Data augmentation with Mobius transformations

(article starts on next page)

PAPER • OPEN ACCESS

Data augmentation with Mobius transformations

View the article online for updates and enhancements.

MACHINE
LEARNING
Science and Technology

PAPER

OPEN ACCESS

Data augmentation with Mobius transformations

Sharon Zhou[1] , Jiequan Zhang[1], Hang Jiang[1], Torbjörn Lundh[2] and Andrew Y Ng[1]

[1] Stanford University, Stanford, CA 94305, United States of America
[2] Chalmers University of Technology, Gothenburg 412 96, Sweden

E-mail: sharonz@cs.stanford.edu

Abstract

Data augmentation has led to substantial improvements in the performance and generalization of deep models, and remains a highly adaptable method to evolving model architectures and varying amounts of data—in particular, extremely scarce amounