppm range, rapid response times, and notable selectivity and resistance to deactivating chemical species, aligning with or surpassing benchmarks set by the US DoE [198], [212]–[214].

In the realm of H_2 sensing, environments can broadly be categorized into two: those with abundant oxygen and the oxygen-starved or "inert" environments. The latter is crucial for safety in large-scale H_2 installations like fuel-cell powered ships and H_2-powered airplanes inerted environments to prevent flammable mixtures. These environments are typically purged with an inert gas, such as nitrogen, but often contain impurities like water, carbon monoxide, or sulfur oxides, which can interfere with sensor function. The presence of these "poisoning" molecules, alongside the lack of oxygen, challenges the efficacy of traditional H_2 sensors, as they require oxygen to operate and are susceptible to surface poisoning which hinders their ability to detect hydrogen [211].

The advancement in this field has traditionally focused on developing new sensing materials, nanostructuring, or refining physical sensing mechanisms [208], [210], [215]–[219]. However, the potential for using deep learning to improve sensor response time, a critical aspect of sensor performance, remains largely untapped, despite its proven effectiveness in enhancing accuracy and sensitivity in gas sensors [220]–[225]. This is explored in papers II and IV.

A primary impetus for the research presented in Paper II arises from the limitation that current H_2 sensing in general, and nanoplasmonic H_2 sensors in particular, do not perform well under conditions of high humidity [198]. This presents a substantial challenge, as sensors that can tolerate such conditions are critically needed for both process monitoring and environmental safety applications, where fluctuations in relative humidity or the inherently humid hydrogen feed in proton-exchange membrane fuel cells are common. The DL-based method introduced in Paper II also allows for the rapid prediction of H_2 concentrations in inert environments, overcoming the limitations imposed by traditional sensors' need for oxygen and their vulnerability to deactivation by contaminants [211], [226], [227].

Our analysis technique involves a more comprehensive utilization of the output data from plasmonic hydrogen sensors by leveraging the full scattering or extinction spectra over time rather than reducing this data to single spectral descriptors as done in the state of the art (Figure 4.4d). This method acknowledges the wealth of information present in the spectrum's temporal evolution,

which can be instrumental in predicting sensor saturation levels or detecting gradual changes in H_2 concentration more quickly in challenging environments [228].

4.2.2 Transformers for Nanoplasmonic Hydrogen Sensing

The transformer architecture used in Paper II is shown in Figure 4.5, and was used to enable nanoplasmonic sensing in highly humid conditions in air [229]. Here, the entire spectral data, i.e, the full LSPR spectrum, across a short time window is considered in every mapping, providing more information compared to standard approaches of reducing spectra into single descriptors like centroid, FWHM, peak position and other similar parameters. To address potential signal drift, we implement min-max normalization based on an initial calibration curve for each measurement series, complemented by batch normalization layers [230] that adjust normalization parameters during the training phase and lock them during inference, as shown in Figure 4.5b. Additionally, to enhance the model's capability to uncover physically meaningful and broadly applicable correlations, Dropout [231] layers (Figure 4.5c) are introduced, periodically resetting weights during training to promote the learning of durable correlations resistant to random fluctuations. The model's computational efficacy and adaptability are further augmented by connecting a series of dense layers, each followed by BatchNorm and Dropout layers, through a skip connection, linking this setup with three similar configurations (Figure 4.5a). These skip connections help counteract the vanishing gradient issue in deep networks, where the gradient during backpropagation in earlier layers disappears in deep networks. The series of extinction spectra generated by the sensor is represented using positional and time-based embeddings, achieved by 1D-convolution and sinusoidal encoding methods, respectively. These encoded forms are then inputted into a multi-head attention module. Within this module, the varied representations formed by the attention heads are combined with the initial embeddings and processed by the dense layer-based computational framework, which is linked to an L1 norm loss and optimized via backpropagation.

The most significant results of Paper II that were generated using the approach outlined above are summarized in Figure 4.6, wherein the aforementioned transformer is trained to predict hydrogen concentration from sequences of acquired nanoplasmonic sensor spectra in humid conditions. We find that such an approach enables an LoD substantially below the DoE goal of 0.1% H_2 for sensor operating temperatures above 30 ^{o}C, compared to standard LSPR sensor readout methods.

4.2.3 Long Short-term Transformer Ensemble Model for Accelerated Sensing

In Paper IV, we extended the transformer architecture shown in Figure 4.5 to consider longer time periods (effectively taking the full measurement history of the sensor into account), include estimates of uncertainty within its own

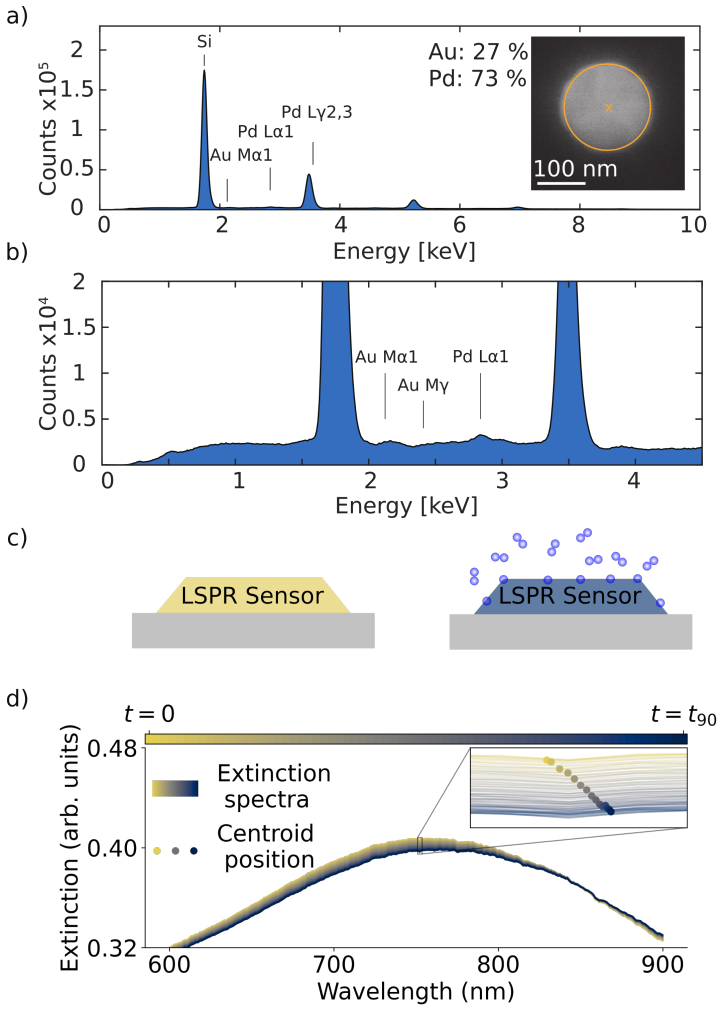

Figure 4.4: *Characterization and working principle of* PdAu *alloy nanoparticle plasmonic H_2 sensors. (a) A single Pd70Au30 alloy nanodisk's Energy-dispersive X-ray (EDX) spectrum corroborating its targeted composition. (b) A detailed view of the EDX spectrum from (a), highlighting the Pd and Au peaks up to 4.5keV. (c) A diagram illustrating the plasmonic H_2 sensor's operating mechanism: the absorption of hydrogen by hydride-forming metal nanoparticles alters their localized surface plasmon resonance frequency, resulting in a detectable color shift during spectroscopic analysis in the visible spectrum. (d) An example of the spectral shift observed in the Pd70Au30 alloy plasmonic sensor's extinction spectrum during hydrogen incorporation into the nanoparticles' crystal structure. Inset: zoom-in reading the subtle spectrum changes induced by hydrogen absorption.*

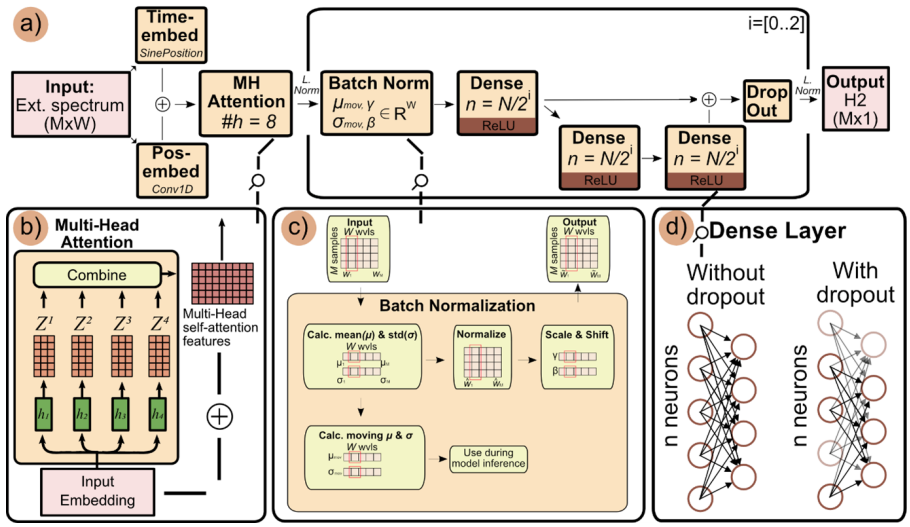

Figure 4.5: *Schematic of the Transformer-based neural network architecture employed for enhancing the detection capability of nanoplasmonic sensors in humid conditions. The architecture integrates multiple advanced techniques: (a) time and position-based embeddings for the input spectral data, (b) multi-head attention for complex pattern recognition within the spectral data, (c) batch normalization to stabilize learning and facilitate model generalization, and (d) layers incorporated within the dense neural network to prevent overfitting. The combination of these elements contributes to a robust learning mechanism capable of handling the intricate nature of nanoplasmonic sensor spectra, particularly in the presence of challenging environmental conditions, such as high humidity.*

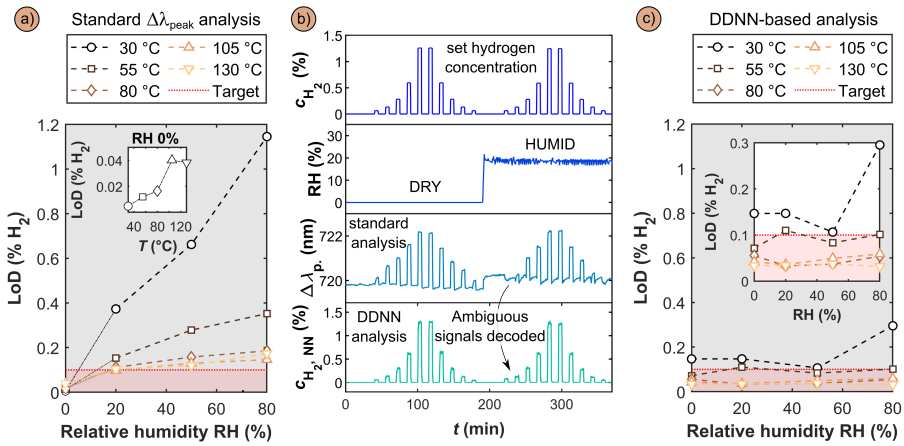

Figure 4.6: *Achieving a Limit of Detection (LoD) below 0.1% (1000 ppm) H_2 in air at 80% RH utilizing deep learning. a) Sensor LoD determined using standard $\Delta\lambda_{peak}$ measurement across various sensor operating temperatures and RH levels. It is noted that at over 20% RH, all sensors do not meet the US Department of Energy (DoE) benchmark of an LoD $< 0.1\%$ H_2. b) A comparison of the sensor's reaction to H_2 concentration pulses under dry and 20% RH conditions at an 80°C operating temperature, recorded via the conventional $\Delta\lambda_{peak}$ method and the DL-enhanced readout, $c_{(H_2,NN)}$. c) LoDs achieved through the DL-enhanced readout indicate an almost RH-independent LoD substantially below the DoE goal of 0.1% H_2 for operating temperatures of 80 °C or higher. An inset provides a closer look at the 0 - 0.3% H_2 LoD range. Figure reproduced with permission from Paper II.*

predictions and enable accelerated prediction of hydrogen concentration compared to the sensor's physical response. Thus, we implement a variant of the transformer known as the long short-term transformer (LSTR), as shown in Figures 4.7 and 4.8 [232].

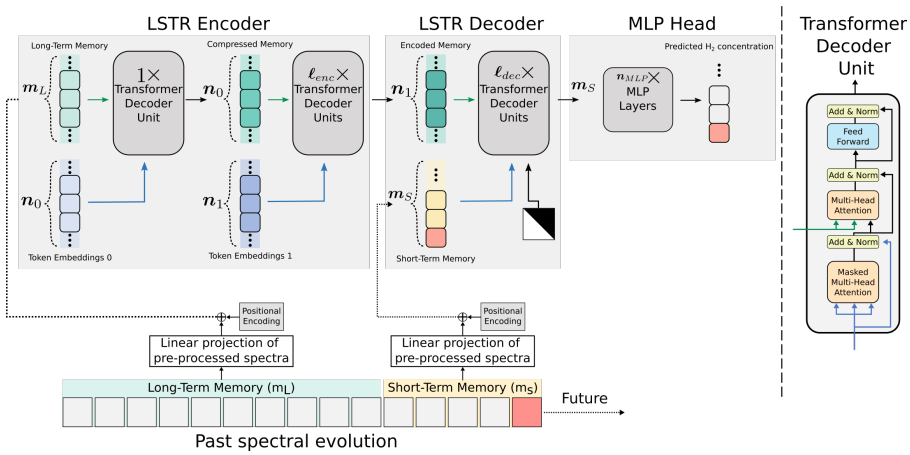

Figure 4.7: *Illustration of the Long Short-Term Transformer (LSTR) structured through an encoder-decoder design. Specifically, the LSTR encoder reduces the extensive long-term memory of size m_L into n_1 encoded latent features, while the LSTR decoder utilizes these encoded memories in conjunction with short-term memory of size m_S to facilitate action recognition in real-time. The construction of both the LSTR encoder and decoder employs transformer decoder units, processing both input tokens (illustrated by dark green arrows) and output tokens (shown as dark blue arrows). During the inference phase, the LSTR evaluates each incoming frame in real-time, operating without access to future context. It sequentially determines the actions within each frame through its encoder-decoder framework, lacking any forward-looking context. The dashed brown arrows depict the flow of data for both long- and short-term memories, adhering to a first-in-first-out (FIFO) principle. Figure adapted from Ref. [233]*

To properly account for long-term drifts inherent to LSPR sensors, the model needs long-term information regarding what influences the shift of spectra. However, the multi-head attention mechanism of a standard transformer means that the memory usage of the architecture scales quadratically with the sequence length, making long sequences, as produced by continuous sensor operation, intractable to compute. The efficacy of the LSTR model is also influenced by how the sensor spectra data is pre-processed. The necessity for pre-processing arises from the sensor response's drift over time, primarily due to the light source intensity's long-term changes, and minor variations in the extinction spectra from different separate and independent measurements caused by slight alterations in the sensor's positioning within the measurement chamber for each test. To address this, we applied multiple pre-processing strategies [234], merging them into a unified dataset. Consequently, the deep learning

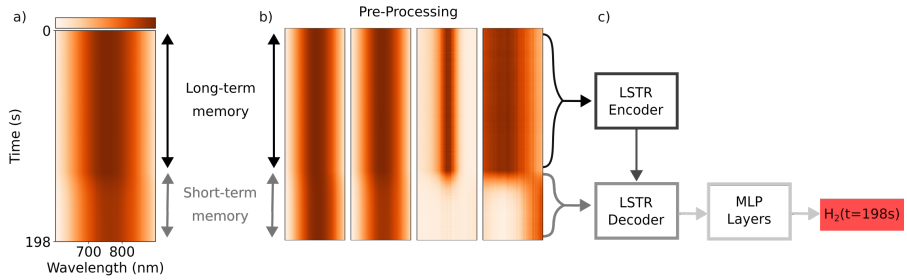

Figure 4.8: *Long Short-term Transformer Ensemble Model for Accelerated Sensing: (a) The model inputs are sequences representing the sensor's spectral response history, with the depicted sequence spanning 600 timesteps or 198 seconds. (b) Initially, the sequence is partitioned into long and short-term segments and undergoes four preprocessing techniques: wavelength-specific min-max normalization, standard normal variate transformation, comprehensive min-max normalization, and level adjustment. These preprocessed segments are then merged and supplied to the LSTR. (c) Within the LSTR, the long-term segment is first condensed into a compact latent form by the encoder. Subsequently, the decoder focuses on the short-term segment, identifying important temporal characteristics while referencing the encoded long-term memory. These characteristics are processed through a series of dense layers, culminating in the prediction of the current H_2 concentration in the sensor environment.*

models received a time series input, where each sequence element represented a combination of these pre-processing methods (Figure 4.8a).

Further, we opted for an ensemble approach with multiple LSTR models to acquire a measure of uncertainty along each prediction. This strategy is driven by the high-stakes nature of hydrogen sensor applications, ensuring more dependable predictions and providing a measure of uncertainty along with each prediction. This is achieved by averaging the forecasts of several LSTR models to determine the mean and standard deviation of predictions. By combining these modeling choices, we developed the Long Short-term Transformer Ensemble Model for Accelerated Sensing (LEMAS) in Paper IV, a system characterized by an ensemble of LSTR models adept at quickly forecasting H_2 concentration and offering a reliability measure based on a long series of pre-processed spectral data.

By implementing neural networks in the form of transformers, we demonstrate in papers II and IV a significant reduction in response times — up to 40 times faster — when exposed to H_2 pulses in inert gas environments and were able to achieve LoDs ranging between 0.01 - 0.06% H_2 (100 - 600 ppm) in 80% RH in synthetic air for sensor operating temperatures of 80-130^{o}C. Thereby, the presented AI-enhanced sensors meet the US DoE performance target of LoD $< 0.1\%$ H_2 in humidity in air with significant margin. Similarly, the ISO 26412:2010 sensor signal robustness standard for operation in humid air was met down to 0.06% H_2 concentration for a sensor operation temperature of 80 ^{o}C.

This technology is vital for early detection of H_2 leaks, allowing ample time for implementing safety measures to prevent accidents [233]. Moreover, LEMAS' ensemble modeling approach provides uncertainty estimates, enhancing its reliability for safety-critical applications, such as H_2 sensing at hand.

Chapter 5

DL for Mass Spectrometry in Single Particle Catalysis

In experimental science, isolating small signals from noise is a crucial and general challenge. One field where this is particularly true is in single particle catalysis, where chemical reaction products formed on single nanoparticles are to be quantitatively measured. These very weak signals make their analysis challenging. We address this in Paper III by combining nanofluidic reactors, quadrupole mass spectrometry (QMS), and a deep learning-based denoising auto-encoder to detect reactions on minuscule surfaces in the form of single Pd nanoparticles, using CO oxidation on Pd as an example. In this chapter, we go through the fundamentals of catalysis, quadrupole mass spectrometry for catalytic applications (and beyond) and deep-learning based auto-encoders for the purposes of denoising.

5.1 Catalysis

Catalysis is the process in which new chemical reaction pathways are enabled, thereby improving the rate of chemical reactions, by the use of a catalyst material which is not consumed during the reaction. Generally speaking, the rate of reaction improves as a result of new intermediate reaction steps enabled by the reactant's interaction with the catalyst surface, with lower activation energies [235]. These crucial effects have made catalysts so pervasive and ubiquitous that an estimated $\approx 90\%$ of commercially produced chemicals involve catalysts at some stage of the manufacturing process [236], and an estimated one-third of the GNP of the US involves a catalytic process somewhere in the production chain. Some of the major applications for catalysts today include energy production, food production and the production of (bulk and fine) chemicals, and one of the most common uses is the catalytic converter within automobile exhaust aftertreatment systems [236]. Motivated by this common application, this project is based on an experimental setup in which

CO and O_2 can be converted to CO_2 on the surface of a few or even a single catalyst nanoparticle.

However, most studies related to the chemical mechanisms of catalysts are historically done on ensembles of catalytic particles, thereby potentially missing intricate subtleties of what occurs on the surfaces of each individual nanoparticle [237]. Indeed, a long-term goal for the field is to develop methods to isolate and investigate single nanoparticles during the catalytic process in order to investigate how the size, shape and structure of individual nanoparticles impact their catalytic efficiency. This, with the goal to eventually make it possible to manufacture catalysts consisting solely of the most effective nanoparticles, leading to significant increases in overall catalytic efficiency. Importantly, we note here that single nanoparticle catalysis has hitherto been an exceedingly elusive sub-field of catalysis science, owing to the decades-long issues of the "materials- and pressure gaps" which is at the heart of the problem we are trying to solve in this thesis in the long-term.

5.1.1 Materials- and pressure gap

The crux of the pressure gap in catalysis is that analysis of the surface catalytic reactions on (single) nanoparticles has been done successfully in high vacuum conditions, and becomes increasingly more difficult at high (and thus practically more relevant) pressures [238]. A key reason is that electrons, the experimental probes used, are incompatible with elevated pressure due to extremely short mean free paths [238]. The materials (and complexities) gap is less fundamental, and is instead the consequence of the gap between simplified linear models and the complex structural nature of real catalyst materials [238].

In Paper III, we present a novel method of approaching the bridging of the pressure and materials gaps at the single nanoparticle level by measuring the activity of catalysts down to single nanoparticles. This is enabled by applying modern deep-learning algorithms to improve the resolution and limits of quantification of QMS data, obtained from nanofluidic reactors decorated with (single) catalyst nanoparticles.

5.1.2 Catalytic Surface Reactions

In general, chemical reactions involve the breaking of chemical bonds between or within molecules and the subsequent forming of new ones. This process is associated with a transfer of energy, related to an activation energy E_{act}, such that the reaction is (effectively) impossible if the excess energy of the reactants is less than E_{act} [239]. If the excess energy of the reactants is provided by thermal energy, then the probability of the reactants having sufficient energy for the reaction to occur is proportional to the Boltzmann distribution $P \propto e^{-E_{act}/k_B T}$, meaning that the overall reaction rate increases with higher temperature and decreases with higher activation energy [239]. According to the Boltzmann distribution, the probability of a system being in a particular state i with

energy E_i is given by:

$$P_i = \frac{1}{Z} e^{-E_i/k_B T}, \tag{5.1}$$

where Z is the partition function, summing over all possible states of the system.

By introducing a catalyst to a reaction, the activation energy can be effectively decreased by providing the reactants with intermediate reaction steps and transition states that are energetically more favorable compared to the uncatalyzed reaction [235].

Important to note is that catalytic reactions take place on the catalyst surface, leading to several catalytic reaction rates being very sensitive to the size, shape and structure of the catalyst's surface [235]. One early example of a single particle catalysis experiment is the study in 2008 by Mulvaney et.al [168]. In this work, surface plasmon spectroscopy was implemented to study the catalytic behaviour on the surface of single gold nanoparticles during oxidation of ascorbic acid. This study, and most studies on single particle catalysts, were carried out in the liquid phase. Hence, a much less studied and understood aspect of single particle catalysis is that of reactions in the gas phase [240]. To fill this gap, here we consider the surface reaction of CO and O_2 in the gas phase, on catalytic palladium nanoparticles.

5.1.3 Oxidation of CO on Palladium Surfaces

The aforementioned CO oxidation reaction over palladium nanoparticles is a suitable model reaction for the purpose of method development aimed at in this thesis. In this section, I thus first introduce the elementary reaction steps of the CO oxidation reaction.

The reaction can be written as:

$$CO \ + \frac{1}{2} \ O_2 \ \rightarrow CO_2. \tag{5.2}$$

In order for this reaction to occur at relatively low temperatures and pressures (relevant to most industrial and automotive applications), it is necessary to include a catalyst in the reaction to reduce its activation energy [241]. In the presence of a catalyst, reaction 5.2 is split into five elementary steps; adsorption of CO and O_2 on the surface of the catalyst, dissociation of O_2 into separate O molecules on the surface, a surface reaction and finally desorption of CO_2. Let $*$ represent an active site on the surface of a catalyst, and X^* represent the adsorption of molecule X on the surface of the catalyst. The catalyzed reaction can then be expressed as [241]

$$\begin{aligned} CO + * &\rightarrow CO^*, \\ O_2 + * &\rightarrow O_2^*, \\ O_2^* + * &\rightarrow 2O^*, \\ CO^* + O^* &\rightarrow CO_2^* + *, \\ CO_2^* &\rightarrow CO_2 + *. \end{aligned} \tag{5.3}$$

This is an example of the Langmuir-Hinshelwood mechanism, whereupon the maximal reaction rate is expected to occur when the surface coverage of carbon monoxide θ_{CO} and oxygen θ_O (molecular oxygen dissociates upon adsorption on the Pd surface) are equal, according to equation 5.4.

$$r = k\theta_{CO}\theta_O, \tag{5.4}$$

where k is the reaction rate constant. An important issue arises from the fact that CO adsorbs onto a single site on the surface, whereas O_2 needs to dissociatively adsorb onto two sites [237]. Since these species compete for the same active sites on the surface, the coverage of CO can become much greater than for O in relation to their gas phase concentrations. This leads to a lowered CO oxidation rate and is known as CO poisoning [237]. This phenomenon contributes to the 'cold-start' problem in automotive catalytic converters, where the catalyst is not sufficiently active at low temperatures immediately after engine start due to CO poisoning, leading to higher emissions of CO until the converter reaches its operating temperature [241]. Increasing the temperature can mitigate CO poisoning by enhancing the desorption rate of CO and thereby reducing its equilibrium surface coverage. This allows more O_2 to access the active sites, improving the CO oxidation rate [241].

5.1.4 Quadrupole Mass Spectrometry

An experimental technique widely used for identifying and quantifying the products of catalytic reactions, is Mass Spectrometry (MS), which ionizes the analyzed species to enable their characterization and quantification based on mass-to-charge ratios.

The specific manner in which incoming particles are ionized may vary between specific MS solutions, but it may for example be achieved by bombarding them with electrons [242]. Upon bombardment, some particles may just become charged whilst others may become fragmented before becoming charged, which is important for distinguishing molecules that have the same mass-to-charge ratio in so-called cracking patterns [242]. When this has been accomplished, the now-ionized particles can be separated by their mass-to-charge ratio through various methods [242]. Finally, the now-separated particles hit a detector that counts them, resulting in the aforementioned mass spectrum representing the number of incident particles of a given mass-to-charge ratio as MS intensity signal [242].

The method used in this thesis is Quadrupole Mass Spectrometry (QMS), whereupon its namesake comes from the detector's use of four parallel rods, through which incoming ions' trajectories are modified to enable mass selection [242]. Two of the parallel rods are put under a combined DC and AC potential, with the other two remaining parallel rods also put under the same combined potential, such that each opposing pair of parallel rods has the same combined potential [242]. This technique utilizes the fact that only particles of a certain mass-to-charge ratio will successfully travel the entire length of the quadrupole for a certain ratio of DC and AC voltages and reach the detector, whereas all

other particles simply collide with the surrounding rods due to their unstable trajectories [242]. This is shown schematically in Figure 5.1. Indeed, through solving the relevant Mathieu's differential equations, the exact trajectories of particles of well-defined mass-to-charge ratio within a given voltage ratio can be calculated [242]. Based on these equations, the voltage controller can be set to "scan" specific ratios of voltages corresponding to the successful transfer of particles with given mass-to-charge ratios, by oscillating continuously between different ratios of voltages [242].

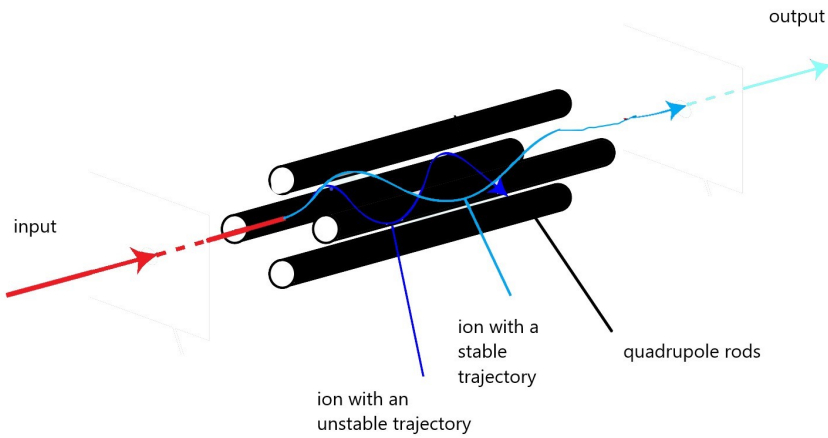

Figure 5.1: *Schematic of the function of the mass filter used in a QMS. Ions pass through four rods which are put under a combined DC and AC potential, with each pair of rods having the same combined potential. Only particles of a certain mass-to-charge ratio will successfully travel through the mass filter and be measured.*

Furthermore, it is important to be aware that the QMS filters can only distinguish ions by their mass-to-charge ratio, which is a problem if sample particles or molecules have the same mass but identical mass-to-charge ratio [242]. This problem can be resolved, as mentioned, by measuring fragments of molecules that form as a result of electronic bombardment [242]. For instance, if the sample molecules are N_2 and CO, which have the same mass, then they can be distinguished by analyzing their cracking patterns of N, C and O particles [242]. In this specific case, then, none of the resulting individual ions have identical mass or mass-to-charge ratio. However, other issues may occur as well, such as higher order ionizing events that result in ions being detected as having higher charge, and therefore lower mass, than they actually do. [241]

Despite its widespread use, technical advancements in quadrupole mass spectrometers (QMS) have plateaued, limiting progress in enhancing detection sensitivity and differentiating weak signals from noise [243], [244]. This stagnation poses challenges in accurately measuring catalytic activity and selectivity from small quantities of catalyst materials like single atom or nanoparticle

catalysts [245]–[248]. Addressing this, single particle catalysis aims to analyze reactions on individual nanoparticles to bypass ensemble averaging [249], yet such endeavors demand exceptionally sensitive detection methods due to the minuscule active surface areas involved. Alternative detection methods focusing on photons or electrons have been explored for their sensitivity and specificity but lack mass spectrometry's broad applicability and molecular weight determination capabilities [240], [249]–[253]. Miniaturizing reactors to micro and nanoreactor scales has shown promise in reducing necessary catalyst surface areas for measurable QMS signals by concentrating reaction products for analysis, achieving high sensitivity overall but still orders of magnitude away from enabling QMS detection of reaction products from single nanoparticles[254]–[259].

To advance beyond this state of the art, we have in Paper III used nanofluidic reactors with QMS readout and integrated them with a constrained convolutional denoising auto-encoder, merging nanofluidic reactors' efficiency with the QMS' high resolution and deep learning's signal processing capabilities. This novel approach significantly lowers the detection limit of state-of-the-art QMS by approximately three orders of magnitude, enabling the analysis of reaction products from single nanoparticles, exemplified by CO oxidation over a Pd catalyst with an active surface area as small as $0.0072 \pm 0.00086 \mu m^2$ [258], [259]. The use of machine learning in mass spectrometry, although a relatively new development, has seen rapid growth [260]–[267]. However, applying machine learning specifically to extract small QMS signals from noise, again as demonstrated in Paper III, is an entirely novel exploration.

5.2 CO Oxidation on Palladium Nanoparticle Catalysts in Nanofluidic Reactors

In Paper III, experiments were conducted on Pd nanoparticle catalysts within nanofluidic reactors using CO oxidation as model reaction [268]. The experimental sequence involved 15 min CO/O_2 mixture pulses at 6% concentration, alternating with 15 min Ar pulses, across temperatures from 280°C to 450°C. The relative CO concentration α_{CO} was varied in steps of 0.05 within the range of $0 \leq \alpha_{CO} \leq 1$ (Figure 5.2A,D,G). For a sample with $n = 1000$ nanoparticles, a clear QMS response was observed for all α_{CO} values at 450°C, with the reaction rate peaking at $\alpha_{CO}^{max} = 0.65$, indicating minimal CO poisoning and negligible mass transport gradients. A decrease in reaction temperature led to reduced reaction rates and a shift in α_{CO}^{max} towards lower values, aligning with established CO oxidation behavior on Pd catalysts [269], [270].

Reducing the active catalyst surface to n=10 and n=1 nanoparticles resulted in CO_2 reaction product pulses that were orders of magnitude smaller than for n=1000, with signals nearing the QMS noise floor at reduced temperatures. The comparison of CO_2 counts vs. α_{CO} for n=1000 and n=10 demonstrated qualitative similarity at 450°C, but a larger variance for n=10, especially at low and high α_{CO} values (Figure 5.2B,E,H). At 280°C, for n=10, the signal

approached the noise floor, indicating the challenge of detecting CO_2 from the reaction at lower particle numbers.

To address signal detection amidst noise, like in the present case, various denoising strategies like digital filtering, Fourier and wavelet transforms, and statistical methods such as PCA were considered. However, these methods have significant limitations, such as the tendency to remove frequency bands of the underlying signals or fail under certain signal/noise assumptions. Consequently, a Denoising Auto-Encoder (DAE), an artificial neural network, was employed to differentiate between complex noisy data and the underlying signal, trained using examples of simulated signals with experimentally measured noise. The DAE's ability to reconstruct the original signal, even with a Signal-to-Noise Ratio (SNR) below 1, was enhanced by constraining the latent space to a step function distribution, ensuring robust representations of the underlying signal and ultimately allowed us to detect the QMS signal from a single Pd nanoparticle in Paper III [271].

5.2.1 Autoencoders for Denoising

Deep auto-encoders are designed around a basic principle that involves a downsampling encoder network coupled with an upsampling decoder network, without any direct connections between the two. The primary role of the encoder is to condense the input data into a more compact representation within a dimensional space known as the "latent space." This space preserves only the most critical features of the data. Subsequently, the decoder utilizes this condensed representation to reconstruct the original input. Whereas traditional auto-encoders strive for an exact replication of the input data, denoising auto-encoders (DAEs) are engineered to recover the significant elements of the input while eliminating noise or other contamination.

DAEs particularly shine in scenarios where the signal and noise are distinctly separable and not intertwined. Such situations often arise in environments where underlying signals are obscured by higher-frequency noise. Neural networks, with their fundamental function as piecewise linear approximators, are inherently more adept at identifying and modeling lower-frequency patterns than higher-frequency ones, rendering them well-suited for most denoising tasks. Specifically, DAEs demonstrate optimal performance when the higher-level representations of a signal remain stable despite data corruption, i.e., when there's no correlation between noise and signal [272].

While understanding the precise mechanics of neural networks can be difficult, adopting a probabilistic perspective offers a more accessible interpretation of DAEs. By learning to invert the process of corruption - effectively deducing the original signal from its noisy counterpart - DAEs can be understood as algorithms trained to reconstruct the underlying patterns or distributions within the data. Table 5.1 captures this concept, illustrating how DAEs are trained to approximate the original data distribution by alternating between sampling from the corrupted and the estimated clean data. This approach suggests that DAEs serve as a potent means to delve into the data-generation

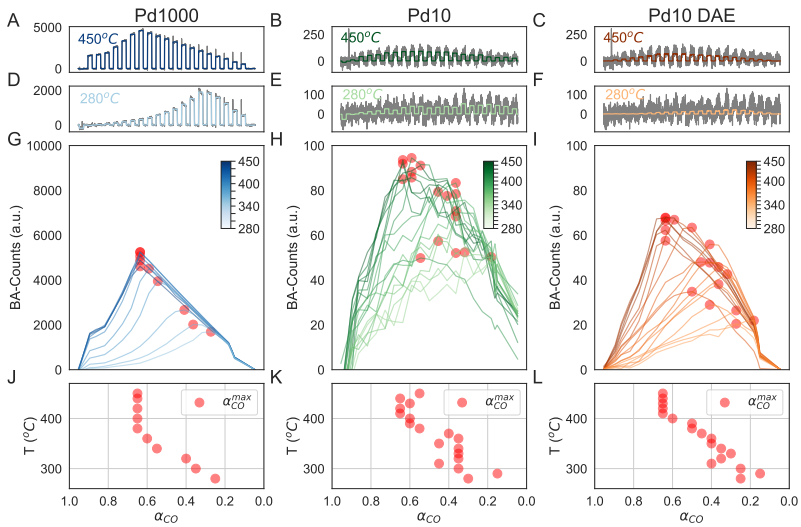

Figure 5.2: *Direct comparison between standard and DAE-enhanced QMS readouts for CO oxidation on 1000 and 10 Pd nanoparticles in a nanofluidic reactor. (A) Baseline-adjusted (BA) raw QMS counts for CO_2 (grey line) alongside the average CO_2 signal (blue line - calculated as the mean BA-count per pulse) for $n = 1000$ Pd nanoparticles over the full α_{CO} range, $\alpha_{CO} \in (0, 1)$, at 450° C. (B) BA-QMS counts for CO_2 (grey line) and the average CO_2 signal (green line) for $n = 10$ Pd nanoparticles throughout the α_{CO} range and at 450 °C. (C) BA-QMS counts for CO_2 (grey line) with the CO_2 signal refined by the DAE (orange line) for $n = 10$ over the complete α_{CO} range, $\alpha_{CO} \in (0, 1)$, and at 450° C. (D-F) Identical to (A-C) but conducted at 280° C. (G) BA-average CO_2 counts for α_{CO} sweeps at reactor temperatures from 280° C to 450° C in 20° C increments for $n = 1000$. Red dots highlight the α_{CO} value at the peak reaction rate, α_{CO}^{max}. (H) Analogous to (G) but for $n = 10$, with 10° C temperature intervals. (I) Similar to (H), but with the QMS signal refined by the DAE. (J) α_{CO}^{max} values derived from (G) for all tested temperatures for $n = 1000$, displayed as a function of α_{CO}. A consistent stoichiometric α_{CO}^{max} value of 0.65 is observed down to $T = 360$° C. Below this temperature, a systematic shift to lower α_{CO}^{max} values is noted, attributed to increased CO poisoning. (K) Comparable to (J) but for $n = 10$. A similar trend as seen with $n = 1000$ is observed, albeit with a greater variance in data points. (L) Equivalent to (K) but using the DAE-refined QMS signal from (I). The uncertainty in α_{CO}^{max} values is notably reduced, and the temperature-dependent trend of α_{CO}^{max} for both $n = 1000$ and $n = 10$ now closely aligns.*

process itself. By corrupting the random variable X (signal) into $\bar{X}$ through the distribution $C(\bar{X}|X)$, DAEs essentially aim to estimate the reverse conditional distribution $P(X|\bar{X})$ [272]. In [272], it is demonstrated that integrating this reverse distribution estimator $P(X|\bar{X})$ with the known corruption process $C(\bar{X}|X)$ enables a straightforward Markov chain, alternating between sampling from $P(X|\bar{X})$ and $C(\bar{X}|X)$ (analogous to encoder/decoder operations), to discover the intrinsic signal distribution $P(X)$. Consequently, DAE networks are interpreted to approximate the underlying data-generating process, with a simple Markov chain that toggles between the denoising model and corrupting distribution converging to this estimator.

Algorithm 1 THE GENERALIZED DENOISING AUTO-ENCODER TRAINING ALGORITHM requires a training set or training distribution $\mathcal{D}$ of examples X, a given corruption process $\mathcal{C}(\tilde{X} \mid X)$ from which one can sample, and with which one trains a conditional distribution $P_\theta(X \mid \tilde{X})$ from which one can sample.
repeat - sample training example $X \sim \mathcal{D}$ - sample corrupted input $\tilde{X} \sim \mathcal{C}(\tilde{X} \mid X)$ - use $(X, \tilde{X})$ as an additional training example towards minimizing the - expected value of log $P_\theta(X \mid \tilde{X})$, e.g., by a gradient step with respect to θ through the negative log likelihood loss, until convergence of training.

Table 5.1: *The Generalized Denoising Auto-encoder Training Algorithm.*

5.2.2 A Constrained Denoising Auto-Encoder to improve the QMS Limit of Detection

The DAE processes the experimentally obtained QMS signal, which includes intrinsic measurement noise, potential noise from the nanoreactor setup (such as slight CO_2 concentration fluctuations due to gas impurities or minor leaks), and the actual CO_2 signal from the catalytic reaction on the Pd nanoparticle(s). This aggregate signal is compressed by an encoder E_θ into a latent space at the bottleneck, where a consistency loss constrains it to adopt a step function distribution. Subsequently, this compressed representation is expanded by a decoder D_θ to reconstruct the underlying signal, effectively separating the true CO_2 signal from the observed noise. Through this architecture, the DAE is adept at learning complex non-linear relations between the noise and the underlying signal, enabling it to manage the types of noise-signal interferences that may challenge traditional data analysis or noise reduction methods. The model used in Paper III is constructed to process inputs matching the entire duration of a single α_{CO} sweep. The encoder portion is built from seven convolutional layers, with each layer comprising 32 neurons and employing a kernel size of 9. Max-pooling steps are interspersed between these layers to ensure efficient data compression. The Leaky Rectified Linear Unit

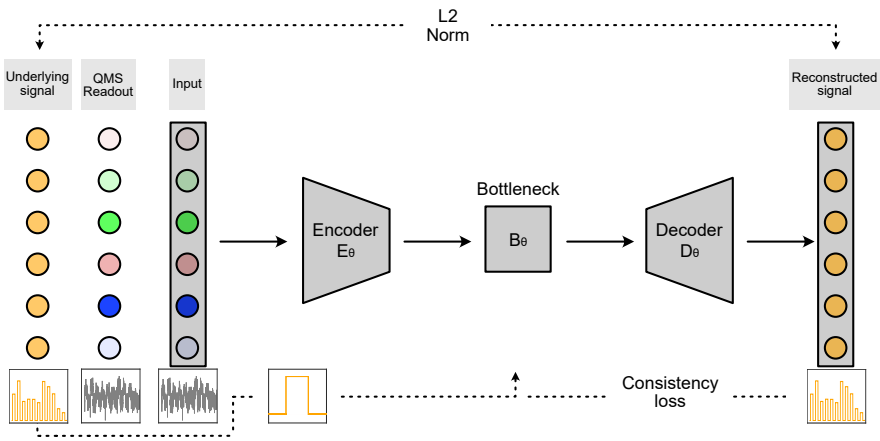

Figure 5.3: *The constrained denoising auto-encoder. We aim to accurately capture the CO_2 production rate's step functions over time, a task complicated by the presence of both correlated and uncorrelated noise in the QMS readings. To mitigate this, a denoising autoencoder is deployed, taking the QMS data as input and distilling it down to a core representation of the signal, followed by its reconstruction without the noise. This approach focuses on restoring the original signal, rather than the noise-embedded QMS output. For training, the denoising autoencoder uses a standard L2 norm to compare the reconstructed signal to the actual underlying signal, while an L1 norm and a consistency loss are applied to the latent space to ensure it accurately represents the signal's structure.*

(LeakyReLU) activation function was used for its ability to manage complex non-linear patterns while preventing issues such as mode collapse and the vanishing or exploding gradient problems. Following these layers, the data is condensed into a latent space representation via a dense layer, which scales down its dimensionality to correspond with the α_{CO} sweep steps. This compact form is then reshaped and merged with the encoder's output, maintaining both the quantitative and structural details of the signal for the decoder. The decoder, mirroring the encoder, reconstructs the signal using seven convolutional layers followed by upsampling operations to ensure precise data recovery. The output is then a reconstructed version of the initial input sans noise. A consistency loss function was also integrated to improve model precision, comparing average values across certain segments of the reconstructed and latent space data to ensure structural alignment with the original input. This structure manages to reduce noise, preserve features, and maintain data consistency.

5.2.3 Synthetic Data Generation

The framework employs two generators: one for the base signal and another for noise. Noise generation precedes signal generation and is characterized by its standard deviation and the morphology and timing of the gas pulses in the experiment to be simulated. Initially, it simulates common stochastic variations observed in QMS readings, such as thermal noise, environmental changes, flicker noise, shot noise, and ion feedback noise, using a Gaussian distribution to cover a wide range of white noise effects.

The signal generator produces a step function with a selectable signal-to-noise ratio $SNR = \frac{\mu_{signal}}{\sigma_{noise}}$, set during training to align with the SNR observed in 1-10 Pd nanoparticle reactor outputs given the specified noise profile. The steps in the function maintain a consistent on-off pattern, though each step carries a randomly assigned signal magnitude, avoiding any biased assumption about the relationship between consecutive pulses within the same function. Some representative samples of this training data are shown in Figure 5.4.

5.2.4 Deep Learning Training

The training process happens in two stages. Initially, a curriculum learning strategy is applied to signals artificially generated atop Gaussian-distributed noise with a pre-set SNR, chosen from a uniform distribution between $(0.95, 1)$ to reflect the peak catalytic activity measured in the $n = 10$ Pd sample at $T = 450°\text{C}$. This SNR range gradually decreases by 0.05 down to a minimum of 0 with each training cycle, facilitated by an early stopping criteria set at 32 epochs. This approach mitigates the risk of vanishing gradients and mode collapse by avoiding direct training on data with extremely low SNRs. Generating data across a spectrum of noises and signal patterns enhances the model's robustness and ability to learn denoising functions that generalize well.

Secondly, all subsequent training on artificial signal patterns within an $SNR \in (0, 1)$ range, spanning the entire spectrum of SNRs considered in this study,

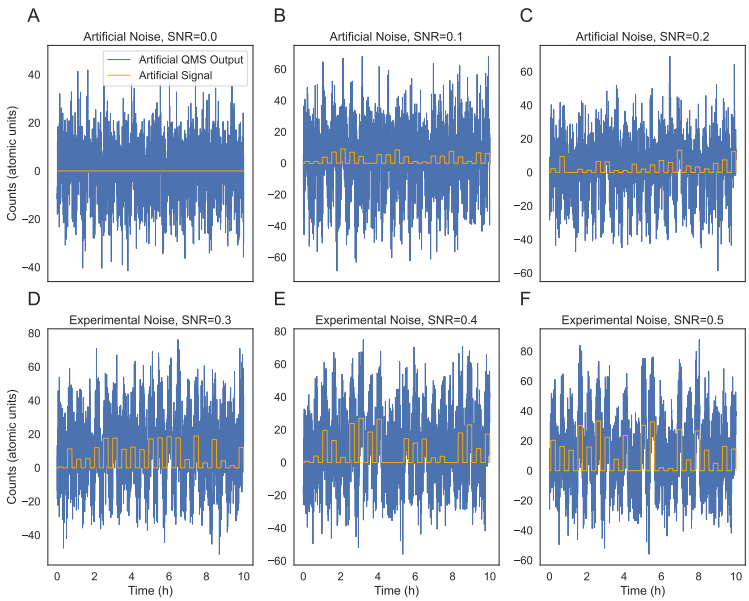

Figure 5.4: *Representative Training Data. A selected sample of the training data, where panels (A-C) feature artificial QMS outputs for* $SNR \in [0, 0.1, 0.2]$*, achieved by overlaying artificially generated signals atop artificially produced noise. Similarly, panels (D-F) present artificial QMS outputs for* $SNR \in [0.3, 0.4, 0.5]$*, constructed by combining artificially generated signals with experimentally recorded noise from a typical CO oxidation sequence conducted at* $T = 40°C$*, a temperature at which (almost) no reaction is expected to occur. It's important to note that the SNR, along with the characteristics of both noise and signal, are varied randomly throughout the training process.*

is overlayed on experimentally measured noise profiles. This refinement step adapts the network to effectively handle noise distributions specific to the experimental settings during inference. Here, the noise model represents the output from a CO oxidation experiment using a $n = 0$ Pd sample chip at $T = 40°\text{C}$.

In machine learning models, the loss function measures the discrepancy between the predicted outputs and actual values and shapes whose functions the models learn to optimize towards. Adopting the framework of a generalized denoising autoencoder as described in [272] and shown in Table 5.1, our primary loss function should be the negative log-likelihood (NLL).

In the context of regression tasks, NLL loss often becomes relevant under certain probabilistic models of noise or prediction errors. The adoption of NLL in regression is typically predicated on the assumption that the residuals (the differences between predicted and actual values) are distributed according to a Gaussian (or normal) distribution. This leads us to the Mean Squared Error (MSE) loss, defined as:

$$\text{MSE}(\theta) = \frac{1}{N}\sum_{i=1}^{N}(y_i - f_{D\theta}(x_i;\theta))^2 \tag{5.5}$$

where y_i denotes the actual value for the i-th data point, $f_{D\theta}(x_i;\theta)$ is the prediction of the decoder for the i-th data point, parameterized by θ, and N represents the total number of data points.

Given the Gaussian noise premise, MSE emerges naturally as the NLL. To see this, the likelihood of a single observation is given by the Gaussian probability density function:

$$\mathcal{L}(y_i|x_i;\theta) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(y_i - f_{D\theta}(x_i;\theta))^2}{2\sigma^2}\right) \tag{5.6}$$

Taking the negative logarithm and disregarding constants not depending on θ (as they do not influence optimization), we obtain a cost that is proportional to MSE [273], also known as the L2 or L^2 norm.

Moreover, to incorporate the consistency loss within the autoencoder's latent space, we apply an MAE (Mean Absolute Error, or L1 norm) directly between a compressed vector representation of the latent space and the actual values of each CO_2 pulse from the synthetic signal.

$$\text{MAE}(\theta) = \frac{1}{n}\sum_{i=1}^{n}|y_i - f_{B\theta}(x_i;\theta)|, \tag{5.7}$$

where y_i represents the actual value for the i-th data point, $f_{B\theta}(x_i;\theta)$ is the representation of the autoencoder's latent space for the i-th data point, parameterized by θ, and n is the number of CO_2 pulses.

Subsequently, we calculate the L1 norm between these estimates and the mean predicted output for each CO_2 pulse, promoting consistency between the decoder's ultimate predicted output and its corresponding representation in the latent space of the autoencoder.

This loss is formalized as:

$$C_{\text{loss}}(\theta) = \text{MAE}\left(f_{B\theta}(x_i;\theta), \mu\left(f_{D\theta}(x_i;\theta)_{\text{pulse}[i]}\right)\right) \tag{5.8}$$

Here, $\mu\left(f_{D\theta}(x_i;\theta)_{\text{pulse}[i]}\right)$ denotes the average output of the decoder across the i-th pulse.

5.2.5 Resolving CO_2 Reaction Product formed on a single Pd Nanoparticle using the DAE

Having verified the Denoising Auto-Encoder (DAE)'s capacity to effectively reduce noise in the CO_2 QMS signal, we proceeded to test its performance on an even smaller scale: a single Pd nanoparticle with an approximate surface area of 17000 nm^2. This experiment involved a nanofluidic chip with a single nanoparticle ($n = 1$) and an empty chip as a control ($n = 0$). At 450°C, the α_{CO} sweeps for both $n = 0$ and $n = 1$ exhibited similar small, stochastic CO_2 pulses in the standard analysis, suggesting background CO_2 levels in the reactant gas mixture and indicating the challenge of discerning CO_2 produced by the single nanoparticle from QMS noise.

Applying the DAE to the $n = 1$ data at 450°C, a clear trend of CO_2 counts was observed along the α_{CO} sweep, peaking at a stoichiometric $\alpha_{CO}^{max} = 0.65$, aligning with results from larger sample sizes ($n = 10$ and $n = 1000$) (Figure 5.5D). For the $n = 0$ control, the DAE processed data displayed nearly flat baselines at zero counts, demonstrating the DAE's ability to filter out noise (Figure 5.5E).

Extending the DAE application to $n = 1$ across temperatures from 330°C to 450°C, we observed the expected reaction rate increase and subsequent decrease along the α_{CO} sweep, with CO_2 counts around 6, roughly ten times lower than for $n = 10$ at 450°C, consistent with the reduction in active surface area (Figure 5.5F). A shift in α_{CO}^{max} to lower values was noted at 410°C, indicating temperature-dependent reaction behavior before dropping below the single count detection limit. Analysis of the $n = 0$ control consistently showed counts near or below the detection limit, validating the DAE's resolution capabilities (Figure 5.5F).

In comparison with $n = 10$ and $n = 1000$ data, α_{CO}^{max} values for the DAE-denoised data across $400 \leq T \leq 450$°C were similar for all samples, closely matching the stoichiometric value of $\alpha_{CO}^{max} = 0.66$,demonstrating consistent reaction conditions across different particle numbers (Figure 5.5G). Additionally, CO_2 counts at α_{CO}^{max}, normalized by the number of particles, indicated direct scaling of QMS counts with catalyst surface area, within expected uncertainty

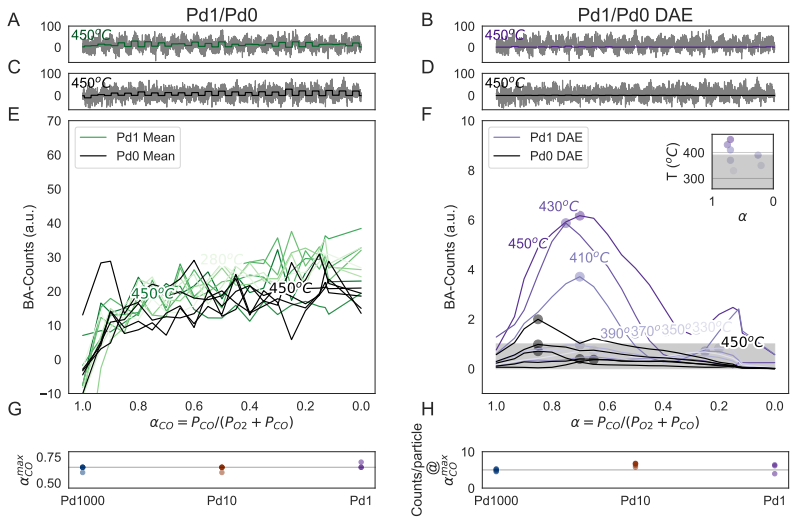

Figure 5.5: *Online Mass Spectrometry from a single Pd Nanoparticle. (A) QMS readings for CO_2 (grey line) combined with the average CO_2 signal (green line), calculated as the mean count for each pulse from a single ($n = 1$) Pd nanoparticle, over the full α_{CO} spectrum, where $\alpha_{CO} \in (0, 1)$, at $450°C$. (B) Identical to (A), but for an empty nanochannel, denoted $n = 0$. Notably, both $n = 1$ and $n = 0$ display a similar stochastic pattern in CO_2 pulse appearance under conventional analysis. (C) Average BA-CO_2 counts per α_{CO} pulse, derived using standard analysis across the entire α_{CO} range, $\alpha_{CO} \in (0, 1)$, and at $450°C$, for both $n = 0$ (black lines) and $n = 1$ (green lines). The procedure involved 7 consecutive α_{CO} sweeps for $n = 1$, and 5 for $n = 0$. A notable finding is the consistent reproducibility of the α_{CO} sweeps for both conditions and the lack of a distinct activity trend based on α_{CO} for the single nanoparticle sample, despite typically showing slightly higher counts than the empty control. (D) Similar to (A), but showcasing DAE-denoised BA-QMS CO_2 readings (purple line), where the DAE reveals significantly smaller, yet more consistent CO_2 pulses compared to standard analysis. (E) As in (D), but for an empty nanochannel ($n = 0$). The DAE analysis yields a consistently flat baseline at zero counts for the empty channel. (F) Analogous to (C), but applying DAE-denoised data, where discernible reaction rate peaks are identified for $n = 1$ (purple curves), with α_{CO}^{max} observed between 0.65 and 0.6 across temperatures from $450°C$ down to $410°C$ (inset), aligning well with prior findings for $n = 10$ and $n = 1000$. At reduced temperatures, and for $n = 0$ (black lines, indicative of several sweeps at $450°C$), the DAE reports counts < 1, deemed physically implausible, thus establishing the detection threshold. (G) α_{CO}^{max} for $n = 1000$ (blue), $n = 10$ with DAE-denoising (orange), and $n = 1$ with DAE-denoising (purple) across the temperature window of $400 \leq T \leq 450°C$, the range wherein the $n = 1$ DAE-signal surpasses the detection floor. (H) Particle-normalized CO_2 counts at α_{CO}^{max} for $n = 1000$ (blue), $n = 10$ post-DAE (orange), and $n = 1$ post-DAE (purple) within the temperature range of $400 \leq T \leq 450°C$.*

due to slight variations in nanoparticle sizes stemming from the fabrication process of the fluidic chips, since different such chips were used for $n = 0$, $n = 1$, $n = 10$ and $n = 1000$, respectively (Figure 5.5H).

Chapter 6

Summary of Included Papers

6.1 Label-free nanofluidic scattering microscopy of size and mass of single diffusing biomolecules and biological nanoparticles

This study presents Nanofluidic Scattering Microscopy (NSM) as a label-free technique designed for real-time imaging and analysis of single biomolecules and biological nanoparticles (BNPs) in solution. NSM addresses the limitations of traditional fluorescence microscopy, which often requires fluorescent labeling that can alter the properties of the target molecules. By confining biomolecules within a nanofluidic channel, NSM significantly enhances their optical contrast, enabling the accurate determination of their molecular weight (MW) and hydrodynamic radius (R_s) through scattering analysis.

NSM's capabilities were demonstrated using a range of biomolecules, including proteins, DNA, and extracellular vesicles (EVs). The technique was able to identify distinct molecular populations, such as monomers and dimers, with high precision. For example, the study identified protein dimers of thyroglobulin and ferritin with peaks at approximately double the expected MW and about 1.3 times the expected R_s, highlighting the sensitivity of NSM in detecting molecular assemblies. Moreover, when applied to complex biological samples like cell culture media, NSM successfully differentiated between lipoprotein particles and EVs based on their size and scattering properties, underscoring its versatility in analyzing mixed samples.

A crucial advancement in this study was the integration of deep learning (DL) methods into the NSM data analysis pipeline. By employing machine learning algorithms, particularly for interpreting the light-scattering patterns from biomolecules, the study achieved significant improvements in the speed, accur-

acy, and precision of MW and R_s measurements. The DL-enhanced analysis was particularly effective for smaller biomolecules, which are challenging to characterize using conventional techniques. This approach not only improved the resolution of NSM but also facilitated the real-time tracking and characterization of biomolecular dynamics, aggregation states, and conformational changes.

The results demonstrate NSM's potential to transform biological research by enabling the label-free characterization of biomolecules, monitoring protein interactions, and analyzing cellular secretomes with unprecedented detail. The combination of NSM and DL opens up new possibilities for studying protein dynamics, enzyme functions, and the behavior of therapeutic agents in biological contexts, making it a valuable tool for molecular diagnostics and drug discovery

6.2 Neural network enabled nanoplasmonic hydrogen sensors with 100 ppm limit of detection in humid air

In this paper, we focused on addressing a critical challenge in the field of hydrogen safety sensors, particularly their performance under high humidity conditions. Recognizing the expanding role of hydrogen technologies and the inherent risks due to hydrogen's high flammability, our work centered on developing an optical nanoplasmonic hydrogen sensor that functions effectively in environments with elevated humidity levels.

The cornerstone of the study was the design and testing of a sensor based on Pd70Au30 alloy nanoparticles. This choice was motivated by the alloy's proven sensing capabilities under dry conditions and its ability to mitigate the intrinsic hysteresis found in pure Pd. The sensor's performance was evaluated in varying humidity conditions, revealing a significant challenge: at ambient sensor operation temperatures, even 20% relative humidity substantially impaired the sensor's performance, with complete deactivation observed at 80% relative humidity.

To overcome this limitation, we first explored the effect of increasing the sensor's operating temperature. This approach was predicated on shifting the surface equilibrium coverage of molecular species present on the sensor's surface in humid conditions, favoring hydrogen over water. This strategy enabled efficient hydrogen dissociation and absorption into the nanoparticles, restoring robust sensor response even at 80% relative humidity. However, a notable drawback of this method was the reduced absolute amplitude of sensor response to specific hydrogen concentrations, a consequence of the temperature-dependent solubility of hydrogen in Pd and its alloys.

To further enhance the sensor's capabilities, I incorporated deep learning techniques, employing a Deep Dense Neural Network (DDNN) and a Transformer architecture. This methodology allowed for a more comprehensive analysis of the sensor's spectral data, significantly improving the limit of detection (LoD). The deep learning-based approach achieved LoDs between 100 – 600 ppm H2 in 80% RH in synthetic air for operating temperatures of 80 – 130 °C. Consequently, the sensor not only met but exceeded the U.S. Department of Energy's performance target for LoD in high humidity conditions.

Additionally, we conducted a long-term stability test of the sensor operating in 80% RH, which showed no signs of degradation over a week, maintaining a directly measured LoD of 0.06% H_2. This test employed a Transformer architecture for its ability to handle complex time-series analysis. One critical insight is that sensor performance, when based on deep learning models like the Transformer, can deteriorate if the sensing environment varies significantly from the conditions used in the training data. Addressing this, I demonstrated that re-training the model with data inclusive of these new conditions significantly improved performance, achieving a record LoD of 100 ppm H_2 at 80% RH.

This study contributes a novel, highly effective solution for hydrogen detection in high humidity environments, addressing a significant challenge in the deployment of hydrogen technologies. This work not only advances the field of hydrogen safety sensors but also has broader implications for ensuring safety in the expanding realm of hydrogen-based applications.

6.3 Deep-learning enabled mass spectrometry of the reaction product from a single catalyst nanoparticle

The key innovation in this research was the combination of nanofluidic reactors with deep learning, particularly a constrained denoising autoencoder (DAE), to enhance the resolution in mass spectrometry for analyzing small amounts of catalytic reaction products in the gas phase. This approach allowed for a substantial reduction in the required catalyst surface area for online mass spectrometric analysis, down to the level of a single nanoparticle.

Unlike traditional auto-encoders that aim to replicate input data precisely, DAEs focus on restoring only the relevant parts of the input by filtering out noise and irrelevant information. This approach was particularly effective in our context, where detecting weak signals amidst high levels of noise was a significant challenge. To demonstrate the superiority of the DAE, we benchmarked its performance against established denoising techniques like Fourier transform-based filtering, Principal Component Analysis (PCA), Total Variation (TV) denoising, and Wavelet denoising. The DAE consistently showed superior noise mitigation while maintaining the fidelity of the original signal.

This approach significantly lowered the detection limit of state-of-the-art QMS by approximately three orders of magnitude, enabling the analysis of reaction products from single nanoparticles, exemplified by CO oxidation over a Pd catalyst with an active surface area as small as $0.0072 \pm 0.00086 \mu m^2$ [258], [259]. The use of machine learning in mass spectrometry, although a relatively new development, has seen rapid growth [260]–[267]. However, applying machine learning specifically to extract small QMS signals from noise is an entirely novel exploration.

This research establishes a new method for online reaction analysis in single particle catalysis, applicable when the quantification of weak signals in overwhelming noise is of interest. The combination of nanofluidic reactors and deep learning, particularly through the use of a DAE, significantly improves the resolution in mass spectrometry, enabling the detection and analysis of reaction products from catalytic processes occurring on single nanoparticles. This advancement not only contributes to the field of catalysis but also demonstrates the potential of deep learning to enhance resolution in analytical techniques across various scientific disciplines.

6.4 Accelerating Plasmonic Hydrogen Sensors in Inert Gas Environments by Transformer-Based Deep Learning

The research focuses in this manuscript on leveraging a state-of-the-art optical plasmonic hydrogen sensor, specifically a Pd70Au30 alloy system, integrated with a deep learning Transformer model we coined Long Short-term Transformer Ensemble Model for Accelerated Sensing (LEMAS). The primary challenge addressed in this study is the slow response time of existing hydrogen sensors in technically relevant conditions, particularly in oxygen-free, inert gas environments. The Pd70Au30 alloy plasmonic hydrogen sensor was chosen due to its suitability in such environments and its ability to eliminate the intrinsic hysteresis characteristic for pure Pd, thus offering a linear optical response to hydrogen concentration changes. The sensing mechanism is based on the localized surface plasmon resonance (LSPR) phenomenon, where hydrogen absorption into the nanoparticle lattice shifts the LSPR peak in the visible light spectrum.

LEMAS accelerates the sensor's response by up to 40 times in scenarios simulating both large and small hydrogen leaks, exploiting the capabilities of deep learning to predict the hydrogen concentration change before the physical sensor's response reaches saturation. This is achieved by learning the relationship between the full spectral response over time and the hydrogen concentration. LEMAS not only accelerates response time but also provides uncertainty estimates, which are crucial for safety-critical applications.

The development of LEMAS involved training and testing on experimental data generated from various hydrogen concentration profiles in an inert gas environment. We demonstrated its ability to rapidly discern and quantify slow, gradual changes in hydrogen concentration, distinguishing these from noise. This ability is critical for detecting hydrogen at as low concentrations as possible and as quickly as possible, allowing sufficient time for implementing safety measures. The results from our study show that LEMAS can significantly accelerate the response time of the plasmonic hydrogen sensor, particularly in the low concentration regime. It has proven effective in scenarios of both large, fast leaks and small, slow leaks. For the smallest H_2 concentrations of 0.06%, the response times ranged between 1.6 s to 3.6 s, compared to 50 s to 85 s for standard analysis methods, marking a 21 to 40-fold improvement.

This research underscores the potential of deep learning to overcome current limitations in hydrogen sensor performance, such as slow response times in technologically relevant sensing environments like inert gas atmospheres. The integration of LEMAS with plasmonic hydrogen sensors represents a significant advancement in the field, providing a faster, more reliable detection method that is critical for the safe implementation of hydrogen technologies.

6.5 Cross modality transforms in biological microscopy enabled by deep learning

The review explores the transformation of microscopy images across different optical modalities through deep learning, advancing multi-modal imaging analysis. It provides a comprehensive overview of cross-modality transformations (CMTs), reviewing applications, datasets, and implementations. The focus is on two primary applications: enhancing image contrast and enhancing resolution. The paper highlights the significant benefit of CMTs for biologists, enabling high-contrast and specificity imaging without chemical staining and achieving super-resolution beyond the limits of diffraction.

Deep learning is central to CMTs, where the networks are trained for image translation tasks using encoder-decoder-style fully convolutional neural networks, such as U-Net, ResNet, and InceptionNet. Loss functions, including adversarial loss, play a crucial role in training, with Generative Adversarial Networks (GANs) being particularly effective. Diffusion models, as an alternative to GANs, offer high-quality images but are computationally expensive. The review discusses the application of CMTs enabled by these neural networks in tissue imaging, histology, and cellular imaging. Histological staining, vital in clinical pathology, encounters challenges with manual chemical staining, which deep learning models can virtually substitute. Techniques like virtual staining networks transform unstained samples into stained images, maximizing information output and easing experimental setups. The use of GANs in cellular staining enables the replacement of chemical and fluorescent staining with virtual fluorescence-stained images. Applications include various imaging modalities like phase-contrast, holographic microscopy and brightfield images.

In molecular imaging, super-resolution (SR) microscopy has seen significant advancements through DL. The review also surveys the physics of SR and details DL techniques for transforming images into super-resolved counterparts, focusing specifically on molecular imaging. Various deep learning approaches for SR are discussed, including end-to-end low-resolution to high-resolution mapping, iterative up- and down-sampling, and sequential upsampling. Prolific architectures like ANNA-PALM, Deep-STORM, and smNet are also highlighted. The review also addresses challenges in deep learning for microscopy, such as the need for large training datasets, overfitting, and the importance of transfer learning. Regularization techniques, data augmentation, and model evaluation are crucial for robust and reliable models. The potential of deep learning in cross-modal transformations is underscored, offering transformative advancements in disease research, drug development, neuroscience, and personalized medicine. The integration of deep learning into biological microscopy offers non-invasive, label-free alternatives, impacting various fields and empowering broader research explorations.

6.6 Deep Learning Microscopy for Label-free Biomolecule Weight-and-Size Characterization in the Single-kDa Regime

In this study, we developed a method that combines Nanofluidic Scattering Microscopy (NSM) with Deep Learning (DL) algorithms in the form of a Hierarchical Vision Transformer (HViT) to enable the label-free characterization of biomolecules, with a particular focus on those in the sub-10-kDa range, which is experimentally extremely challenging. This approach addresses the limitations of traditional optical microscopy techniques, which often require fluorescent labeling or surface binding, by facilitating real-time, label-free analysis of small biomolecules.

We enhanced the capability of NSM to accurately determine the molecular weight and hydrodynamic radius of single biomolecules. By applying the HViT to interpret complex light scattering patterns from nanofluidic channels, we achieved detailed study capabilities for smaller biomolecules, including specific proteins and peptides, which were previously challenging to analyze due to their size and behavior in solution.

The methodology not only extends the capabilities of label-free biomolecule characterization but also contributes to biological research, potentially facilitating new discoveries in areas such as protein dynamics, enzyme functions, and therapeutic agent development. Additionally, our approach addresses the limitations of spatial and temporal resolution and the challenges of the ill-posed nature of microscopy image transformation tasks encountered with heuristic-based methods in optical microscopy. Through DL, our method effectively denoises low-frequency trends obscured by high-frequency noise, thereby enabling a more precise characterization of small biomolecules.

This work signifies a contribution to the fields of biological sciences and medical research, offering a novel method for the analysis of diseases, including cancer and diabetes, through the study of binding events and environmental effects on biomolecules.

Chapter 7

Discussion and Future Work

7.1 DL for Nanofluidic Scattering Microscopy

NSM has recently been introduced as a technique that enables the label-free real-time characterization of freely diffusing single biomolecules down to the tens-of-kDa regime inside nano-sized channels. The focus of near-term projects in this system can be split in two; firstly, applying the deep learning methods developed in this thesis to analyze single molecule trajectories to a high degree of accuracy, precision and detection limit, and secondly to develop new deep learning methods to enable analysis beyond the single trajectory regime that, e.g., would enable studies of mixed samples. All the examples brought up in the sections below are already being worked on or are in early planning stages, through collaboration with a broad range of research groups and institutions.

7.1.1 Single Molecule Analysis

All projects within this sub-section share one thing in common; they can in principle be solved by the deep learning tools introduced in this thesis as-is, including the high precision, accuracy and detection limits said tools entail. The only required modification is to train the models in the corresponding parameter range of noise and generated optical response that each inividual project requires.

7.1.1.1 Insulin

Following the theoretical limits of detection and quantification defined in Paper VI, accurate characterization of single molecule trajectories below the 10 kDa regime is possible. Since the iOC of molecules scales linearly with decreasing nano-channel cross section, smaller channels may also be nanofabricated to

improve the accuracy of said characterization. Thus, we should expect to be able to accurately characterize even the protein-hormone insulin with $MW \approx 6$ kDa, particularly important for diabetes research.

7.1.1.2 λ-DNA

Over the past two decades, nanofluidic structures have become a potent platform for in-depth investigation of DNA on the kilo-bp length scale [274]. The combination of excluded volume and DNA stiffness causes DNA to be stretched to almost its full length when contained in nano-sized channels [274]. NSM is therefore an ideal complementary method to study these long DNA structures, and preliminary results on comparing fluorescently stained λ-DNA to unstained λ-DNA show that the staining process indeed affects the length and optical contrast of said molecules. Therefore, to study their properties without significantly altering them, label-free methods such as NSM are crucial. In this project, all deep learning models were used as described in this thesis, with the only difference being that the models were instead trained on the generated simulated response of λ-DNA in our nano-sized channels. Further measurements will corroborate these results, and lead to the natural continuation of the project in studying DNA-protein binding events as outlined in section 7.1.2.

7.1.1.3 Stress Granules

Eukaryotic cells respond to stress by forming stress granules, which are effectively transient membraneless organelles [275]. They have a core-shell substructure, are made up of messenger ribonucleoprotein complexes, and are crucial for controlling the rate of protein translation and signaling under stress [275]. This has significant implications for cancer and neurological illnesses [275]. The molecular structure of stress granules and their biochemical function in plant growth and fitness are the subjects of active research, yet no methods have so far been able to measure each granule individually in a satisfactory manner [275]. Stress granules, as high-density protein and RNA aggregates formed by organisms under stress, may be analyzed in an effectively equivalent fashion to EVs in Paper I through single molecule trajectory analysis in large channels.

7.1.1.4 Extracellular Vesicles

In Paper I, we introduce NSM as a powerful characterization technique for EVs, and we expect it to remain a key method for label-free EV analysis in the future. In the short term, we will use NSM not only to characterize EVs, but also their secreted contents by use of nanovesicle traps built directly into the nano-sized channels. Further, as noted in Paper I, precise conversion of iOC into MW is more complicated for biological nanoparticles, such as EVs, than for single molecules due to a large variety of molecular constituents with different optical properties whose representation and spatial distribution might be different for each BNP. Therefore, we did not further explore the content of the EVs via the measured iOC. Further theoretical work suggests that

correcting for such effects is possible through modelling the electromagnetic properties of EVs, which would enable more informative BNP characterization in general.

7.1.2 Beyond Single Molecule Analysis

Going beyond the analysis of isolated trajectories of single molecules or particles requires the development of new deep learning models. Unpublished results on simulated data show that node-regression-based graph neural networks (GNNs) are ideal for high-density samples which require high-fidelity output temporally and spatially through the channel [151]. In principle, GNNs retain the inherent benefit of a CNN approach in being able to consider an entire trajectory simultaneously in its prediction of its properties, without the disadvantage of having to consider an entire kymograph (or portion of a kymograph). Thus, each frame of a trajectory can be given a separate prediction of its properties, whilst still considering the entire connected trajectory in the prediction. This approach is ideal in cases where we expect molecule trajectories to change over time, as is the case for enzyme-molecule interactions, DNA-protein binding events and for investigations of conformational changes in trajectories over time, among other applications.

7.1.2.1 Protein-Protein Interactions

When performing their tasks *in vivo*, proteins hardly ever behave as separate species [276]. In fact, it is estimated that over 80% of proteins have been found to function in complexes rather than alone [276]. Studying how protein-protein interactions impact their function is crucial for determining how proteins work in their natural environment, i.e. in the inside of cells. Further, to understand the cell's biochemistry, it would be incredibly helpful to characterize the interactions between the proteins in said cell [276]. There are numerous approaches to determine the outcome of an interaction between two or more proteins that has a clear functional goal.

Cell-to-cell contacts, metabolic regulation, and control of developmental processes are only a few of the biological activities that are handled by said protein-protein interactions [277]. These interactions can be categorized in a number of ways based on their differing structural and functional traits [277]. For instance, they can be homo- or heterooligomeric depending on their interaction surface, obligate or non-obligate depending on their stability, and transient or permanent depending on their persistence [277]. Any combination of these three distinct pairs may be used to characterize a certain protein-protein interaction. For instance, while temporary contacts likely create signaling pathways, long-term interactions are more likely to result in a stable protein complex. NSM is a promising method to study such interactions, and the aforementioned GNN approach is an ideal analysis tool for such an application. Another project along the same focus is to study enzyme-molecular interactions, i.e., how an enzyme (catalyst) affects the properties of a biomolecule when bound to it, as is the case with CRISPR/Cas-9 binding events.

7.1.2.2 λ-DNA - CRISPR/Cas-9 Protein Binding Events

Proteins are lengthy chains of amino acids connected by peptide bonds and, as established in section 3.1.3, are versatile and serve a number of purposes as building blocks of the cell. Transporter proteins, receptor proteins, structural proteins, and enzymes are a few of the various protein subgroups [278]. The latter work by speeding up chemical reactions and highly selectively controlling a wide range of functions within living cells [278]. They are also crucial to the replication, repair, and alteration of DNA [278]. The Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) associated protein-9 nuclease (Cas9) is one example of an enzyme involved in DNA modification [279]. Cas9, effectivelly a gene-editing tool, can catalyze extremely sequence-specific double stranded breaks by cleaving links in the DNA backbone [279]. The enzyme has many scientific uses, such as gene targeting, control, and involvement in the adaptive immune system of bacteria and archaea [279]. The Cas9 enzyme uses an RNA molecule, known as Guide RNA (gRNA), with a 20 bp motif corresponding to the site of interest to identify the precise DNA sequence at which it will cause a double-stranded break with very high precision. Direct detection and characterization of these binding events, as principally enabled by NSM, is a highly relevant and sought-after feature in the fields of gene editing and genomic mapping [278]. To study how active each binding site is over time, GNNs are again an ideal solution.

7.1.2.3 Temperature Influence on Oligomeric State of Protease

Protein quality control is a crucial biological task primarily carried out by a wide variety of distinct proteases [280]. One of these, the *DegP* protease, is essential to the Escherichia coli protein quality control system because it eliminates unfolded proteins or stops them from aggregating [280]. Upon temperature activation, this protease is expected to go through extensive oligomeric state modifications, including massive structural transitions from its hexameric resting form via a trimeric intermediate state toward the higher oligomeric states [280]. However, despite being thoroughly studied and having a number of high-resolution crystal structures of DegP in its many states, there are still many unanswered concerns regarding how its structural transitions are actually accomplished [280]. Prior studies using NMR diffusion and Cryo-EM have been unable to elucidate these questions [280], but temperature-controlled NSM is in contrast quite promising. To study the structural transitions over time (with varying temperature), the aforementioned GNNs are ideal.

7.1.3 Long-term Future Outlook

In terms of DL advancement, recent progress in geometric DL for particle tracking and characterization in terms of GNNs represents promising complementary methods for use in high-fidelity high-density studies [281], where the CV pipeline introduced in this thesis will struggle. It has recently been independently shown [151] that reconstructing the trajectories of diffusive particles is not a necessary requisite to accurately calculate the properties of

the underlying stochastic process which gave rise to the trajectory. Machine-learned models, like the ones trained in this thesis, have the potential to exploit this hidden information to push the boundaries of particle characterization methods in the regime of low SNR where heuristic-based methods often struggle significantly. Since the models are trained entirely on simulated data, their performance is also technically limited by available computational power and are thereby expected to improve in tandem with the breakneck rates at which newer generations of parallelizable (GPU) and tensor-processing-optimized (TPU) chips operate and are developed. Therefore, we expect that beyond enabling the analysis of biomolecules in the single-digit kDa regime, these DL tools will become ever-more crucial in the complex and versatile analysis that NSM and other modern methods of optical characterization enable.

In its current form, NSM already enables highly accurate label-free and tether-free characterization of individual biomolecules and biological nanoparticles in a wide range of biofluids. Advances in terms of instrumentation and DL tools is expected to push the performance even further, and particularly we expect that high-throughput and resolution in the single-digit kDa regime will find numerous bioanalytical applications requiring analysis of highly heterogenous samples, such as exosome characterization and direct detection of small-molecule binding events [282]. Further, long-term monitoring of individual biomolecules diffusing in solution represents a yet unexplored opportunity for studies of conformational changes, aggregation processes and interactions between individual biomolecules. In addition, due to minimized dilution of the sample by the nanofluidic platform, NSM is highly efficient for transporting ultra-low volumes, such as intracellular content or secreted metabolites of a single cell, thereby paving the way to real-time label-free single-cell studies. Looking further forward, we expect as mentioned that NSM will find applications outside the field of bioanalysis, such as characterization of inorganic nanoparticles, particle counting, and single particle catalysis [142].

7.2 DL for Nanoplasmonic Hydrogen Sensing

In our recent works, we have embarked on an exploration of the intersection between artificial intelligence (AI) and nanoplasmonic sensing, particularly focusing on the realm of hydrogen sensors in inert gas environments. Our studies demonstrate the profound impact of integrating tailored AI models, specifically the Long Short-term Transformer Ensemble Model for Accelerated Sensing (LEMAS), with plasmonic hydrogen sensors. This approach significantly enhances sensor responsiveness and accuracy, enabling up to a forty-fold acceleration in response times and eradicating the intrinsic pressure dependence that has traditionally limited metal hydride-based sensors. Such advancements are crucial for applications in environments devoid of oxygen, where rapid and reliable hydrogen detection is essential for safety.

Looking ahead, the implications of our research extend far beyond hydrogen sensing. The principles underlying the AI-enhanced methodologies we've

developed can be adapted and applied across a wide spectrum of chemical and environmental sensing tasks. The core achievements—markedly improved sensor response times and operational efficacy under technologically relevant conditions—pave the way for more secure and efficient hydrogen energy technologies and related applications.

Furthermore, our work lays a foundational blueprint for future research aimed at leveraging AI to push the boundaries of traditional sensor technology. By demonstrating the versatility of LEMAS and similar AI models across different sensing platforms and environments, we highlight the potential for a transformative impact on gas sensing and beyond. The integration of AI-driven enhancements into sensors for a broad range of applications—from environmental monitoring and industrial safety to healthcare diagnostics—holds promise for significant breakthroughs.

Our future research endeavors will likely concentrate on refining AI models to achieve even quicker response times, enhancing the robustness and dependability of sensors across an expanded range of environmental conditions, and broadening the spectrum of detectable substances. The synergy between materials science, chemistry, and AI research will be instrumental in driving these advancements, as will the development of sophisticated, miniaturized sensor platforms capable of seamless integration with AI algorithms.

Our work in combining AI with nanoplasmonic sensing marks a significant stride toward overcoming longstanding limitations in sensor technology that we postulate cannot be overcome by traditional means of sensing material or transduction mechanism engineering alone. As we look forward, the continued evolution of this interdisciplinary approach promises not only to enhance the safety and efficacy of hydrogen energy systems but also to spur innovations across diverse domains where precise and rapid detection of chemical species is of paramount importance.

7.3 DL for Single Particle Catalysis

In our study, we introduce an innovative methodology that employs nanofluidic reactors in conjunction with deep learning to significantly advance the ability to study the reaction output from single nanoparticle catalysts. Specifically, by leveraging a constrained denoising autoencoder, we enable the analysis of gas-phase reaction products via mass spectrometry at the nanoscale. This technique focuses on catalyst surface areas as diminutive as individual nanoparticles, approximately 60-70 nm in diameter. The use of nanofluidic reactors concentrates reaction products, while the autoencoder isolates relevant signal data from noise. This approach is particularly applied to the CO oxidation on Pd nanoparticles with an area of $\approx 0.0072 \pm 0.00086 \mu m^2$. Our experiments, conducted over a temperature range from 280 °C to 450 °C and with varying numbers of Pd particles, have delineated a clear dependence of reaction rates on the $CO{:}O_2$ ratios. The integration of deep learning has been instrumental in accurately analyzing data from experiments involving even a single nanoparticle.

This capability to scrutinize catalytic activity at such a granular level could potentially lead to the development of catalysts composed solely of highly efficient nanoparticles, thereby enhancing overall catalytic efficiency. Looking forward, the research opens several promising avenues for exploration. Notably, the potential to correlate the nanoplasmonic response of catalytic particles directly with their catalytic activity offers a novel insight into the interplay between the physical and chemical properties of nanoparticles and their efficacy in catalysis [283]. The versatility of our platform suggests it could be adapted for a wider range of reactions beyond CO oxidation, including those facilitated by metals with significant plasmonic properties, such as Ag in ethylene epoxidation.

Moreover, our methodology presents opportunities for optimizing reactor geometries to facilitate the detection of reaction products from even smaller nanoparticle groups. The exploration of alternative sensing arrangements, such as tailored nanoplasmonic structures for enhancing field strengths, could extend our ability to study catalytic activities on nanoparticles smaller than currently possible. Additionally, enhancing the nanoreactor platform to accommodate high-resolution imaging techniques, like TEM, could provide valuable morphological insights into catalyst particles at different reaction stages.

Incorporating chemically synthesized colloidal nanoparticles into our experiments could further refine our understanding of catalyst performance, leveraging the control over particle crystallinity and structure afforded by chemical synthesis. By continuing to develop and refine these techniques, we aim to achieve a comprehensive understanding of catalyst behavior at the single-particle level, thereby making significant contributions to the field of catalysis and offering potential pathways for the development of more efficient and sustainable catalytic processes.

Acknowledgment

I would like to acknowledge my supervisor Christoph Langhammer and co-supervisor Giovanni Volpe, whose initiative and constant engagement has been absolutely vital to the successful completion of this thesis.

I would like to acknowledge Barbora Špačková, whose insightful descriptions of the underlying biochemistry and physics was absolutely vital for the successful completion of this project, and for providing several useful figures and graphs that were featured in this thesis.

I would like to acknowledge Daniel Midtvedt for his technical expertise within machine learning, without which the successful implementation of these tools would have been impossible.

I would like to acknowledge my fantastic colleagues, whose support and engaging discussions has made my PhD journey an enjoyable one.

I would like to acknowledge the computer cluster Alvis, through which some of the computations in this thesis were made possible. These computations were enabled by resources provided by the Swedish National Infrastructure for Computing (SNIC), partially funded by the Swedish Research Council through grant agreement no. 2018-05973

I would like to acknowledge financial support from the Swedish Foundation for Strategic Research project FFL15-0087, the Swedish Research Council (VR) Consolidator Grant project 2018-00329 and the European Research Council (ERC) under the European Union's Horizon Europe research and innovation program (101043480/NACAREI). Part of this research has been executed at the Chalmers Nanofabrication Laboratory MC2.

Finally, I would like to acknowledge my unnamed yet invaluable friends and family, without the persistent support of whom this thesis would never have been possible.

Henrik Klein Moberg

Gothenburg, November 2024

Bibliography

[1] D. Hassabis, *The promise of artificial intelligence with demis hassabis - deepmind: The podcast (s2, ep9)*, 2022. [Online]. Available: https://www.youtube.com/watch?v=GdeY-MrXD74 (cit. on p. 1).

[2] Y. Huang and Y. Chen, *Autonomous driving with deep learning: A survey of state-of-art technologies*, 2020. DOI: 10.48550/ARXIV.2006.06091. [Online]. Available: https://arxiv.org/abs/2006.06091 (cit. on p. 1).

[3] Z. Hu, J. Tang, Z. Wang, K. Zhang, L. Zhang and Q. Sun, "Deep learning for image-based cancer detection and diagnosis - a survey," *Pattern Recognition*, vol. 83, pp. 134–149, 2018, ISSN: 0031-3203. DOI: https://doi.org/10.1016/j.patcog.2018.05.014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0031320318301845 (cit. on p. 1).

[4] J. Schmidt, M. R. G. Marques, S. Botti and M. A. L. Marques, "Recent advances and applications of machine learning in solid-state materials science," *npj Computational Materials*, vol. 5, no. 1, p. 83, 2019, ISSN: 2057-3960. DOI: 10.1038/s41524-019-0221-0. [Online]. Available: https://doi.org/10.1038/s41524-019-0221-0 (cit. on p. 1).

[5] D. Baron, *Machine learning in astronomy: A practical overview*, 2019. DOI: 10.48550/ARXIV.1904.07248. [Online]. Available: https://arxiv.org/abs/1904.07248 (cit. on p. 1).

[6] J. A. Weyn, D. R. Durran and R. Caruana, "Improving data-driven global weather prediction using deep convolutional neural networks on a cubed sphere," *Journal of Advances in Modeling Earth Systems*, vol. 12, no. 9, e2020MS002109, 2020, e2020MS002109 10.1029/2020MS002109. DOI: https://doi.org/10.1029/2020MS002109. eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2020MS002109. [Online]. Available: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020MS002109 (cit. on p. 1).

[7] O. Garcia-Garin, T. Monleón-Getino, P. López-Brosa *et al.*, "Automatic detection and quantification of floating marine macro-litter in aerial images: Introducing a novel deep learning approach connected to a web application in R," *Environmental Pollution*, vol. 273, p. 116 490,

2021, ISSN: 0269-7491. DOI: https://doi.org/10.1016/j.envpol.2021.116490. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0269749121000683 (cit. on p. 1).

[8] OpenAI, : J. Achiam *et al.*, *GPT-4 technical report*, 2023. arXiv: 2303.08774 [cs.CL] (cit. on p. 1).

[9] Y. LeCun, B. Boser, J. S. Denker *et al.*, "Backpropagation applied to handwritten zip code recognition," *Neural Computation*, vol. 1, no. 4, pp. 541–551, 1989. DOI: 10.1162/neco.1989.1.4.541 (cit. on pp. 1, 8, 9).

[10] Y. LeCun, Y. Bengio and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015, ISSN: 1476-4687. DOI: 10.1038/nature14539. [Online]. Available: https://doi.org/10.1038/nature14539 (cit. on pp. 2, 13).

[11] H. Wang, Y. Rivenson, Y. Jin *et al.*, "Deep learning enables cross-modality super-resolution in fluorescence microscopy," *Nature Methods*, vol. 16, no. 1, pp. 103–110, 2019, ISSN: 1548-7105. DOI: 10.1038/s41592-018-0239-0. [Online]. Available: https://doi.org/10.1038/s41592-018-0239-0 (cit. on p. 2).

[12] B. Midtvedt, E. Olsén, F. Eklund *et al.*, "Fast and accurate nanoparticle characterization using deep-learning-enhanced off-axis holography," *ACS Nano*, vol. 15, no. 2, pp. 2240–2250, 2021, ISSN: 1936-0851. DOI: 10.1021/acsnano.0c06902. [Online]. Available: https://doi.org/10.1021/acsnano.0c06902 (cit. on pp. 2, 21).

[13] S. Helgadottir, B. Midtvedt, J. Pineda *et al.*, "Extracting quantitative biological information from bright-field cell images using deep learning," *Biophysics Reviews*, vol. 2, no. 3, p. 031 401, 2021. DOI: 10.1063/5.0044782. eprint: https://aip.scitation.org/doi/pdf/10.1063/5.0044782. [Online]. Available: https://aip.scitation.org/doi/abs/10.1063/5.0044782 (cit. on pp. 2, 22, 28).

[14] M. Lelek, M. T. Gyparaki, G. Beliu *et al.*, "Single-molecule localization microscopy," *Nature Reviews Methods Primers*, vol. 1, no. 1, p. 39, 2021, ISSN: 2662-8449. DOI: 10.1038/s43586-021-00038-x. [Online]. Available: https://doi.org/10.1038/s43586-021-00038-x (cit. on pp. 2, 30, 34).

[15] G. Storz, Y. I. Wolf and K. S. Ramamurthi, "Small proteins can no longer be ignored," eng, *Annual review of biochemistry*, vol. 83, pp. 753–777, 2014, 24606146[pmid], ISSN: 1545-4509. DOI: 10.1146/annurev-biochem-070611-102400. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/24606146 (cit. on pp. 3, 39).

[16] H. K. Moberg, "Pushing the boundaries of biomolecule characterization through deep learning," *Chalmers University of Technology*, 2022. [Online]. Available: https://research.chalmers.se/en/publication/533941 (cit. on p. 3).

[17] D. A. Roberts, S. Yaida and B. Hanin, *The Principles of Deep Learning Theory*. Cambridge University Press, 2022. DOI: 10.1017/9781009023405. [Online]. Available: https://doi.org/10.1017%2F9781009023405 (cit. on p. 7).

[18] P. Baldi, *Deep Learning in Science*. Cambridge University Press, 2021. DOI: 10.1017/9781108955652 (cit. on pp. 7, 22).

[19] A. D'Amour, K. Heller, D. Moldovan *et al.*, *Underspecification presents challenges for credibility in modern machine learning*, 2020. DOI: 10.48550/ARXIV.2011.03395. [Online]. Available: https://arxiv.org/abs/2011.03395 (cit. on pp. 7, 12, 20, 21).

[20] C. Dwork and M. Louise Minow, "Distrust of artificial intelligence: Sources responses from computer science law," in 2022 (cit. on p. 7).

[21] B. Mehlig, *Machine Learning with Neural Networks*. Cambridge University Press, 2021. DOI: 10.1017/9781108860604. [Online]. Available: https://doi.org/10.1017%2F9781108860604 (cit. on p. 8).

[22] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *The bulletin of mathematical biophysics*, vol. 5, no. 4, pp. 115–133, 1943, ISSN: 1522-9602. DOI: 10.1007/BF02478259. [Online]. Available: https://doi.org/10.1007/BF02478259 (cit. on p. 8).

[23] C. C. Tappert, "Who is the father of deep learning?" In *2019 International Conference on Computational Science and Computational Intelligence (CSCI)*, 2019, pp. 343–348. DOI: 10.1109/CSCI49370.2019.00067 (cit. on p. 8).

[24] H. Wang, B. Raj and E. P. Xing, "On the origin of deep learning," *CoRR*, vol. abs/1702.07800, 2017. arXiv: 1702.07800. [Online]. Available: http://arxiv.org/abs/1702.07800 (cit. on p. 8).

[25] J. Kelleher, *Deep Learning* (The MIT Press Essential Knowledge series). MIT Press, 2019, ISBN: 9780262537551. [Online]. Available: https://books.google.se/books?id=b06qDwAAQBAJ (cit. on p. 8).

[26] A. Krizhevsky, I. Sutskever and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, F. Pereira, C. J. C. Burges, L. Bottou and K. Q. Weinberger, Eds., vol. 25, Curran Associates, Inc., 2012. [Online]. Available: https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf (cit. on pp. 8, 13).

[27] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248–255. DOI: 10.1109/CVPR.2009.5206848 (cit. on pp. 8, 19).

[28] O. Ronneberger, P.Fischer and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, ser. LNCS, (available on arXiv:1505.04597 [cs.CV]), vol. 9351, Springer, 2015, pp. 234–241. [Online]. Available: http://lmb.informatik.uni-freiburg.de/Publications/2015/RFB15a (cit. on pp. 8, 23, 40, 42).

[29] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *CoRR*, vol. abs/1804.02767, 2018. arXiv: 1804.02767. [Online]. Available: http://arxiv.org/abs/1804.02767 (cit. on pp. 8, 43, 44).

[30] K. He, X. Zhang, S. Ren and J. Sun, "Deep residual learning for image recognition," *CoRR*, vol. abs/1512.03385, 2015. arXiv: 1512.03385. [Online]. Available: http://arxiv.org/abs/1512.03385 (cit. on pp. 9, 24).

[31] J. Hawkins and R. Dawkins, *A Thousand Brains: A New Theory of Intelligence*. Basic Books, 2021, ISBN: 9781541675803. [Online]. Available: https://books.google.se/books?id=hYrvDwAAQBAJ (cit. on p. 9).

[32] K. Hornik, M. Stinchcombe and H. White, "Multilayer feedforward networks are universal approximators," *Neural Networks*, vol. 2, no. 5, pp. 359–366, 1989, ISSN: 0893-6080. DOI: https://doi.org/10.1016/0893-6080(89)90020-8. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0893608089900208 (cit. on p. 9).

[33] I. C. Education, *Convolutional neural networkss*, 2020. [Online]. Available: https://www.ibm.com/cloud/learn/convolutional-neural-networks1 (cit. on p. 11).

[34] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Computer Vision – ECCV 2014*, D. Fleet, T. Pajdla, B. Schiele and T. Tuytelaars, Eds., Cham: Springer International Publishing, 2014, pp. 818–833, ISBN: 978-3-319-10590-1 (cit. on pp. 11, 13, 21).

[35] R. Bellman, R. Corporation and K. M. R. Collection, *Dynamic Programming* (Rand Corporation research study). Princeton University Press, 1957, ISBN: 9780691079516. [Online]. Available: https://books.google.se/books?id=wdtoPwAACAAJ (cit. on pp. 12, 13).

[36] J. Long, E. Shelhamer and T. Darrell, "Fully convolutional networks for semantic segmentation," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 3431–3440. DOI: 10.1109/CVPR.2015.7298965 (cit. on p. 12).

[37] H. Gholamalinezhad and H. Khosravi, "Pooling methods in deep neural networks, a review," *CoRR*, vol. abs/2009.07485, 2020. arXiv: 2009.07485. [Online]. Available: https://arxiv.org/abs/2009.07485 (cit. on p. 12).

[38] M. D. Zeiler and R. Fergus, "Stochastic pooling for regularization of deep convolutional neural networks," in *1st International Conference on Learning Representations, ICLR 2013, Scottsdale, Arizona, USA, May 2-4, 2013, Conference Track Proceedings*, Y. Bengio and Y. LeCun, Eds., 2013. [Online]. Available: http://arxiv.org/abs/1301.3557 (cit. on p. 12).

[39] K. He, X. Zhang, S. Ren and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *CoRR*, vol. abs/1406.4729, 2014. arXiv: 1406.4729. [Online]. Available: http://arxiv.org/abs/1406.4729 (cit. on p. 12).

[40] A. Voulodimos, N. Doulamis, A. Doulamis and E. Protopapadakis, "Deep learning for computer vision: A brief review," *Computational Intelligence and Neuroscience*, vol. 2018, p. 7 068 349, 2018, ISSN: 1687-5265. DOI: 10.1155/2018/7068349. [Online]. Available: https://doi.org/10.1155/2018/7068349 (cit. on pp. 13, 21).

[41] R. Girshick, J. Donahue, T. Darrell and J. Malik, *Rich feature hierarchies for accurate object detection and semantic segmentation*, 2013. DOI: 10.48550/ARXIV.1311.2524. [Online]. Available: https://arxiv.org/abs/1311.2524 (cit. on p. 13).

[42] F. Chollet *et al.*, *Keras*, https://keras.io, 2015 (cit. on pp. 13, 20).

[43] A. Paszke, S. Gross, S. Chintala *et al.*, "Automatic differentiation in pytorch," 2017 (cit. on pp. 13, 38).

[44] A. Vaswani, N. Shazeer, N. Parmar *et al.*, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017 (cit. on pp. 14, 15).

[45] J. Devlin, M.-W. Chang, K. Lee and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," in *North American Chapter of the Association for Computational Linguistics*, 2019. [Online]. Available: https://api.semanticscholar.org/CorpusID:52967399 (cit. on p. 15).

[46] A. Dosovitskiy, L. Beyer, A. Kolesnikov *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *ArXiv*, vol. abs/2010.11929, 2020. [Online]. Available: https://api.semanticscholar.org/CorpusID:225039882 (cit. on p. 15).

[47] A. Dosovitskiy, L. Beyer, A. Kolesnikov *et al.*, *An image is worth 16x16 words: Transformers for image recognition at scale*, 2021. arXiv: 2010.11929 [cs.CV]. [Online]. Available: https://arxiv.org/abs/2010.11929 (cit. on pp. 17, 48).

[48] K. Weiss, T. M. Khoshgoftaar and D. Wang, "A survey of transfer learning," *Journal of Big Data*, vol. 3, no. 1, p. 9, 2016, ISSN: 2196-1115. DOI: 10.1186/s40537-016-0043-6. [Online]. Available: https://doi.org/10.1186/s40537-016-0043-6 (cit. on p. 20).

[49] M. D. Zeiler and R. Fergus, *Visualizing and understanding convolutional networks*, D. Fleet, T. Pajdla, B. Schiele and T. Tuytelaars, Eds., Cham, 2013. DOI: 10.48550/ARXIV.1311.2901. [Online]. Available: https://arxiv.org/abs/1311.2901 (cit. on p. 20).

[50] L. Rokach, "Ensemble methods for classifiers," in *Data Mining and Knowledge Discovery Handbook*, O. Maimon and L. Rokach, Eds. Boston, MA: Springer US, 2005, pp. 957–980, ISBN: 978-0-387-25465-4. DOI: 10.1007/0-387-25465-X_45. [Online]. Available: https://doi.org/10.1007/0-387-25465-X_45 (cit. on p. 20).

[51] O. Sagi and L. Rokach, "Ensemble learning: A survey," *WIREs Data Mining and Knowledge Discovery*, vol. 8, no. 4, e1249, 2018. DOI: https://doi.org/10.1002/widm.1249. eprint: https://wires.onlinelibrary.wiley.com/doi/pdf/10.1002/widm.1249. [Online]. Available: https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/widm.1249 (cit. on p. 20).

[52] F. Chollet, *Deep Learning with Python*. Manning, Nov. 2017, ISBN: 9781617294433 (cit. on p. 21).

[53] Y. Bengio, J. Louradour, R. Collobert and J. Weston, "Curriculum learning," in *Proceedings of the 26th Annual International Conference on Machine Learning*, ser. ICML '09, Montreal, Quebec, Canada: Association for Computing Machinery, 2009, 41–48, ISBN: 9781605585161. DOI: 10.1145/1553374.1553380. [Online]. Available: https://doi.org/10.1145/1553374.1553380 (cit. on p. 21).

[54] M. Csikszentmihalyi, "Flow: The psychology of optimal experience," in Jan. 1990 (cit. on p. 21).

[55] B. Settles, "Active learning literature survey," 2009 (cit. on p. 21).

[56] D. Heaven, "Why deep-learning AIs are so easy to fool," *Nature*, vol. 574, no. 7777, pp. 163–166, Oct. 2019. DOI: 10.1038/d41586-019-03013-5. [Online]. Available: https://doi.org/10.1038/d41586-019-03013-5 (cit. on p. 21).

[57] S. Kaplan, D. Handelman and A. Handelman, "Sensitivity of neural networks to corruption of image classification," *AI and Ethics*, vol. 1, no. 4, pp. 425–434, 2021, ISSN: 2730-5961. DOI: 10.1007/s43681-021-00049-0. [Online]. Available: https://doi.org/10.1007/s43681-021-00049-0 (cit. on p. 21).

[58] J. Grauer, F. Schmidt, J. Pineda *et al.*, "Active droploids," *Nature Communications*, vol. 12, no. 1, p. 6005, 2021, ISSN: 2041-1723. DOI: 10.1038/s41467-021-26319-3. [Online]. Available: https://doi.org/10.1038/s41467-021-26319-3 (cit. on p. 21).

[59] R. Penrose, *Fashion, Faith, and Fantasy in the New Physics of the Universe*. Princeton University Press, Jun. 2016, ISBN: 9780691178530. DOI: 10.2307/j.ctvc775bn. [Online]. Available: https://doi.org/10.2307/j.ctvc775bn (cit. on pp. 22, 38).

[60] G. Volpe, C. Wählby, L. Tian *et al.*, "Roadmap on deep learning for microscopy," en, *ArXiv*, Mar. 2023 (cit. on p. 22).

[61] T. Lin, M. Maire, S. J. Belongie *et al.*, "Microsoft COCO: common objects in context," *CoRR*, vol. abs/1405.0312, 2014. arXiv: 1405.0312. [Online]. Available: http://arxiv.org/abs/1405.0312 (cit. on p. 23).

[62] D. Shen, G. Wu and H.-I. Suk, "Deep learning in medical image analysis," *Annual Review of Biomedical Engineering*, vol. 19, no. 1, pp. 221–248, 2017, PMID: 28301734. DOI: 10.1146/annurev-bioeng-071516-044442. eprint: https://doi.org/10.1146/annurev-bioeng-071516-044442. [Online]. Available: https://doi.org/10.1146/annurev-bioeng-071516-044442 (cit. on p. 22).

[63] K. Sung and T. Poggio, "Example based learning for view-based human face detection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 20, pp. 39–51, Feb. 1998. DOI: 10.1109/34.655648 (cit. on p. 24).

[64] P. Dollar, C. Wojek, B. Schiele and P. Perona, "Pedestrian detection: An evaluation of the state of the art," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 4, pp. 743–761, 2012. DOI: 10.1109/TPAMI.2011.155 (cit. on p. 24).

[65] C. Chen, A. Seff, A. Kornhauser and J. Xiao, "Deepdriving: Learning affordance for direct perception in autonomous driving," in *2015 IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 2722–2730. DOI: 10.1109/ICCV.2015.312 (cit. on p. 24).

[66] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, *You only look once: Unified, real-time object detection*, 2016. arXiv: 1506.02640 [cs.CV] (cit. on p. 24).

[67] A. B. Oktay and A. Gurses, "Automatic detection, localization and segmentation of nano-particles with deep learning in microscopy images," *Micron*, vol. 120, pp. 113–119, 2019, ISSN: 0968-4328. DOI: https://doi.org/10.1016/j.micron.2019.02.009. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0968432818304013 (cit. on p. 25).

[68] X. wang, T. Xu, J. Zhang, S. Chen and Y. Zhang, "So-yolo based wbc detection with fourier ptychographic microscopy," *IEEE Access*, vol. PP, pp. 1–1, Aug. 2018. DOI: 10.1109/ACCESS.2018.2865541 (cit. on p. 24).

[69] P. Cunningham, M. Cord and S. J. Delany, "Supervised learning," in *Machine Learning Techniques for Multimedia: Case Studies on Organization and Retrieval*, M. Cord and P. Cunningham, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 21–49, ISBN: 978-3-540-75171-7. DOI: 10.1007/978-3-540-75171-7_2. [Online]. Available: https://doi.org/10.1007/978-3-540-75171-7_2 (cit. on p. 26).

[70] D. A. Roberts, S. Yaida and B. Hanin, "The principles of deep learning theory," *CoRR*, vol. abs/2106.10165, 2021. arXiv: 2106.10165. [Online]. Available: https://arxiv.org/abs/2106.10165 (cit. on p. 26).

[71] B. Midtvedt, J. Pineda, F. Skärberg *et al.*, "Single-shot self-supervised object detection in microscopy," *Nature Communications*, vol. 13, no. 1, p. 7492, 2022, ISSN: 2041-1723. DOI: 10.1038/s41467-022-35004-y. [Online]. Available: https://doi.org/10.1038/s41467-022-35004-y (cit. on pp. 26, 27).

[72] J. Chen, Q. Ye, X. Yang, S. K. Zhou, X. Hong and L. Zhang, "Few-shot learning for multi-modality tasks," in *Proceedings of the 29th ACM International Conference on Multimedia*. New York, NY, USA: Association for Computing Machinery, 2021, 5673–5674, ISBN: 9781450386517. [Online]. Available: https://doi.org/10.1145/3474085.3478873 (cit. on p. 26).

[73] H.-H. Bock, "Clustering methods: A history of k-means algorithms," in *Selected Contributions in Data Analysis and Classification*, P. Brito, G. Cucumel, P. Bertrand and F. de Carvalho, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 161–172, ISBN: 978-3-540-73560-1. DOI: 10.1007/978-3-540-73560-1_15. [Online]. Available: https://doi.org/10.1007/978-3-540-73560-1_15 (cit. on p. 26).

[74] J. An and S. Cho, "Variational autoencoder based anomaly detection using reconstruction probability," 2015 (cit. on p. 26).

[75] T. D. Gebhard, N. Kilbertus, I. Harry and B. Schölkopf, "Convolutional neural networks: A magic bullet for gravitational-wave detection?" *Physical Review D*, vol. 100, no. 6, 2019. DOI: 10.1103/physrevd.100.063015. [Online]. Available: https://doi.org/10.1103%2Fphysrevd.100.063015 (cit. on p. 26).

[76] H. Shen, D. George, E. A. Huerta and Z. Zhao, "Denoising gravitational waves with enhanced deep recurrent denoising auto-encoders," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2019, pp. 3237–3241, ISBN: 9781479981311. DOI: 10.1109/icassp.2019.8683061. [Online]. Available: https://doi.org/10.1109%2Ficassp.2019.8683061 (cit. on p. 26).

[77] T. B. Brown, B. Mann, N. Ryder *et al.*, *Language models are few-shot learners*, 2020. DOI: 10.48550/ARXIV.2005.14165. [Online]. Available: https://arxiv.org/abs/2005.14165 (cit. on p. 27).

[78] A. D. Cohen, A. Roberts, A. Molina *et al.*, "Lamda: Language models for dialog applications," in *arXiv*, 2022 (cit. on p. 27).

[79] M. Caron, H. Touvron, I. Misra *et al.*, "Emerging properties in self-supervised vision transformers," in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021, pp. 9630–9640. DOI: 10.1109/ICCV48922.2021.00951 (cit. on p. 27).

[80] D. Silver, A. Huang, C. J. Maddison *et al.*, "Mastering the game of go with deep neural networks and tree search," *Nature*, vol. 529, pp. 484–503, 2016. [Online]. Available: http://www.nature.com/nature/journal/v529/n7587/full/nature16961.html (cit. on p. 27).

[81] O. Vinyals, I. Babuschkin, W. M. Czarnecki *et al.*, "Grandmaster level in starcraft II using multi-agent reinforcement learning," *Nature*, vol. 575, no. 7782, pp. 350–354, 2019, ISSN: 1476-4687. DOI: 10.1038/s41586-019-1724-z. [Online]. Available: https://doi.org/10.1038/s41586-019-1724-z (cit. on p. 27).

[82] OpenAI, : C. Berner *et al.*, *Dota 2 with large scale deep reinforcement learning*, 2019. DOI: 10.48550/ARXIV.1912.06680. [Online]. Available: https://arxiv.org/abs/1912.06680 (cit. on p. 27).

[83] A. Kadian, J. Truong, A. Gokaslan *et al.*, "Sim2real predictivity: Does evaluation in simulation predict real-world performance?" *IEEE Robotics and Automation Letters*, vol. 5, no. 4, pp. 6670–6677, 2020. DOI: 10.1109/LRA.2020.3013848 (cit. on p. 27).

[84] K. Arulkumaran, M. P. Deisenroth, M. Brundage and A. A. Bharath, "Deep reinforcement learning: A brief survey," *IEEE Signal Processing Magazine*, vol. 34, no. 6, pp. 26–38, 2017. DOI: 10.1109/msp.2017.2743240. [Online]. Available: https://doi.org/10.1109%2Fmsp.2017.2743240 (cit. on p. 27).

[85] A. Mirhoseini, A. Goldie, M. Yazgan *et al.*, "A graph placement methodology for fast chip design," *Nature*, vol. 594, no. 7862, pp. 207–212, 2021, ISSN: 1476-4687. DOI: 10.1038/s41586-021-03544-w. [Online]. Available: https://doi.org/10.1038/s41586-021-03544-w (cit. on p. 27).

[86] R. D. King, K. E. Whelan, F. M. Jones *et al.*, "Functional genomic hypothesis generation and experimentation by a robot scientist," *Nature*, vol. 427, no. 6971, pp. 247–252, 2004, ISSN: 1476-4687. DOI: 10.1038/nature02236. [Online]. Available: https://doi.org/10.1038/nature02236 (cit. on p. 27).

[87] L. Wang, W. Chen, W. Yang, F. Bi and F. Yu, "A state-of-the-art review on image synthesis with generative adversarial networks," *IEEE Access*, vol. PP, pp. 1–1, Mar. 2020. DOI: 10.1109/ACCESS.2020.2982224 (cit. on p. 27).

[88] M. Hossam, T. Le, M. Papasimeon, V. Huynh and D. Phung, *Text generation with deep variational GAN*, 2021. DOI: 10.48550/ARXIV.2104.13488. [Online]. Available: https://arxiv.org/abs/2104.13488 (cit. on p. 27).

[89] P. Baniukiewicz, E. J. Lutton, S. Collier and T. Bretschneider, "Generative adversarial networks for augmenting training data of microscopic cell images," *Frontiers in Computer Science*, vol. 1, 2019, ISSN: 2624-9898. DOI: 10.3389/fcomp.2019.00010. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fcomp.2019.00010 (cit. on p. 27).

[90] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza *et al.*, *Generative adversarial networks*, 2014. DOI: 10.48550/ARXIV.1406.2661. [Online]. Available: https://arxiv.org/abs/1406.2661 (cit. on p. 28).

[91] R. Rizzuto, M. Brini, P. Pizzo, M. Murgia and T. Pozzan, "Chimeric green fluorescent protein as a tool for visualizing subcellular organelles in living cells," *Current biology*, vol. 5, no. 6, pp. 635–642, 1995 (cit. on p. 28).

[92] O. Kepp, L. Galluzzi, M. Lipinski, J. Yuan and G. Kroemer, "Cell death assays for drug discovery," *Nature reviews Drug discovery*, vol. 10, no. 3, pp. 221–237, 2011 (cit. on p. 28).

[93] V. Lulevich, Y.-P. Shih, S. H. Lo and G.-y. Liu, "Cell tracing dyes significantly change single cell mechanics," *The Journal of Physical Chemistry B*, vol. 113, no. 18, pp. 6511–6519, 2009 (cit. on p. 28).

[94] C. Ounkomol, S. Seshamani, M. M. Maleckar, F. Collman and G. R. Johnson, "Label-free prediction of three-dimensional fluorescence images from transmitted-light microscopy," *Nature methods*, vol. 15, no. 11, pp. 917–920, 2018 (cit. on p. 28).

[95] N. Pillar and A. Ozcan, "Virtual tissue staining in pathology using machine learning," *Expert Review of Molecular Diagnostics*, vol. 22, no. 11, pp. 987–989, 2022, PMID: 36440487. DOI: 10.1080/14737159.2022.2153040. eprint: https://doi.org/10.1080/14737159.2022.2153040. [Online]. Available: https://doi.org/10.1080/14737159.2022.2153040 (cit. on p. 28).

[96] X. Li, G. Zhang, H. Qiao *et al.*, "Unsupervised content-preserving transformation for optical microscopy," *Light: Science & Applications*, vol. 10, no. 1, pp. 1–11, 2021 (cit. on p. 28).

[97] T. C. Nguyen, V. Bui, A. Thai *et al.*, "Virtual organelle self-coding for fluorescence imaging via adversarial learning," *Journal of biomedical optics*, vol. 25, no. 9, p. 096 009, 2020 (cit. on p. 28).

[98] Z. Korczak, J. Pineda, S. Helgadottir *et al.*, "Dynamic live/apoptotic cell assay using phase-contrast imaging and deep learning," *bioRxiv*, 2022 (cit. on p. 28).

[99] M. E. Kandel, E. Kim, Y. J. Lee, G. Tracy, H. J. Chung and G. Popescu, "Multiscale assay of unlabeled neurite dynamics using phase imaging with computational specificity," *ACS sensors*, vol. 6, no. 5, pp. 1864–1874, 2021 (cit. on p. 28).

[100] Y. N. Nygate, M. Levi, S. K. Mirsky *et al.*, "Holographic virtual staining of individual biological cells," *Proceedings of the National Academy of Sciences*, vol. 117, no. 17, pp. 9223–9231, 2020 (cit. on p. 28).

[101] Y. Liu, H. Yuan, Z. Wang and S. Ji, "Global pixel transformers for virtual staining of microscopy images," *IEEE Transactions on Medical Imaging*, vol. 39, no. 6, pp. 2256–2266, 2020 (cit. on p. 28).

[102] E. Nehme, L. E. Weiss, T. Michaeli and Y. Shechtman, "Deep-storm: Super-resolution single-molecule microscopy by deep learning," *Optica*, vol. 5, no. 4, pp. 458–464, 2018. DOI: 10.1364/OPTICA.5.000458. [Online]. Available: http://opg.optica.org/optica/abstract.cfm?URI=optica-5-4-458 (cit. on p. 31).

[103] W. Ouyang, A. Aristov, M. Lelek, X. Hao and C. Zimmer, "Deep learning massively accelerates super-resolution localization microscopy," *Nature Biotechnology*, vol. 36, no. 5, pp. 460–468, 2018, ISSN: 1546-1696. DOI: 10.1038/nbt.4106. [Online]. Available: https://doi.org/10.1038/nbt.4106 (cit. on p. 31).

[104] M. J. Rust, M. Bates and X. Zhuang, "Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM)," en, *Nat Methods*, vol. 3, no. 10, pp. 793–795, Aug. 2006 (cit. on p. 30).

[105] H. Shroff, H. White and E. Betzig, "Photoactivated localization microscopy (PALM) of adhesion complexes," en, *Curr Protoc Cell Biol*, vol. Chapter 4, Unit4.21, Mar. 2013 (cit. on p. 30).

[106] S. T. Hess, T. P. Girirajan and M. D. Mason, "Ultra-high resolution imaging by fluorescence photoactivation localization microscopy," *Biophysical Journal*, vol. 91, no. 11, pp. 4258–4272, 2006, ISSN: 0006-3495. DOI: https://doi.org/10.1529/biophysj.106.091116. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0006349506721403 (cit. on p. 30).

[107] M. Jung, D. Kim and J. Y. Mun, "Direct visualization of actin filaments and Actin-Binding proteins in neuronal cells," en, *Front Cell Dev Biol*, vol. 8, p. 588 556, Nov. 2020 (cit. on p. 30).

[108] V. Mannam, Y. Zhang, X. Yuan and S. Howard, "Deep learning-based super-resolution fluorescence microscopy on small datasets," in *Single Molecule Spectroscopy and Superresolution Imaging XIV*, I. Gregor, R. Erdmann and F. Koberling, Eds., SPIE, 2021. DOI: 10.1117/12.2578519 (cit. on p. 30).

[109] E. Wolf. Elsevier, 2008, p. 355, ISBN: 978-0-08-055768-7 (cit. on p. 30).

[110] R. Weissleder and U. Mahmood, "Molecular imaging," *Radiology*, vol. 219, no. 2, pp. 316–333, 2001, PMID: 11323453. DOI: 10.1148/radiology.219.2.r01ma19316. eprint: https://doi.org/10.1148/radiology.219.2.r01ma19316. [Online]. Available: https://doi.org/10.1148/radiology.219.2.r01ma19316 (cit. on p. 33).

[111] E. Britannica, *History of the Optical Microscope*. Encyclopedia Britannica, Inc. (cit. on p. 33).

[112] X. Zhuang, "Nano-imaging with storm," *Nature Photonics*, vol. 3, no. 7, pp. 365–367, 2009, ISSN: 1749-4893. DOI: 10.1038/nphoton.2009.101. [Online]. Available: https://doi.org/10.1038/nphoton.2009.101 (cit. on p. 33).

[113] K. Thorn, "A quick guide to light microscopy in cell biology," eng, *Molecular biology of the cell*, vol. 27, no. 2, pp. 219–222, 2016, 26768859[pmid], ISSN: 1939-4586. DOI: 10.1091/mbc.E15-02-0088. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/26768859 (cit. on p. 34).

[114] S. Shashkova and M. C. Leake, "Single-molecule fluorescence microscopy review: Shedding new light on old problems," eng, *Bioscience reports*, vol. 37, no. 4, BSR20170031, 2017, 28694303[pmid], ISSN: 1573-4935. DOI: 10.1042/BSR20170031. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/28694303 (cit. on p. 34).

[115] B. Špačková, H. Klein Moberg, J. Fritzsche *et al.*, "Label-free nanofluidic scattering microscopy of size and mass of single diffusing molecules and nanoparticles," *Nature Methods*, vol. 19, no. 6, pp. 751–758, 2022, ISSN: 1548-7105. DOI: 10.1038/s41592-022-01491-6. [Online]. Available: https://doi.org/10.1038/s41592-022-01491-6 (cit. on pp. 35–37, 40, 42, 44, 45, 53, 54).

[116] M. Piliarik and V. Sandoghdar, "Direct optical sensing of single unlabelled proteins and super-resolution imaging of their binding sites," *Nature Communications*, vol. 5, no. 1, p. 4495, 2014, ISSN: 2041-1723. DOI: 10.1038/ncomms5495. [Online]. Available: https://doi.org/10.1038/ncomms5495 (cit. on p. 35).

[117] M. Piliarik and V. Sandoghdar, "Direct optical sensing of single unlabelled proteins and super-resolution imaging of their binding sites," *Nature communications*, vol. 5, no. 1, pp. 1–8, 2014, ISSN: 2041-1723 (cit. on p. 35).

[118] P. Dechadilok and W. M. Deen, "Hindrance factors for diffusion and convection in pores," *Industrial & engineering chemistry research*, vol. 45, no. 21, pp. 6953–6959, 2006 (cit. on p. 36).

[119] J. A. Marsh and S. A. Teichmann, "Structure, dynamics, assembly, and evolution of protein complexes," *Annual Review of Biochemistry*, vol. 84, no. 1, pp. 551–575, 2015, PMID: 25494300. DOI: 10.1146/annurev-biochem-060614-034142. eprint: https://doi.org/10.1146/annurev-biochem-060614-034142. [Online]. Available: https://doi.org/10.1146/annurev-biochem-060614-034142 (cit. on p. 37).

[120] A. Bruce, J. Alexander, L. Julian *et al.*, *Molecular Biology of the Cell*. 4th edition, 2002 (cit. on p. 37).

[121] M. Prpić, M. Franceschi, M. Romić, T. Jukić and Z. Kusić, "Thyroglobulin as a tumor marker in differentiated thyroid cancer - clinical considerations," en, *Acta Clin Croat*, vol. 57, no. 3, pp. 518–527, Sep. 2018 (cit. on p. 37).

[122] R. E. Casiday and R. F. Frey, "Iron use and storage in the body : Ferritin and molecular representations iron in biology : Study of the iron content in ferritin , the iron-storage protein .," 2004 (cit. on p. 37).

[123] C. B. Chen, B. Hammo, J. Barry and K. Radhakrishnan, "Overview of albumin physiology and its role in pediatric diseases," *Current Gastroenterology Reports*, vol. 23, no. 8, p. 11, 2021, ISSN: 1534-312X. DOI: 10.1007/s11894-021-00813-6. [Online]. Available: https://doi.org/10.1007/s11894-021-00813-6 (cit. on p. 37).

[124] H. J. Edenberg and J. N. McClintick, "Alcohol dehydrogenases, aldehyde dehydrogenases, and alcohol use disorders: A critical review," en, *Alcohol Clin Exp Res*, vol. 42, no. 12, pp. 2281–2297, Nov. 2018 (cit. on p. 37).

[125] S. T. Kalinowski, M. J. Leonard and T. M. Andrews, "Nothing in evolution makes sense except in the light of DNA," en, *CBE Life Sci Educ*, vol. 9, no. 2, pp. 87–97, 2010 (cit. on p. 37).

[126] V. Müller and F. Westerlund, "Optical DNA mapping in nanofluidic devices: Principles and applications," *Lab on a Chip*, vol. 17, no. 4, pp. 579–590, 2017 (cit. on p. 37).

[127] B. TA., "Genomes," *2nd edition. Oxford: Wiley-Liss; 2002. Chapter*, vol. 1, 2002 (cit. on p. 37).

[128] R. J. Simpson, J. W. Lim, R. L. Moritz and S. Mathivanan, "Exosomes: Proteomic insights and diagnostic potential," *Expert Review of Proteomics*, vol. 6, no. 3, pp. 267–283, 2009, PMID: 19489699. DOI: 10.1586/epr.09.17. eprint: https://doi.org/10.1586/epr.09.17. [Online]. Available: https://doi.org/10.1586/epr.09.17 (cit. on pp. 37, 38).

[129] E. van der Pol, A. N. Böing, P. Harrison, A. Sturk and R. Nieuwland, "Classification, functions, and clinical relevance of extracellular vesicles," *Pharmacological Reviews*, vol. 64, no. 3, M. P. Mattson, Ed., pp. 676–705, 2012, ISSN: 0031-6997. DOI: 10.1124/pr.112.005983. eprint: https://pharmrev.aspetjournals.org/content/64/3/676.full.pdf. [Online]. Available: https://pharmrev.aspetjournals.org/content/64/3/676 (cit. on p. 38).

[130] B. Berne and R. Pecora, *Dynamic Light Scattering: With Applications to Chemistry, Biology, and Physics* (Dover Books on Physics). Dover Publications, 2013, ISBN: 9780486320243. [Online]. Available: https://books.google.se/books?id=xg3CAgAAQBAJ (cit. on p. 38).

[131] V. Filipe, A. Hawe and W. Jiskoot, "Critical evaluation of nanoparticle tracking analysis (nta) by nanosight for the measurement of nanoparticles and protein aggregates," eng, *Pharmaceutical research*, vol. 27, no. 5, pp. 796–810, 2010, 20204471[pmid], ISSN: 1573-904X. DOI: 10.1007/s11095-010-0073-2. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/20204471 (cit. on p. 38).

[132] M. Abadi, A. Agarwal, P. Barham *et al.*, *TensorFlow: Large-scale machine learning on heterogeneous systems*, Software available from tensorflow.org, 2015. [Online]. Available: https://www.tensorflow.org/ (cit. on p. 38).

[133] F. Rapaport, R. Khanin, Y. Liang *et al.*, "Comprehensive evaluation of differential gene expression analysis methods for rna-seq data," *Genome Biology*, vol. 14, no. 9, p. 3158, 2013, ISSN: 1474-760X. DOI: 10.1186/gb-2013-14-9-r95. [Online]. Available: https://doi.org/10.1186/gb-2013-14-9-r95 (cit. on p. 39).

[134] J. P. Hughes, S Rees, S. B. Kalindjian and K. L. Philpott, "Principles of early drug discovery," en, *Br J Pharmacol*, vol. 162, no. 6, pp. 1239–1249, Mar. 2011 (cit. on p. 39).

[135] S Leonelli, "What difference does quantity make? on the epistemology of big data in biology," *Big Data & Society*, vol. 1, no. 1, p. 2 053 951 714 534 395, 2014, PMID: 25729586. DOI: 10.1177/2053951714534395. eprint: https://doi.org/10.1177/2053951714534395. [Online]. Available: https://doi.org/10.1177/2053951714534395 (cit. on p. 39).

[136] J. Zrimec, C. S. Börlin, F. Buric *et al.*, "Deep learning suggests that gene expression is encoded in all parts of a co-evolving interacting gene regulatory structure," *Nature Communications*, vol. 11, no. 1, p. 6141, 2020, ISSN: 2041-1723. DOI: 10.1038/s41467-020-19921-4. [Online]. Available: https://doi.org/10.1038/s41467-020-19921-4 (cit. on p. 39).

[137] J. Jumper, R. Evans, A. Pritzel *et al.*, "Highly accurate protein structure prediction with alphafold," *Nature*, vol. 596, no. 7873, pp. 583–589, 2021, ISSN: 1476-4687. DOI: 10.1038/s41586-021-03819-2. [Online]. Available: https://doi.org/10.1038/s41586-021-03819-2 (cit. on p. 39).

[138] N. M. Luscombe, S. E. Austin, H. M. Berman and J. M. Thornton, "An overview of the structures of protein-dna complexes," *Genome Biology*, vol. 1, no. 1, reviews001.1, 2000, ISSN: 1474-760X. DOI: 10.1186/gb-2000-1-1-reviews001. [Online]. Available: https://doi.org/10.1186/gb-2000-1-1-reviews001 (cit. on p. 39).

[139] C. Corbo, R. Molinaro, A. Parodi, N. E. Toledano Furman, F. Salvatore and E. Tasciotti, "The impact of nanoparticle protein corona on cytotoxicity, immunotoxicity and target drug delivery," en, *Nanomedicine (Lond)*, vol. 11, no. 1, pp. 81–100, Dec. 2015 (cit. on p. 39).

[140] H. K. Moberg, B. Yeroshenko, D. Midtvedt *et al.*, "Deep learning for nanofluidic scattering microscopy," in *Emerging Topics in Artificial Intelligence (ETAI) 2023*, G. Volpe, J. B. Pereira, D. Brunner and A. Ozcan, Eds., International Society for Optics and Photonics, vol. PC12655, SPIE, 2023, PC126550J. DOI: 10.1117/12.2676769. [Online]. Available: https://doi.org/10.1117/12.2676769 (cit. on p. 39).

[141] A. Vianello, R. L. Jensen, L. Liu and J. Vollertsen, "Simulating human exposure to indoor airborne microplastics using a breathing thermal manikin," *Scientific Reports*, vol. 9, no. 1, p. 8670, 2019, ISSN: 2045-2322. DOI: 10.1038/s41598-019-45054-w. [Online]. Available: https://doi.org/10.1038/s41598-019-45054-w (cit. on p. 39).

[142] J. Fritzsche, D. Albinsson, M. Fritzsche, T. J. Antosiewicz, F. Westerlund and C. Langhammer, "Single particle nanoplasmonic sensing in individual nanofluidic channels," *Nano Letters*, vol. 16, no. 12, pp. 7857–7864, 2016, ISSN: 1530-6984. DOI: 10.1021/acs.nanolett.6b04124. [Online]. Available: https://doi.org/10.1021/acs.nanolett.6b04124 (cit. on pp. 39, 117).

[143] L. Roeder, *Netron*, https://github.com/lutzroeder/netron, 2020 (cit. on pp. 42, 44, 45).

[144] D. P. Kingma and J. Ba, *Adam: A method for stochastic optimization*, 2014. DOI: 10.48550/ARXIV.1412.6980. [Online]. Available: https://arxiv.org/abs/1412.6980 (cit. on pp. 41, 45, 46).

[145] J. Redmon, *Darknet: Open Source Neural Networks in C*. http://pjreddie.com/darknet/ (20132016). [Online]. Available: http://pjreddie.com/darknet/ (cit. on pp. 44, 45).

[146] E. Noren, *Pytorch-yolov3*, https://github.com/eriklindernoren/PyTorch-YOLOv3, 2019 (cit. on p. 43).

[147] Z. Lu, J. Lu, Q. Ge and T. Zhan, "Multi-object detection method based on yolo and resnet hybrid networks," in *2019 IEEE 4th International Conference on Advanced Robotics and Mechatronics (ICARM)*, 2019, pp. 827–832 (cit. on p. 43).

[148] Z.-Q. Zhao, P. Zheng, S. tao Xu and X. Wu, *Object detection with deep learning: A review*, 2019. arXiv: 1807.05511 [cs.CV] (cit. on p. 43).

[149] E. Linder Noren, *Pytorch-yolov3*, https://github.com/eriklindernoren/PyTorch-YOLOv3, 2019 (cit. on p. 43).

[150] A. Ghosh, B. G. Sumpter, O. Dyck, S. V. Kalinin and M. Ziatdinov, "Ensemble learning-iterative training machine learning for uncertainty quantification and automated experiment in atom-resolved microscopy," *npj Computational Materials*, vol. 7, no. 1, p. 100, 2021, ISSN: 2057-3960. DOI: 10.1038/s41524-021-00569-7. [Online]. Available: https://doi.org/10.1038/s41524-021-00569-7 (cit. on p. 46).

[151] J. Pineda, B. Midtvedt, H. Bachimanchi *et al.*, *Geometric deep learning reveals the spatiotemporal fingerprint of microscopic motion*, 2022. DOI: 10.48550/ARXIV.2202.06355. [Online]. Available: https://arxiv.org/abs/2202.06355 (cit. on pp. 48, 115, 116).

[152] N. Rahaman, A. Baratin, D. Arpit *et al.*, "On the spectral bias of neural networks," in *Proceedings of the 36th International Conference on Machine Learning*, K. Chaudhuri and R. Salakhutdinov, Eds., ser. Proceedings of Machine Learning Research, vol. 97, PMLR, 2019, pp. 5301–5310. [Online]. Available: https://proceedings.mlr.press/v97/rahaman19a.html (cit. on p. 48).

[153] H. Cramér, *Mathematical Methods of Statistics*. Princeton University Press, 1946 (cit. on p. 55).

[154] C. R. Rao, "Information and the accuracy attainable in the estimation of statistical parameters," *Bulletin of the Calcutta Mathematical Society*, vol. 37, pp. 81–91, 1945 (cit. on p. 55).

[155] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*. Prentice Hall, 1993 (cit. on pp. 55, 57).

[156] W. Feller, *An Introduction to Probability Theory and Its Applications*, 3rd. John Wiley & Sons, 1968, vol. 1 (cit. on p. 55).

[157] E. L. Lehmann and G. Casella, *Theory of Point Estimation*, 2nd. Springer, 1998 (cit. on p. 56).

[158] D. Curran-Everett, "Explorations in statistics: The bootstrap," *Advances in Physiology Education*, vol. 33, no. 4, pp. 286–292, 2009, PMID: 19948676. DOI: 10.1152/advan.00062.2009. eprint: https://doi.org/10.1152/advan.00062.2009. [Online]. Available: https://doi.org/10.1152/advan.00062.2009 (cit. on p. 65).

[159] B Prieto-Simín, M. Campàs and J.-L Marty, "Biomolecule immobilization in biosensor development: Tailored strategies based on affinity interactions," *Protein and peptide letters*, vol. 15, pp. 757–63, Feb. 2008. DOI: 10.2174/092986608785203791 (cit. on p. 65).

[160] S. S. Jensen, H. Jensen, C. Cornett, E. H. Møller and J. Østergaard, "Insulin diffusion and self-association characterized by real-time UV imaging and taylor dispersion analysis," en, *J Pharm Biomed Anal*, vol. 92, pp. 203–210, Jan. 2014 (cit. on p. 65).

[161] C. Langhammer, I. Zorić, B. Kasemo and B. M. Clemens, "Hydrogen storage in pd nanodisks characterized with a novel nanoplasmonic sensing scheme," *Nano Letters*, vol. 7, no. 10, pp. 3122–3127, 2007, ISSN: 1530-6984. DOI: 10.1021/nl071664a. eprint: https://pubs.aip.org/aip/jcp/article-pdf/doi/10.1063/1.2734550/13494151/194702_1_online.pdf. [Online]. Available: https://doi.org/10.1021/nl071664a (cit. on p. 69).

[162] J. N. Anker, W. P. Hall, O. Lyandres, N. C. Shah, J. Zhao and R. P. Van Duyne, "Biosensing with plasmonic nanosensors," *Nature Materials*, vol. 7, no. 6, pp. 442–453, 2008, ISSN: 1476-4660. DOI: 10.1038/nmat2162. [Online]. Available: https://doi.org/10.1038/nmat2162 (cit. on pp. 69, 73, 77).

[163] F. A. A. Nugroho, D. Albinsson, T. J. Antosiewicz and C. Langhammer, "Plasmonic metasurface for spatially resolved optical sensing in three dimensions," *ACS Nano*, vol. 14, no. 2, pp. 2345–2353, 2020, ISSN: 1936-0851. DOI: 10.1021/acsnano.9b09508. [Online]. Available: https://doi.org/10.1021/acsnano.9b09508 (cit. on p. 69).

[164] K. A. Willets and R. P. Van Duyne, "Localized surface plasmon resonance spectroscopy and sensing," *Annual Review of Physical Chemistry*, vol. 58, no. 1, pp. 267–297, 2007, PMID: 17067281. DOI: 10.1146/annurev.physchem.58.032806.104607. eprint: https://doi.org/10.1146/annurev.physchem.58.032806.104607. [Online]. Available: https://doi.org/10.1146/annurev.physchem.58.032806.104607 (cit. on p. 69).

[165] A. R. Sepulveda, O. Kyriacou and J. Treasure, "Development and validation of the accommodation and enabling scale for eating disorders (aesed) for caregivers in eating disorders," *BMC Health Services Research*, vol. 9, no. 1, p. 171, 2009, ISSN: 1472-6963. DOI: 10.1186/1472-6963-9-171. [Online]. Available: https://doi.org/10.1186/1472-6963-9-171 (cit. on p. 69).

[166] S. Eustis and M. A. El-Sayed, "Why gold nanoparticles are more precious than pretty gold: Noble metal surface plasmon resonance and its enhancement of the radiative and nonradiative properties of nanocrystals of different shapes," *Chem. Soc. Rev.*, vol. 35, pp. 209–217, 3 2006. DOI: 10.1039/B514191E. [Online]. Available: http://dx.doi.org/10.1039/B514191E (cit. on pp. 70, 71).

[167] C. Langhammer and E. M. Larsson, "Nanoplasmonic in situ spectroscopy for catalysis applications," *ACS Catalysis*, vol. 2, no. 9, pp. 2036–2045, 2012. DOI: 10.1021/cs300423a. eprint: https://doi.org/10.1021/cs300423a. [Online]. Available: https://doi.org/10.1021/cs300423a (cit. on pp. 72, 73, 77).

[168] C. Novo, A. M. Funston and P. Mulvaney, "Direct observation of chemical reactions on single gold nanocrystals using surface plasmon spectroscopy," *Nature Nanotechnology*, vol. 3, no. 10, pp. 598–602, 2008, ISSN: 1748-3395. DOI: 10.1038/nnano.2008.246. [Online]. Available: https://doi.org/10.1038/nnano.2008.246 (cit. on pp. 73, 77, 91).

[169] E. M. Larsson, S. Syrenova and C. Langhammer, "Nanoplasmonic sensing for nanomaterials science," *Nanophotonics*, vol. 1, no. 3-4, pp. 249–266, 2012. DOI: doi:10.1515/nanoph-2012-0029. [Online]. Available: https://doi.org/10.1515/nanoph-2012-0029 (cit. on p. 73).

[170] A. B. Taylor and P. Zijlstra, "Single-Molecule plasmon sensing: Current status and future prospects," en, *ACS Sens*, vol. 2, no. 8, pp. 1103–1122, Aug. 2017 (cit. on pp. 73, 77).

[171] G. Mie, "Beiträge zur optik trüber medien, speziell kolloidaler metallösungen," *Annalen der Physik*, vol. 330, no. 3, pp. 377–445, 1908. DOI: https://doi.org/10.1002/andp.19083300302. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/andp.19083300302. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/andp.19083300302 (cit. on p. 73).

[172] S. A. Maier, *Plasmonics: Fundamentals and Applications*. Springer, 2007 (cit. on pp. 73, 74).

[173] A. Bansal and S. Verma, "Optical response of noble metal alloy nanostructures," *Physics Letters A*, vol. 379, no. 3, pp. 163–169, 2015, ISSN: 0375-9601. DOI: https://doi.org/10.1016/j.physleta.2014.11.018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0375960114011463 (cit. on p. 76).

[174] F. A. A. Nugroho, B. Iandolo, J. B. Wagner and C. Langhammer, "Bottom-up nanofabrication of supported noble metal alloy nanoparticle arrays for plasmonics," *ACS Nano*, vol. 10, no. 2, pp. 2871–2879, 2016, ISSN: 1936-0851. DOI: 10.1021/acsnano.5b08057. [Online]. Available: https://doi.org/10.1021/acsnano.5b08057 (cit. on p. 76).

[175] A. Lalisse, G. Tessier, J. Plain and G. Baffou, "Plasmonic efficiencies of nanoparticles made of metal nitrides (tin, zrn) compared with gold," *Scientific Reports*, vol. 6, no. 1, p. 38 647, 2016, ISSN: 2045-2322. DOI: 10.1038/srep38647. [Online]. Available: https://doi.org/10.1038/srep38647 (cit. on p. 76).

[176] A. Wokaun, J. P. Gordon and P. F. Liao, "Radiation damping in surface-enhanced raman scattering," *Phys. Rev. Lett.*, vol. 48, pp. 957–960, 14 1982. DOI: 10.1103/PhysRevLett.48.957. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.48.957 (cit. on p. 76).

[177] K. L. Kelly, E. Coronado, L. L. Zhao and G. C. Schatz, "The optical properties of metal nanoparticles: the influence of size, shape, and dielectric environment," *The Journal of Physical Chemistry B*, vol. 107, no. 3, pp. 668–677, 2003, ISSN: 1520-6106. DOI: 10.1021/jp026731y. [Online]. Available: https://doi.org/10.1021/jp026731y (cit. on p. 76).

[178] J. Zhao, A. O. Pinchuk, J. M. McMahon *et al.*, "Methods for describing the electromagnetic properties of silver and gold nanoparticles," *Accounts of Chemical Research*, vol. 41, no. 12, pp. 1710–1720, 2008, ISSN: 0001-4842. DOI: 10.1021/ar800028j. [Online]. Available: https://doi.org/10.1021/ar800028j (cit. on p. 76).

[179] A. Taflove, S. Hagness and M. Piket-May, "Computational electromagnetics: The finite-difference time-domain method," English (US), in *The Electrical Engineering Handbook*. Elsevier Inc, Dec. 2005, pp. 629–670, ISBN: 9780121709600. DOI: 10.1016/B978-012170960-0/50046-3 (cit. on p. 76).

[180] F. J. García de Abajo and A. Howie, "Retarded field calculation of electron energy loss in inhomogeneous dielectrics," *Phys. Rev. B*, vol. 65, p. 115 418, 11 2002. DOI: 10.1103/PhysRevB.65.115418. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevB.65.115418 (cit. on p. 76).

[181] I. Zorić, M. Zäch, B. Kasemo and C. Langhammer, "Gold, platinum, and aluminum nanodisk plasmons: Material independence, subradiance, and damping mechanisms," *ACS Nano*, vol. 5, no. 4, pp. 2535–2546, 2011, ISSN: 1936-0851. DOI: 10.1021/nn102166t. [Online]. Available: https://doi.org/10.1021/nn102166t (cit. on p. 77).

[182] T. Jensen, M. Malinsky, C. Haynes and R. Duyne, "Nanosphere lithography: Tunable localized surface plasmon resonance spectra of silver nanoparticles," *Journal of Physical Chemistry B - J PHYS CHEM B*, vol. 104, Nov. 2000. DOI: 10.1021/jp002435e (cit. on p. 77).

[183] T. R. Jensen, M. L. Duval, K. L. Kelly, A. A. Lazarides, G. C. Schatz and R. P. Van Duyne, "Nanosphere lithography: effect of the external dielectric medium on the surface plasmon resonance spectrum of a periodic array of silver nanoparticles," *The Journal of Physical Chemistry B*, vol. 103, no. 45, pp. 9846–9853, 1999, ISSN: 1520-6106. DOI: 10.1021/jp9926802. [Online]. Available: https://doi.org/10.1021/jp9926802 (cit. on p. 77).

[184] P. Englebienne, "Use of colloidal gold surface plasmon resonance peak shift to infer affinity constants from the interactions between protein antigens and antibodies specific for single or multiple epitopes," *Analyst*, vol. 123, pp. 1599–1603, 7 1998. DOI: 10.1039/A804010I. [Online]. Available: http://dx.doi.org/10.1039/A804010I (cit. on p. 77).

[185] B. Foerster, J. Rutten, H. Pham, S. Link and C. Sönnichsen, "Particle plasmons as dipole antennas: State representation of relative observables," *The Journal of Physical Chemistry C*, vol. 122, no. 33, pp. 19 116–19 123, 2018, ISSN: 1932-7447. DOI: 10.1021/acs.jpcc.8b06350. [Online]. Available: https://doi.org/10.1021/acs.jpcc.8b06350 (cit. on p. 77).

[186] A. B. Dahlin, J. O. Tegenfeldt and F. Höök, "Improving the instrumental resolution of sensors based on localized surface plasmon resonance," *Analytical Chemistry*, vol. 78, no. 13, pp. 4416–4423, 2006, PMID: 16808449, ISSN: 0003-2700. DOI: 10.1021/ac0601967. eprint: https://doi.org/10.1021/ac0601967. [Online]. Available: https://doi.org/10.1021/ac0601967 (cit. on p. 77).

[187] J. A. Jackman, A. Rahim Ferhan and N.-J. Cho, "Nanoplasmonic sensors for biointerfacial science," *Chem. Soc. Rev.*, vol. 46, pp. 3615–3660, 12 2017. DOI: 10.1039/C6CS00494F. [Online]. Available: http://dx.doi.org/10.1039/C6CS00494F (cit. on p. 77).

[188] A. J. Haes, S. Zou, G. C. Schatz and R. P. Van Duyne, "Nanoscale optical biosensor: short range distance dependence of the localized surface plasmon resonance of noble metal nanoparticles," *The Journal of Physical Chemistry B*, vol. 108, no. 22, pp. 6961–6968, 2004, ISSN: 1520-6106. DOI: 10.1021/jp036261n. [Online]. Available: https://doi.org/10.1021/jp036261n (cit. on p. 77).

[189] C. Wadell and C. Langhammer, "Drift-corrected nanoplasmonic hydrogen sensing by polarization," *Nanoscale*, vol. 7, pp. 10 963–10 969, 25 2015. DOI: 10.1039/C5NR01818H. [Online]. Available: http://dx.doi.org/10.1039/C5NR01818H (cit. on p. 77).

[190] C. Wadell, S. Syrenova and C. Langhammer, "Plasmonic hydrogen sensing with nanostructured metal hydrides," *ACS Nano*, vol. 8, no. 12, pp. 11 925–11 940, 2014, ISSN: 1936-0851. DOI: 10.1021/nn505804f. [Online]. Available: https://doi.org/10.1021/nn505804f (cit. on p. 77).

[191] C. Langhammer, V. P. Zhdanov, familyi=c., given=Igor, giveni=I., ,, and B. Kasemo, "Size-dependent kinetics of hydriding and dehydriding of pd nanoparticles," *Phys. Rev. Lett.*, vol. 104, p. 135 502, 13 2010. DOI: 10.1103/PhysRevLett.104.135502. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.104.135502 (cit. on p. 77).

[192] T. Shegai and C. Langhammer, "Hydride formation in single palladium and magnesium nanoparticles studied by nanoplasmonic dark-field scattering spectroscopy," *Advanced Materials*, vol. 23, no. 38, pp. 4409–4414, 2011. DOI: https://doi.org/10.1002/adma.201101976. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/adma.201101976. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/adma.201101976 (cit. on p. 77).

[193] X. Duan, S. Kamin, F. Sterl, H. Giessen and N. Liu, "Hydrogen-regulated chiral nanoplasmonics," *Nano Letters*, vol. 16, no. 2, pp. 1462–1466, 2016, ISSN: 1530-6984. DOI: 10.1021/acs.nanolett.5b05105. [Online]. Available: https://doi.org/10.1021/acs.nanolett.5b05105 (cit. on p. 77).

[194] F. Sterl, H. Linnenbank, T. Steinle, F. Mörz, N. Strohfeldt and H. Giessen, "Nanoscale hydrogenography on single magnesium nanoparticles," *Nano Letters*, vol. 18, no. 7, pp. 4293–4302, 2018, ISSN: 1530-6984. DOI: 10.1021/acs.nanolett.8b01277. [Online]. Available: https://doi.org/10.1021/acs.nanolett.8b01277 (cit. on p. 77).

[195] K. P. Rice, E. J. Walker Jr., M. P. Stoykovich and A. E. Saunders, "Solvent-dependent surface plasmon response and oxidation of copper nanocrystals," *The Journal of Physical Chemistry C*, vol. 115, no. 5, pp. 1793–1799, 2011, ISSN: 1932-7447. DOI: 10.1021/jp110483z. [Online]. Available: https://doi.org/10.1021/jp110483z (cit. on p. 77).

[196] M. D. Susman, Y. Feldman, T. A. Bendikov, A. Vaskevich and I. Rubinstein, "Real-time plasmon spectroscopy study of the solid-state oxidation and kirkendall void formation in copper nanoparticles," *Nanoscale*, vol. 9, pp. 12 573–12 589, 34 2017. DOI: 10.1039/C7NR04256F. [Online]. Available: http://dx.doi.org/10.1039/C7NR04256F (cit. on p. 77).

[197] K. M. Mayer and J. H. Hafner, "Localized surface plasmon resonance sensors," *Chemical Reviews*, vol. 111, no. 6, pp. 3828–3857, 2011, ISSN: 0009-2665. DOI: 10.1021/cr100313v. [Online]. Available: https://doi.org/10.1021/cr100313v (cit. on p. 77).

[198] I. Darmadi, F. A. A. Nugroho and C. Langhammer, "High-performance nanostructured palladium-based hydrogen sensors—current limitations and strategies for their mitigation," *ACS sensors*, vol. 5, no. 11, pp. 3306–3327, 2020 (cit. on pp. 78–80).

[199] S. Alekseeva, I. I. Nedrygailov and C. Langhammer, "Single particle plasmonics for materials science and single particle catalysis," *ACS Photonics*, vol. 6, no. 6, pp. 1319–1330, 2019 (cit. on p. 79).

[200] T. Hübert, L Boon-Brett, G. Black and U. Banach, "Hydrogen sensors–a review," *Sensors and Actuators B: Chemical*, vol. 157, no. 2, pp. 329–352, 2011 (cit. on p. 79).

[201] U Sripriya and U. S. Meda, "Market study of hydrogen sensors and sensing systems," *ECS Transactions*, vol. 107, no. 1, p. 4489, 2022 (cit. on p. 79).

[202] C. Wadell, S. Syrenova and C. Langhammer, "Plasmonic hydrogen sensing with nanostructured metal hydrides," *ACS nano*, vol. 8, no. 12, pp. 11 925–11 940, 2014 (cit. on p. 79).

[203] F. A. A. Nugroho, I. Darmadi, V. P. Zhdanov and C. Langhammer, "Universal scaling and design rules of hydrogen-induced optical properties in pd and pd-alloy nanoparticles," *Acs Nano*, vol. 12, no. 10, pp. 9903–9912, 2018 (cit. on p. 80).

[204] M. A. Poyli, V. M. Silkin, I. Chernov, P. M. Echenique, R. D. Muino and J. Aizpurua, "Multiscale theoretical modeling of plasmonic sensing of hydrogen uptake in palladium nanodisks," *The Journal of Physical Chemistry Letters*, vol. 3, no. 18, pp. 2556–2561, 2012 (cit. on p. 80).

[205] I. Zoric, E. M. Larsson, B. Kasemo and C. Langhammer, "Localized surface plasmons shed light on nanoscale metal hydrides," *Adv. Mater*, vol. 22, no. 41, pp. 4628–4633, 2010 (cit. on p. 80).

[206] *U.s. department of energy; 2015 safety, codes and standards section*, 2015. [Online]. Available: https://www.energy.gov/sites/default/files/2015/06/f23/fcto_myrdd_safety_codes.pdf (cit. on p. 80).

[207] W. Buttner, R. Burgess, M. Post and C. Rivkin, "Summary and findings from the nrel/doe hydrogen sensor workshop (june 8, 2011)," DOI: 10.2172/1048994. [Online]. Available: https://www.osti.gov/biblio/1048994 (cit. on p. 80).

[208] F. A. Nugroho, I. Darmadi, L. Cusinato *et al.*, "Metal–polymer hybrid nanomaterials for plasmonic ultrafast hydrogen detection," *Nature Materials*, vol. 18, no. 5, pp. 489–495, 2019. DOI: 10.1038/s41563-019-0325-4 (cit. on p. 80).

[209] H. M. Luong, M. T. Pham, T. Guin *et al.*, "Sub-second and ppm-level optical sensing of hydrogen using templated control of nano-hydride geometry and composition," *Nature communications*, vol. 12, no. 1, p. 2414, 2021. DOI: 10.1038/s41467-021-22697-w (cit. on p. 80).

[210] L. Bannenberg, H. Schreuders and B. Dam, "Tantalum-palladium: Hysteresis-free optical hydrogen sensor over 7 orders of magnitude in pressure with sub-second response," *Advanced Functional Materials*, vol. 31, no. 16, p. 2 010 483, 2021. DOI: 10.1002/adfm.202010483 (cit. on p. 80).

[211] I. Darmadi, F. A. A. Nugroho and C. Langhammer, "High-Performance Nanostructured Palladium-Based Hydrogen Sensors–Current Limitations and Strategies for Their Mitigation," *ACS Sensors*, vol. 5, no. 11, pp. 3306–3327, 2020. DOI: 10.1021/acssensors.0c02019 (cit. on p. 80).

[212] I. Darmadi, F. A. A. Nugroho, S. Kadkhodazadeh, J. B. Wagner and C. Langhammer, "Rationally designed pdaucu ternary alloy nanoparticles for intrinsically deactivation-resistant ultrafast plasmonic hydrogen sensing," *ACS sensors*, vol. 4, no. 5, pp. 1424–1432, 2019 (cit. on p. 80).

[213] F. A. Nugroho, I. Darmadi, L. Cusinato *et al.*, "Metal–polymer hybrid nanomaterials for plasmonic ultrafast hydrogen detection," *Nature materials*, vol. 18, no. 5, pp. 489–495, 2019 (cit. on p. 80).

[214] F. A. A. Nugroho, P. Bai, I. Darmadi *et al.*, "Inverse designed plasmonic metasurface with parts per billion optical hydrogen detection," *Nature Communications*, vol. 13, no. 1, p. 5737, 2022 (cit. on p. 80).

[215] I. Darmadi, F. A. A. Nugroho, S. Kadkhodazadeh, J. B. Wagner and C. Langhammer, "Rationally Designed PdAuCu Ternary Alloy Nanoparticles for Intrinsically Deactivation-Resistant Ultrafast Plasmonic Hydrogen Sensing," *ACS Sensors*, vol. 4, no. 5, pp. 1424–1432, May 2019. DOI: 10.1021/acssensors.9b00610 (cit. on p. 80).

[216] E. K. Alenezy, Y. M. Sabri, A. E. Kandjani *et al.*, "Low-temperature hydrogen sensor: Enhanced performance enabled through photoactive pd-decorated tio2 colloidal crystals," *ACS Sensors*, vol. 5, no. 12, pp. 3902–3914, 2020. DOI: 10.1021/acssensors.0c01387 (cit. on p. 80).

[217] S. Kabcum, D. Channei, A. Tuantranont, A. Wisitsoraat, C. Liewhiran and S. Phanichphant, "Ultra-responsive hydrogen gas sensors based on pdo nanoparticle-decorated wo3 nanorods synthesized by precipitation and impregnation methods," *Sensors and Actuators B: Chemical*, vol. 226, pp. 76–89, 2016. DOI: 10.1016/j.snb.2015.11.120 (cit. on p. 80).

[218] K. Hassan, A. Iftekhar Uddin and G.-S. Chung, "Fast-response hydrogen sensors based on discrete pt/pd bimetallic ultra-thin films," *Sensors and Actuators B: Chemical*, vol. 234, pp. 435–445, 2016, ISSN: 0925-4005. DOI: 10.1016/j.snb.2016.05.013 (cit. on p. 80).

[219] M. Losurdo, Y. Gutiérrez, A. Suvorova *et al.*, "Gallium plasmonic nanoantennas unveiling multiple kinetics of hydrogen sensing, storage, and spillover," *Advanced Materials*, vol. 33, no. 29, p. 2 100 500, 2021 (cit. on p. 80).

[220] S. Huang, A. Croy, B. Ibarlucea and G. Cuniberti, "Machine learning-driven gas identification in gas sensors," in *Machine Learning for Advanced Functional Materials*, N. Joshi, V. Kushvaha and P. Madhushri, Eds. Singapore: Springer Nature Singapore, 2023, pp. 21–41, ISBN: 978-981-99-0393-1. DOI: 10.1007/978-981-99-0393-1_2. [Online]. Available: https://doi.org/10.1007/978-981-99-0393-1_2 (cit. on p. 80).

[221] M. Kang, I. Cho, J. Park *et al.*, "High accuracy real-time multi-gas identification by a batch-uniform gas sensor array and deep learning algorithm," *ACS Sensors*, vol. 7, no. 2, pp. 430–440, 2022. DOI: 10.1021/acssensors.1c01204 (cit. on p. 80).

[222] T. Hayasaka, A. Lin, V. C. Copa *et al.*, "An electronic nose using a single graphene fet and machine learning for water, methanol, and ethanol," *Microsystems & nanoengineering*, vol. 6, no. 1, p. 50, 2020 (cit. on p. 80).

[223] D. R. Wijaya, F. Afianti, A. Arifianto, D. Rahmawati and V. S. Kodogiannis, "Ensemble machine learning approach for electronic nose signal processing," *Sensing and Bio-Sensing Research*, vol. 36, p. 100 495, 2022, ISSN: 2214-1804. DOI: https://doi.org/10.1016/j.sbsr.2022.100495. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2214180422000241 (cit. on p. 80).

[224] N. X. Thai, M. Tonezzer, L. Masera, H. Nguyen, N. V. Duy and N. D. Hoa, "Multi gas sensors using one nanomaterial, temperature gradient, and machine learning algorithms for discrimination of gases and their concentration," *Analytica Chimica Acta*, vol. 1124, pp. 85–93, 2020, ISSN: 0003-2670. DOI: https://doi.org/10.1016/j.aca.2020.05.015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0003267020305407 (cit. on p. 80).

[225] S. Kanaparthi and S. G. Singh, "Discrimination of gases with a single chemiresistive multi-gas sensor using temperature sweeping and machine learning," *Sensors and Actuators B: Chemical*, vol. 348, p. 130 725, 2021, ISSN: 0925-4005. DOI: https://doi.org/10.1016/j.snb.2021.130725. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925400521012934 (cit. on p. 80).

[226] B. Ai, Y. Sun and Y. Zhao, "Plasmonic hydrogen sensors," *Small*, vol. 18, no. 25, p. 2 107 882, 2022. DOI: 10.1002/smll.202107882 (cit. on p. 80).

[227] C. Wadell, S. Syrenova and C. Langhammer, "Plasmonic hydrogen sensing with nanostructured metal hydrides," *ACS Nano*, vol. 8, no. 12, pp. 11 925–11 940, 2014. DOI: 10.1021/nn505804f (cit. on p. 80).

[228] A. B. Dahlin, J. O. Tegenfeldt and F. Höök, "Improving the instrumental resolution of sensors based on localized surface plasmon resonance," *Analytical Chemistry*, vol. 78, no. 13, pp. 4416–4423, 2006. DOI: 10.1021/ac0601967 (cit. on p. 81).

[229] D. Tomeček, H. K. Moberg, S. Nilsson *et al.*, "Neural network enabled nanoplasmonic hydrogen sensors with 100 ppm limit of detection in humid air," *Nature Communications*, vol. 15, no. 1, p. 1208, 2024, ISSN: 2041-1723. DOI: 10.1038/s41467-024-45484-9. [Online]. Available: https://doi.org/10.1038/s41467-024-45484-9 (cit. on p. 81).

[230] S. Ioffe and C. Szegedy, *Batch normalization: Accelerating deep network training by reducing internal covariate shift*, 2015. arXiv: 1502.03167 [cs.LG] (cit. on p. 81).

[231] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, 1929–1958, 2014, ISSN: 1532-4435 (cit. on p. 81).

[232] V. Martvall, H. K. Moberg, A. Theodoridis *et al.*, *Accelerating plasmonic hydrogen sensors for inert gas environments by transformer-based deep learning*, 2023. arXiv: 2312.15372 [physics.comp-ph]. [Online]. Available: https://arxiv.org/abs/2312.15372 (cit. on p. 85).

[233] M. Xu, Y. Xiong, H. Chen *et al.*, "Long short-term transformer for online action detection," in *NeurIPS 2021*, 2021. [Online]. Available: https://www.amazon.science/publications/long-short-term-transformer-for-online-action-detection (cit. on pp. 85, 87).

[234] P. Mishra, D. Passos, F. Marini *et al.*, "Deep learning for near-infrared spectral data modelling: Hypes and benefits," *TrAC Trends in Analytical Chemistry*, vol. 157, p. 116 804, 2022, ISSN: 0165-9936. DOI: 10.1016/j.trac.2022.116804. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0165993622002874 (cit. on p. 85).

[235] G. Ertl, *Reactions at surfaces: From atoms to complexity*, 2007 (cit. on pp. 89, 91).

[236] C. A. B. et. al., "Transition metal catalysis in the pharmaceutical industry," 2005. [Online]. Available: https://content.schweitzer-online.de/static/catalog_manager/live/media_files/representation/zd_std_orig__zd_schw_orig/002/059/499/9780470631324_content_pdf_1.pdf (cit. on p. 89).

[237] M. A. van Spronsen, J. W. M. Frenken and I. M. N. Groot, "Surface science under reaction conditions: Co oxidation on pt and pd model catalysts," *Chem. Soc. Rev.*, vol. 46, pp. 4347–4374, 14 2017. DOI: 10.1039/C7CS00045F. [Online]. Available: http://dx.doi.org/10.1039/C7CS00045F (cit. on pp. 90, 92).

[238] "Mind the gap," *Nature Catalysis*, vol. 1, no. 11, pp. 807–808, 2018, ISSN: 2520-1158. DOI: 10.1038/s41929-018-0188-0. [Online]. Available: https://doi.org/10.1038/s41929-018-0188-0 (cit. on p. 90).

[239] D. V. Schroeder, *An introduction to thermal physics*. San Francisco, CA : Addison Wesley, [2000] ©2000, [2000], Includes bibliographical references (pages 397-405) and index. [Online]. Available: https://search.library.wisc.edu/catalog/999908452302121 (cit. on p. 90).

[240] S. Alekseeva, I. I. Nedrygailov and C. Langhammer, "Single particle plasmonics for materials science and single particle catalysis," *ACS Photonics*, vol. 6, no. 6, pp. 1319–1330, 2019. DOI: 10.1021/acsphotonics.9b00339. [Online]. Available: https://doi.org/10.1021/acsphotonics.9b00339 (cit. on pp. 91, 94).

[241] I. Chorkendorff and J. Niemantsverdriet, *Concepts of Modern Catalysis and Kinetics*, English, 3rd ed. Germany: Wiley-VCH, 2017, ISBN: 9783527332687 (cit. on pp. 91–93).

[242] V. S. Edmond de Hoffmann, *Mass Spectrometry: Principles and Applications*. WILEY, 2004, p. 65 (cit. on pp. 92, 93).

[243] J. M. Dahl, A. Maiolica, M. Gstaiger and R. Aebersold, "Mass spectrometry for the study of protein–protein interactions," *Methods*, vol. 24, no. 3, pp. 230–239, 2001 (cit. on p. 93).

[244] W. Abushareeda, E. Lyris, S. Kraiem, A. Thomas and M. Thevis, "Mass spectrometry in doping control: Structure characterization and analytical assays," *Mass spectrometry reviews*, vol. 36, no. 2, pp. 196–217, 2017 (cit. on p. 93).

[245] Y. Li, F. Xu, F. Zhang *et al.*, "DLBI: Deep learning guided Bayesian inference for structure reconstruction of super-resolution fluorescence microscopy," *Bioinformatics*, vol. 34, no. 13, pp. i284–i294, 2018, ISSN: 14602059. DOI: 10.1093/bioinformatics/bty241. arXiv: 1805.07777. [Online]. Available: https://doi.org/10.1038/s41929-018-0150-1 (cit. on p. 94).

[246] B. Qiao, A. Wang, X. Yang *et al.*, "Single-atom catalysis of CO oxidation using Pt1/FeOx," *Nature chemistry*, vol. 3, no. 8, pp. 634–641, 2011 (cit. on p. 94).

[247] X. Zhang, Z. Sun, X. Wang *et al.*, "Single-atom catalysts: Synthetic strategies and electrochemical applications," *Joule*, vol. 2, no. 7, pp. 1242–1264, 2018 (cit. on p. 94).

[248] L. Liu and A. Corma, "Single-atom catalysts: A new frontier in heterogeneous catalysis," *Accounts of chemical research*, vol. 51, no. 4, pp. 789–799, 2018 (cit. on p. 94).

[249] J. B. Sambur, T. Chen, E. Choudhary *et al.*, "Sub-particle reaction and photocurrent mapping to optimize catalyst-modified photoanodes," en, *Nature*, vol. 530, no. 7591, pp. 77–80, Jan. 2016 (cit. on p. 94).

[250] H. Wei and H. Xu, "Shell-isolated nanoparticle-enhanced raman spectroscopy: Principle and applications," *Chemical Society Reviews*, vol. 40, no. 5, pp. 2494–2507, 2011 (cit. on p. 94).

[251] W. Xu, Y. Zhang and T. Chen, *Single Particle Nanocatalysis: Fundamentals and Applications*. Apr. 2019, ISBN: 9783527343294 (cit. on p. 94).

[252] T. Hartman, R. G. Geitenbeek, C. S. Wondergem, W. van der Stam and B. M. Weckhuysen, "Operando nanoscale sensors in catalysis: All eyes on catalyst particles," *ACS Nano*, vol. 14, no. 4, pp. 3725–3735, 2020, ISSN: 1936-0851 (cit. on p. 94).

[253] T. Hartman, C. S. Wondergem, N. Kumar, A. van den Berg and B. M. Weckhuysen, "Surface- and tip-enhanced raman spectroscopy in catalysis," *The Journal of Physical Chemistry Letters*, vol. 7, no. 8, pp. 1570–1584, 2016. DOI: 10.1021/acs.jpclett.6b00147. [Online]. Available: https://doi.org/10.1021/acs.jpclett.6b00147 (cit. on p. 94).

[254] G. Henriksen, P. Hoff, R. Larsen and J. Alstad, "Microreactor and method for preparing a radiolabeled complex or a biomolecule conjugate," *US Patent*, no. 7,553,463, 2009 (cit. on p. 94).

[255] J. C. Schouten, T. A. Nijhuis, L. C. van den Oetelaar and C. M. van den Bleek, "Microreactors for kinetic studies of heterogeneous catalytic reactions," *Applied Catalysis A: General*, vol. 221, no. 1-2, pp. 253–264, 2001 (cit. on p. 94).

[256] P. C. Vesborg, J. L. Olsen, T. R. Henriksen, I. Chorkendorff and O. Hansen, "Gas-phase photocatalysis in micro-reactors," *Chemical Engineering Journal*, vol. 160, no. 2, pp. 738–741, 2010, ISSN: 1385-8947 (cit. on p. 94).

[257] K. J. P. Schouten, E. Pérez-Gallent and M. T. Koper, "Nanoreactors: Engineering catalyst particles with enzyme-like behavior," *Chemical Science*, vol. 8, no. 7, pp. 5159–5166, 2017 (cit. on p. 94).

[258] D. Albinsson, A. Boje, S. Nilsson *et al.*, "Copper catalysis at operando conditions—bridging the gap between single nanoparticle probing and catalyst-bed-averaging," *Nature Communications*, vol. 11, no. 1, p. 4832, 2020, ISSN: 2041-1723 (cit. on pp. 94, 108).

[259] D. Albinsson, S. Bartling, S. Nilsson, H. Ström, J. Fritzsche and C. Langhammer, "Operando detection of single nanoparticle activity dynamics inside a model pore catalyst material," *Science Advances*, vol. 6, no. 25, eaba7678, 2020. eprint: https://www.science.org/doi/pdf/10.1126/sciadv.aba7678 (cit. on pp. 94, 108).

[260] D. A. Boiko, K. S. Kozlov, J. V. Burykina, V. V. Ilyushenkova and V. P. Ananikov, "Fully automated unconstrained analysis of high-resolution mass spectrometry data with machine learning," *Journal of the American Chemical Society*, vol. 144, no. 32, pp. 14 590–14 606, 2022, ISSN: 0002-7863. DOI: 10.1021/jacs.2c03631. [Online]. Available: https://doi.org/10.1021/jacs.2c03631 (cit. on pp. 94, 108).

[261] W. M. Abdelmoula, B. G.-C. Lopez, E. C. Randall *et al.*, "Peak learning of mass spectrometry imaging data using artificial neural networks," *Nature Communications*, vol. 12, no. 1, p. 5544, 2021, ISSN: 2041-1723. DOI: 10.1038/s41467-021-25744-8. [Online]. Available: https://doi.org/10.1038/s41467-021-25744-8 (cit. on pp. 94, 108).

[262] U. W. Liebal, A. N. T. Phan, M. Sudhakar, K. Raman and L. M. Blank, "Machine learning applications for mass Spectrometry-Based metabolomics," en, *Metabolites*, vol. 10, no. 6, Jun. 2020 (cit. on pp. 94, 108).

[263] H. Ji, H. Deng, H. Lu and Z. Zhang, "Predicting a molecular fingerprint from an electron ionization mass spectrum with deep neural networks," *Analytical Chemistry*, vol. 92, no. 13, pp. 8649–8653, 2020, ISSN: 0003-2700. DOI: 10.1021/acs.analchem.0c01450. [Online]. Available: https://doi.org/10.1021/acs.analchem.0c01450 (cit. on pp. 94, 108).

[264] A. Jetybayeva, N. Borodinov, A. V. Ievlev *et al.*, "A review on recent machine learning applications for imaging mass spectrometry studies," *Journal of Applied Physics*, vol. 133, no. 2, p. 020 702, Jan. 2023, ISSN: 0021-8979. DOI: 10.1063/5.0100948. eprint: https://pubs.aip.org/aip/jap/article-pdf/doi/10.1063/5.0100948/16680211/020702_1_online.pdf. [Online]. Available: https://doi.org/10.1063/5.0100948 (cit. on pp. 94, 108).

[265] Z. Zhou and R. N. Zare, "Personal information from latent fingerprints using desorption electrospray ionization mass spectrometry and machine learning," *Analytical Chemistry*, vol. 89, no. 2, pp. 1369–1372, 2017, ISSN: 0003-2700. DOI: 10.1021/acs.analchem.6b04498. [Online]. Available: https://doi.org/10.1021/acs.analchem.6b04498 (cit. on pp. 94, 108).

[266] S. Gessulat, T. Schmidt, D. P. Zolg *et al.*, "Prosit: Proteome-wide prediction of peptide tandem mass spectra by deep learning," *Nature Methods*, vol. 16, no. 6, pp. 509–518, 2019, ISSN: 1548-7105. DOI: 10.1038/s41592-019-0426-7. [Online]. Available: https://doi.org/10.1038/s41592-019-0426-7 (cit. on pp. 94, 108).

[267] N. H. Tran, R. Qiao, L. Xin *et al.*, "Deep learning enables de novo peptide sequencing from data-independent-acquisition mass spectrometry," *Nature Methods*, vol. 16, no. 1, pp. 63–66, 2019, ISSN: 1548-7105. DOI: 10.1038/s41592-018-0260-3. [Online]. Available: https://doi.org/10.1038/s41592-018-0260-3 (cit. on pp. 94, 108).

[268] H. K. Moberg, I. Nedrygailov, J. Fritzsche, D. Albinsson and C. Langhammer, "Deep-learning enabled online mass spectrometry of the reaction product from a single catalyst nanoparticle," *ResearchSquare*, 2024, PREPRINT (Version 1) available at Research Square. [Online]. Available: https://doi.org/10.21203/rs.3.rs-3797128/v1 (cit. on p. 94).

[269] Y. Zhou, Z. Wang and C. Liu, "Perspective on CO oxidation over Pd-based catalysts," *Catal. Sci. Technol.*, vol. 5, pp. 69–81, 1 2015 (cit. on p. 94).

[270] M. A. van Spronsen, J. W. M. Frenken and I. M. N. Groot, "Surface science under reaction conditions: CO oxidation on Pt and Pd model catalysts," *Chem. Soc. Rev.*, vol. 46, pp. 4347–4374, 14 2017 (cit. on p. 94).

[271] P. Vincent, H. Larochelle, Y. Bengio and P.-A. Manzagol, "Extracting and composing robust features with denoising autoencoders," in *Proceedings of the 25th International Conference on Machine Learning*, ser. ICML '08, Helsinki, Finland: Association for Computing Machinery, 2008, 1096–1103, ISBN: 9781605582054 (cit. on p. 95).

[272] Y. Bengio, L. Yao, G. Alain and P. Vincent, *Generalized denoising auto-encoders as generative models*, 2013 (cit. on pp. 95, 97, 101).

[273] C. M. Bishop, "Pattern recognition and machine learning," in Springer, 2006, ch. 3, Linear Models for Regression, pp. 137–211, ISBN: 978-0-387-31073-2 (cit. on p. 101).

[274] K. Frykholm, V. Müller, S. KK, K. D. Dorfman and F. Westerlund, "Dna in nanochannels: Theory and applications," *Quarterly Reviews of Biophysics*, vol. 55, e12, 2022. DOI: 10.1017/S0033583522000117 (cit. on p. 114).

[275] E. Gutierrez-Beltran, P. N. Moschou, A. P. Smertenko and P. V. Bozhkov, "Tudor staphylococcal nuclease links formation of stress granules and processing bodies with mRNA catabolism in arabidopsis," en, *Plant Cell*, vol. 27, no. 3, pp. 926–943, Mar. 2015 (cit. on p. 114).

[276] M. Yanagida, "Functional proteomics; current achievements," *Journal of Chromatography B*, vol. 771, no. 1-2, pp. 89–106, 2002 (cit. on p. 115).

[277] V. S. Rao, K Srinivas, G. N. Sujini and G. N. S. Kumar, "Protein-protein interaction detection: Methods and analysis," en, *Int J Proteomics*, vol. 2014, p. 147 648, Feb. 2014 (cit. on p. 115).

[278] V. Müller, F. Rajer, K. Frykholm *et al.*, "Direct identification of antibiotic resistance genes on single plasmid molecules using CRISPR/Cas9 in combination with optical DNA mapping," *Scientific Reports*, vol. 6, no. 1, p. 37 938, Dec. 2016 (cit. on p. 116).

[279] P. Horvath and R. Barrangou, "CRISPR/Cas, the immune system of bacteria and archaea," en, *Science*, vol. 327, no. 5962, pp. 167–170, Jan. 2010 (cit. on p. 116).

[280] D. Šulskis, J. Thoma and B. M. Burmann, "Structural basis of degp protease temperature-dependent activation," *Science Advances*, vol. 7, no. 50, eabj1816, 2021. DOI: 10.1126/sciadv.abj1816. eprint: https://www.science.org/doi/pdf/10.1126/sciadv.abj1816. [Online]. Available: https://www.science.org/doi/abs/10.1126/sciadv.abj1816 (cit. on p. 116).

[281] M. M. Bronstein, J. Bruna, Y. LeCun, A. Szlam and P. Vandergheynst, "Geometric deep learning: Going beyond euclidean data," *IEEE Signal Processing Magazine*, vol. 34, no. 4, pp. 18–42, 2017. DOI: 10.1109/msp.2017.2693418. [Online]. Available: https://doi.org/10.1109\%2Fmsp.2017.2693418 (cit. on p. 116).

[282] G. Young, N. Hundt, D. Cole *et al.*, "Quantitative mass imaging of single biological macromolecules," *Science*, vol. 360, no. 6387, pp. 423–427, 2018. DOI: 10.1126/science.aar5839. eprint: https://www.science.org/doi/pdf/10.1126/science.aar5839. [Online]. Available: https://www.science.org/doi/abs/10.1126/science.aar5839 (cit. on p. 117).

[283] H. Klein Moberg, "Single nanoparticle catalysis-implementing machine learning algorithms for the identification of kinetic phase transitions on the surface of individual catalyst nanoparticles during surface chemical reactions through nanoplasmonic spectroscopy," *Chalmers ODR*, 2020 (cit. on p. 119).

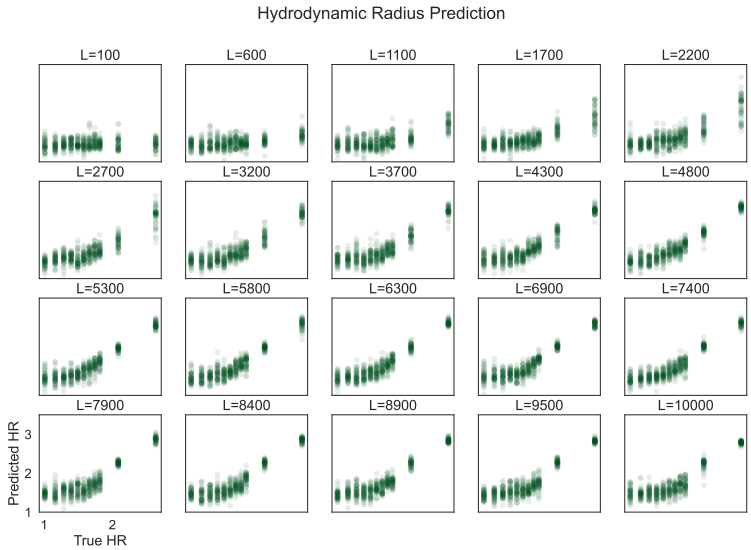

Figure 4: Predictions underlying the results in Figure 2 of the main text, consisting of predicting hydrodynamic radius of simulated molecules for $MW \in [**]$ kDa for reference channel Ch_B for trajectory lengths L between 100 and 10000 frames.

variance among all unbiased estimators for that parameter. The amount of data influences the Fisher information—generally, an increase in data leads to higher Fisher information, which implies a tighter bound and hence, a reduced potential estimation error.

5.0.1 Define the Statistical Model

We begin by specifying the probability density function (pdf) or probability mass function (pmf) of our data, including the scale parameter σ that we aim to estimate. The choice of model depends on the nature of our data (e.g., normal distribution, exponential distribution).

Example: Consider a normal distribution with mean zero and unknown scale (standard deviation) σ:

$$f(x;\sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x^2}{2\sigma^2}\right)$$

This model is appropriate for many natural phenomena due to the Central Limit Theorem [35].

5.0.2 Find the Likelihood Function

For a sample of independent and identically distributed observations $X = (X_1, X_2, \ldots, X_n)$, we construct the likelihood function by taking the product of individual pdfs:

$$L(\sigma; X) = \prod_{i=1}^{n} f(x_i; \sigma)$$

Example:

$$L(\sigma; X) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left(-\frac{\sum_{i=1}^n x_i^2}{2\sigma^2}\right)$$

5.0.3 Compute the Log-Likelihood Function

Next, we take the natural logarithm of the likelihood function to simplify differentiation:

$$\ell(\sigma; X) = \ln L(\sigma; X)$$

Example:

$$\ell(\sigma; X) = -n \ln \sigma - \frac{\sum_{i=1}^n x_i^2}{2\sigma^2} + \text{constant terms}$$

5.0.4 Calculate the First Derivative (Score Function)

We differentiate the log-likelihood function with respect to σ to obtain the score function:

$$\frac{\partial \ell(\sigma; X)}{\partial \sigma} = -\frac{n}{\sigma} + \frac{\sum_{i=1}^n x_i^2}{\sigma^3}$$

5.0.5 Compute the Second Derivative

We now differentiate the score function with respect to σ to get the observed information:

$$\frac{\partial^2 \ell(\sigma; X)}{\partial \sigma^2} = \frac{n}{\sigma^2} - \frac{3\sum_{i=1}^n x_i^2}{\sigma^4}$$

5.0.6 Calculate the Fisher Information $I(\sigma)$

The Fisher Information quantifies the amount of information that our observable data carries about the unknown parameter σ:

$$I(\sigma) = -\mathrm{E}\left[\frac{\partial^2 \ell(\sigma; X)}{\partial \sigma^2}\right]$$

Given that $X_i \sim N(0, \sigma^2)$ and $\mathrm{E}[X_i^2] = \sigma^2$, we have:

$$I(\sigma) = -\left(\frac{n}{\sigma^2} - \frac{3n\sigma^2}{\sigma^4}\right) = \frac{2n}{\sigma^2}$$

This result is consistent with standard statistical texts [36].

5.0.7 Compute the Cramér-Rao Lower Bound

Finally, the CRLB provides the minimum variance bound for any unbiased estimator $\hat{\sigma}$ of σ:

$$\mathrm{Var}(\hat{\sigma}) \geq \frac{1}{I(\sigma)} = \frac{\sigma^2}{2n}$$

This implies that no unbiased estimator of σ can have a variance lower than $\frac{\sigma^2}{2n}$, as established in estimation theory [34].

5.1 Cramér-Rao Lower Bound for NSM

In NSM, the intensity of scattered light from biomolecules diffusing in a nanofluidic channel provides insights into molecular weight and hydrodynamic radius. For NSM, the observed intensity I_t of the scattered light from a biomolecule within a nanochannel can be modeled as:

$$I_t = cI_0 L|\alpha_t|^2 \frac{k^3}{4}$$

where:

- $\alpha_t = \alpha_c + \frac{\alpha_m}{L^2}$ represents the total polarizability, composed of the polarizability of the nanochannel α_c and the biomolecule α_m.
- α_m, α_c: Polarizabilities of the biomolecule and the nanochannel.
- I_0: Incident light intensity.
- k: Wavenumber of the light.
- L: Length of the illuminated part of the nanochannel.
- c: Collection efficiency.

5.1.1 Assumptions for the Likelihood Function

Assume the measured intensity I_t is subject to Gaussian noise. Thus, the probability density function of observing I_t given the parameters α_m is modeled as:

$$f(I_t; \alpha_m) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(I_t - \mu(\alpha_m))^2}{2\sigma^2}\right)$$

where:

$$\mu(\alpha_m) = cI_0L\left(\alpha_c + \frac{\alpha_m}{L^2}\right)^2 \frac{k^3}{4}$$

and σ is the standard deviation of the measurement noise.

The log-likelihood function $\log f(I_t; \alpha_m)$ is:

$$\log f(I_t; \alpha_m) = -\frac{1}{2}\log(2\pi\sigma^2) - \frac{(I_t - \mu(\alpha_m))^2}{2\sigma^2}$$

5.1.2 Derivation of the Fisher Information

The Fisher Information $I(\alpha_m)$ is defined as the negative expected value of the second derivative of the log-likelihood with respect to the parameter α_m. In the Gaussian noise model with known variance, the Fisher Information simplifies to:

$$I(\alpha_m) = \frac{1}{\sigma^2}\left(\frac{\partial \mu(\alpha_m)}{\partial \alpha_m}\right)^2$$

Compute the first derivative of $\mu(\alpha_m)$ with respect to α_m:

$$\frac{\partial \mu(\alpha_m)}{\partial \alpha_m} = cI_0L \cdot 2\left(\alpha_c + \frac{\alpha_m}{L^2}\right) \cdot \left(\frac{1}{L^2}\right) \cdot \frac{k^3}{4}$$

Simplify:

$$\frac{\partial \mu(\alpha_m)}{\partial \alpha_m} = \frac{cI_0k^3}{2L}\left(\alpha_c + \frac{\alpha_m}{L^2}\right)$$

Note that $\alpha_t = \alpha_c + \frac{\alpha_m}{L^2}$, so:

$$\frac{\partial \mu(\alpha_m)}{\partial \alpha_m} = \frac{cI_0k^3}{2L}\alpha_t$$

Therefore, the Fisher Information is:

$$I(\alpha_m) = \frac{1}{\sigma^2}\left(\frac{cI_0k^3}{2L}\alpha_t\right)^2$$

5.1.3 Cramér-Rao Lower Bound (CRLB)

The CRLB provides a lower bound on the variance of any unbiased estimator of α_m:

$$\mathrm{Var}(\hat{\alpha}_m) \geq \frac{1}{I(\alpha_m)} = \frac{\sigma^2}{\left(\frac{cI_0k^3}{2L}\alpha_t\right)^2}$$

Simplify:

$$\mathrm{Var}(\hat{\alpha}_m) \geq \sigma^2 \left(\frac{2L}{cI_0k^3\alpha_t}\right)^2$$

This expression shows that the variance of the estimator depends on the true value of α_t, which includes the parameter α_m we aim to estimate.

5.1.4 Model for Multiple Measurements

For N independent measurements, the total Fisher Information is:

$$I_{\text{total}}(\alpha_m) = N \times I(\alpha_m)$$

Thus, the CRLB for estimating α_m from N measurements is:

$$\mathrm{Var}(\hat{\alpha}_m) \geq \frac{1}{I_{\text{total}}(\alpha_m)} = \frac{\sigma^2}{N}\left(\frac{2L}{cI_0k^3\alpha_t}\right)^2$$

5.1.5 CRLB for Diffusivity in NSM

The relationship between the diffusivity D and the hydrodynamic radius R_s is given by the Stokes-Einstein equation:

$$D = \frac{k_B T}{6\pi\eta R_s}$$

Considering the displacement x of a diffusing particle over time t, the probability density function is:

$$f(x; D) = \frac{1}{\sqrt{4\pi Dt}} \exp\left(-\frac{x^2}{4Dt}\right)$$

The log-likelihood function is:

$$\log f(x; D) = -\frac{1}{2}\log(4\pi Dt) - \frac{x^2}{4Dt}$$

Compute the first derivative with respect to D:

$$\frac{\partial}{\partial D}\log f = -\frac{1}{2D} + \frac{x^2}{4D^2t}$$

Compute the second derivative:

$$\frac{\partial^2}{\partial D^2}\log f = \frac{1}{2D^2} - \frac{x^2}{2D^3t}$$

Calculate the Fisher Information by taking the negative expected value of the second derivative. Since $E[x^2] = 2Dt$, we have:

$$I(D) = -E\left[\frac{\partial^2}{\partial D^2}\log f\right] = \frac{1}{2D^2}$$

Thus, the CRLB for a single measurement is:

$$\mathrm{Var}(\hat{D}) \geq \frac{1}{I(D)} = 2D^2$$

For N independent measurements:

$$I_{\text{total}}(D) = N \times I(D) = \frac{N}{2D^2}$$

So the CRLB becomes:

$$\mathrm{Var}(\hat{D}) \geq \frac{1}{I_{\text{total}}(D)} = \frac{2D^2}{N}$$

5.1.6 Final CRLB Expressions

$$\boxed{\mathrm{Var}(\hat{\alpha}_m) \geq \frac{\sigma^2}{N}\left(\frac{2L}{cI_0k^3\alpha_t}\right)^2}$$

$$\boxed{\mathrm{Var}(\hat{D}) \geq \frac{2D^2}{N}}$$

5.2 Including Localization Error in CRLB

So far, we have assumed a theoretically optimal but practically impossible localization error of 0. If we include the theoretical limits of localization error,for n independent displacement measurements over a total observation time $T = n\Delta t$, the CRLB is derived as follows.

Let σ_L be the standard deviation of the localization error. The measured position $x_m(t)$ at time t is:

$$x_m(t) = x(t) + \epsilon$$

where $x(t)$ is the true position, and ϵ is a zero-mean Gaussian random variable with variance σ_L^2.

The observed displacement Δx_m between two time points separated by Δt is:

$$\Delta x_m = x_m(t + \Delta t) - x_m(t) = \Delta x + \epsilon_2 - \epsilon_1$$

where $\Delta x = x(t + \Delta t) - x(t)$ is the true displacement due to diffusion.

The variance of the observed displacement Δx_m is:

$$\mathrm{Var}(\Delta x_m) = 2D\Delta t + 2\sigma_L^2$$

5.2.1 Likelihood Function and Fisher Information

Assuming that the observed displacements Δx_m are independent and normally distributed, the likelihood function for observing a displacement Δx_m given diffusivity D is:

$$f(\Delta x_m; D) = \frac{1}{\sqrt{4\pi(D\Delta t + \sigma_L^2)}} \exp\left(-\frac{(\Delta x_m)^2}{4(D\Delta t + \sigma_L^2)}\right)$$

The log-likelihood function is:

$$\log f(\Delta x_m; D) = -\frac{1}{2}\log\left(4\pi(D\Delta t + \sigma_L^2)\right) - \frac{(\Delta x_m)^2}{4(D\Delta t + \sigma_L^2)}$$

The first derivative with respect to D is:

$$\frac{\partial}{\partial D}\log f = -\frac{\Delta t}{2(D\Delta t + \sigma_L^2)} + \frac{(\Delta x_m)^2 \Delta t}{4(D\Delta t + \sigma_L^2)^2}$$

The second derivative is:

$$\frac{\partial^2}{\partial D^2}\log f = \frac{\Delta t^2}{2(D\Delta t + \sigma_L^2)^2} - \frac{(\Delta x_m)^2 \Delta t^2}{2(D\Delta t + \sigma_L^2)^3}$$

The Fisher information $I(D)$ from a single displacement measurement is the negative expected value of the second derivative:

$$I(D) = -E\left[\frac{\partial^2}{\partial D^2}\log f\right]$$

Since $E\left[(\Delta x_m)^2\right] = 2(D\Delta t + \sigma_L^2)$, we have:

$$I(D) = -\left(\frac{\Delta t^2}{2(D\Delta t + \sigma_L^2)^2} - \frac{2(D\Delta t + \sigma_L^2)\Delta t^2}{2(D\Delta t + \sigma_L^2)^3}\right) = \frac{\Delta t^2}{2(D\Delta t + \sigma_L^2)^2}$$

For n independent measurements:

$$I_{\text{total}}(D) = n \times I(D) = \frac{n\Delta t^2}{2(D\Delta t + \sigma_L^2)^2}$$

5.2.2 Cramér-Rao Lower Bound

The CRLB for diffusivity D, considering the localization error, is:

$$\text{Var}(\hat{D}) \geq \frac{1}{I_{\text{total}}(D)} = \frac{2(D\Delta t + \sigma_L^2)^2}{n\Delta t^2}$$

Simplifying:

$$\text{Var}(\hat{D}) \geq \frac{2}{n}\left(\frac{D\Delta t + \sigma_L^2}{\Delta t}\right)^2$$

The localization precision σ_L is related to the measurement noise σ and the experimental parameters:

$$\sigma_L = \frac{\sigma}{\sqrt{cI_0L|\alpha_t|^2\frac{k^3}{4}}}$$

Substituting σ_L into the CRLB expression:

$$\sigma_L^2 = \left(\frac{\sigma}{\sqrt{cI_0L|\alpha_t|^2\frac{k^3}{4}}}\right)^2 = \frac{4\sigma^2}{cI_0L|\alpha_t|^2k^3}$$

Thus, the CRLB becomes:

$$\mathrm{Var}(\hat{D}) \geq \frac{2}{n}\left(\frac{D\Delta t + \frac{4\sigma^2}{cI_0L|\alpha_t|^2k^3}}{\Delta t}\right)^2$$

Simplify the numerator inside the parentheses:

$$D\Delta t + \frac{4\sigma^2}{cI_0L|\alpha_t|^2k^3} = \Delta t\left(D + \frac{4\sigma^2}{cI_0L|\alpha_t|^2k^3\Delta t}\right)$$

Therefore, the CRLB simplifies to:

$$\mathrm{Var}(\hat{D}) \geq \frac{2}{n}\left(D + \frac{4\sigma^2}{cI_0L|\alpha_t|^2k^3\Delta t}\right)^2$$

5.2.3 Final CRLB Expression for Diffusivity in NSM

The modified CRLB for diffusivity D in NSM, incorporating the localization precision, is:

$$\boxed{\mathrm{Var}(\hat{D}) \geq \frac{2}{n}\left(D + \frac{4\sigma^2}{cI_0L|\alpha_t|^2k^3\Delta t}\right)^2}$$

The term $\frac{4\sigma^2}{cI_0L|\alpha_t|^2k^3\Delta t}$ represents the influence of localization error on the variance of $\hat{D}$. Reducing measurement noise σ or increasing the detected signal improves localization precision, thus reducing this term. Increasing Δt reduces the impact of localization error relative to diffusion, enhancing estimation precision. The variance decreases inversely with the number of measurements, emphasizing the benefit of collecting more data.

In the absence of localization error ($\sigma_L^2 = 0$), the CRLB reduces to:

$$\mathrm{Var}(\hat{D}) \geq \frac{2D^2}{n}$$

which matches the standard result derived without considering localization error.

References

[1] X. Huang, J. Fan, L. Li, *et al.*, "Fast, long-term, super-resolution imaging with hessian structured illumination microscopy," *Nature Biotechnology*, vol. 36, no. 5, pp. 451–459, May 2018, ISSN: 1546-1696. DOI: 10.1038/nbt.4115. [Online]. Available: https://doi.org/10.1038/nbt.4115 (cit. on p. 1).

[2] B. Špačková, H. Klein Moberg, J. Fritzsche, *et al.*, "Label-free nanofluidic scattering microscopy of size and mass of single diffusing molecules and nanoparticles," *Nature Methods*, vol. 19, no. 6, pp. 751–758, Jun. 2022, ISSN: 1548-7105. DOI: 10.1038/s41592-022-01491-6. [Online]. Available: https://doi.org/10.1038/s41592-022-01491-6 (cit. on p. 1).

[3] W. Zhao, S. Zhao, L. Li, *et al.*, "Sparse deconvolution improves the resolution of live-cell super-resolution fluorescence microscopy," *Nature Biotechnology*, vol. 40, no. 4, pp. 606–617, Apr. 2022, ISSN: 1546-1696. DOI: 10.1038/s41587-021-01092-2. [Online]. Available: https://doi.org/10.1038/s41587-021-01092-2 (cit. on p. 1).

[4] K. de Haan, Y. Rivenson, Y. Wu, and A. Ozcan, "Deep-learning-based image reconstruction and enhancement in optical microscopy," *Proceedings of the IEEE*, vol. 108, no. 1, pp. 30–50, 2020. DOI: 10.1109/JPROC.2019.2949575 (cit. on p. 2).

[5] C. Belthangady and L. A. Royer, "Applications, promises, and pitfalls of deep learning for fluorescence image reconstruction," *Nature Methods*, vol. 16, no. 12, pp. 1215–1225, Dec. 2019, ISSN: 1548-7105. DOI: 10.1038/s41592-019-0458-z. [Online]. Available: https://doi.org/10.1038/s41592-019-0458-z (cit. on p. 2).

[6] N. Rahaman, A. Baratin, D. Arpit, F. Draxler, M. Lin, F. Hamprecht, Y. Bengio, and A. Courville, "On the spectral bias of neural networks," in *Proceedings of the 36th International Conference on Machine Learning*, K. Chaudhuri and R. Salakhutdinov, Eds., ser. Proceedings of Machine Learning Research, vol. 97, PMLR, Sep. 2019, pp. 5301–5310. [Online]. Available: https://proceedings.mlr.press/v97/rahaman19a.html (cit. on p. 2).

[7] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, ISSN: 1476-4687. DOI: 10.1038/nature14539. [Online]. Available: https://doi.org/10.1038/nature14539 (cit. on p. 2).

[8] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Computer Vision – ECCV 2014*, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds., Cham: Springer International Publishing, 2014, pp. 818–833, ISBN: 978-3-319-10590-1 (cit. on pp. 2, 6).

[9] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, "Deep learning for computer vision: A brief review," *Computational Intelligence and Neuroscience*, vol. 2018, p. 7 068 349, Feb. 2018, ISSN: 1687-5265. DOI: 10.1155/2018/7068349. [Online]. Available: https://doi.org/10.1155/2018/7068349 (cit. on pp. 2, 3).

[10] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Pro-*

cessing Systems, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds., vol. 25, Curran Associates, Inc., 2012. [Online]. Available: https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf (cit. on p. 3).

[11] R. Girshick, J. Donahue, T. Darrell, and J. Malik, *Rich feature hierarchies for accurate object detection and semantic segmentation*, 2013. DOI: 10.48550/ARXIV.1311.2524. [Online]. Available: https://arxiv.org/abs/1311.2524 (cit. on p. 3).

[12] D. Heaven, "Why deep-learning AIs are so easy to fool," *Nature*, vol. 574, no. 7777, pp. 163–166, Oct. 2019. DOI: 10.1038/d41586-019-03013-5. [Online]. Available: https://doi.org/10.1038/d41586-019-03013-5 (cit. on p. 3).

[13] A. D'Amour, K. Heller, D. Moldovan, *et al.*, *Underspecification presents challenges for credibility in modern machine learning*, 2020. DOI: 10.48550/ARXIV.2011.03395. [Online]. Available: https://arxiv.org/abs/2011.03395 (cit. on pp. 3, 6).

[14] S. Kaplan, D. Handelman, and A. Handelman, "Sensitivity of neural networks to corruption of image classification," *AI and Ethics*, vol. 1, no. 4, pp. 425–434, Nov. 2021, ISSN: 2730-5961. DOI: 10.1007/s43681-021-00049-0. [Online]. Available: https://doi.org/10.1007/s43681-021-00049-0 (cit. on p. 3).

[15] K. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey of transfer learning," *Journal of Big Data*, vol. 3, no. 1, p. 9, May 2016, ISSN: 2196-1115. DOI: 10.1186/s40537-016-0043-6. [Online]. Available: https://doi.org/10.1186/s40537-016-0043-6 (cit. on p. 5).

[16] M. D. Zeiler and R. Fergus, *Visualizing and understanding convolutional networks*, 2013. DOI: 10.48550/ARXIV.1311.2901. [Online]. Available: https://arxiv.org/abs/1311.2901 (cit. on p. 5).

[17] L. Rokach, "Ensemble methods for classifiers," in *Data Mining and Knowledge Discovery Handbook*, O. Maimon and L. Rokach, Eds. Boston, MA: Springer US, 2005, pp. 957–980, ISBN: 978-0-387-25465-4. DOI: 10.1007/0-387-25465-X_45. [Online]. Available: https://doi.org/10.1007/0-387-25465-X_45 (cit. on p. 6).

[18] F. Chollet *et al.*, *Keras*, https://keras.io, 2015 (cit. on p. 6).

[19] O. Sagi and L. Rokach, "Ensemble learning: A survey," *WIREs Data Mining and Knowledge Discovery*, vol. 8, no. 4, e1249, 2018. DOI: https://doi.org/10.1002/widm.1249. eprint: https://wires.onlinelibrary.wiley.com/doi/pdf/10.1002/widm.1249. [Online]. Available: https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/widm.1249 (cit. on p. 6).

[20] F. Chollet, *Deep Learning with Python*. Manning, Nov. 2017, ISBN: 9781617294433 (cit. on p. 6).

[21] Y. Bengio, J. Louradour, R. Collobert, and J. Weston, "Curriculum learning," in *Proceedings of the 26th Annual International Conference on Machine Learning*, ser. ICML '09, Montreal, Quebec, Canada: Association for Computing Machinery, 2009, pp. 41–48, ISBN: 9781605585161. DOI: 10.1145/1553374.

1553380. [Online]. Available: https://doi.org/10.1145/1553374.1553380 (cit. on p. 6).
[22] M. Csikszentmihalyi, "Flow: The psychology of optimal experience," in Jan. 1990 (cit. on p. 6).
[23] B. Settles, "Active learning literature survey," 2009 (cit. on p. 6).
[24] A. Ghosh, B. G. Sumpter, O. Dyck, S. V. Kalinin, and M. Ziatdinov, "Ensemble learning-iterative training machine learning for uncertainty quantification and automated experiment in atom-resolved microscopy," *npj Computational Materials*, vol. 7, no. 1, p. 100, Jul. 2021, ISSN: 2057-3960. DOI: 10.1038/s41524-021-00569-7. [Online]. Available: https://doi.org/10.1038/s41524-021-00569-7 (cit. on p. 7).
[25] B. Midtvedt, S. Helgadottir, A. Argun, J. Pineda, D. Midtvedt, and G. Volpe, "Quantitative digital microscopy with deep learning," *Applied Physics Reviews*, vol. 8, no. 1, p. 011 310, Feb. 2021, ISSN: 1931-9401. DOI: 10.1063/5.0034891. eprint: https://pubs.aip.org/aip/apr/article-pdf/doi/10.1063/5.0034891/14577703/011310_1_online.pdf. [Online]. Available: https://doi.org/10.1063/5.0034891 (cit. on p. 7).
[26] D. P. Kingma and J. Ba, *Adam: A method for stochastic optimization*, 2014. DOI: 10.48550/ARXIV.1412.6980. [Online]. Available: https://arxiv.org/abs/1412.6980 (cit. on p. 8).
[27] O. Ronneberger, P.Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, ser. LNCS, (available on arXiv:1505.04597 [cs.CV]), vol. 9351, Springer, 2015, pp. 234–241. [Online]. Available: http://lmb.informatik.uni-freiburg.de/Publications/2015/RFB15a (cit. on p. 9).
[28] L. Roeder, *Netron*, https://github.com/lutzroeder/netron, 2020 (cit. on p. 9).
[29] Y. Rivenson, T. Liu, Z. Wei, Y. Zhang, K. de Haan, and A. Ozcan, "Phasestain: The digital staining of label-free quantitative phase microscopy images using deep learning," *Light: Science & Applications*, vol. 8, no. 1, p. 23, Feb. 2019, ISSN: 2047-7538. DOI: 10.1038/s41377-019-0129-y. [Online]. Available: https://doi.org/10.1038/s41377-019-0129-y (cit. on p. 9).
[30] P. Dechadilok and W. M. Deen, "Hindrance factors for diffusion and convection in pores," *Industrial & engineering chemistry research*, vol. 45, no. 21, pp. 6953–6959, 2006 (cit. on p. 8).
[31] P. Virtanen, R. Gommers, T. E. Oliphant, *et al.*, "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python," *Nature Methods*, vol. 17, pp. 261–272, 2020. DOI: 10.1038/s41592-019-0686-2 (cit. on p. 8).
[32] H. Cramér, *Mathematical Methods of Statistics*. Princeton University Press, 1946 (cit. on p. 11).
[33] C. R. Rao, "Information and the accuracy attainable in the estimation of statistical parameters," *Bulletin of the Calcutta Mathematical Society*, vol. 37, pp. 81–91, 1945 (cit. on p. 11).

[34] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*. Prentice Hall, 1993 (cit. on pp. 11, 14).

[35] W. Feller, *An Introduction to Probability Theory and Its Applications*, 3rd. John Wiley & Sons, 1968, vol. 1 (cit. on p. 12).

[36] E. L. Lehmann and G. Casella, *Theory of Point Estimation*, 2nd. Springer, 1998 (cit. on p. 14).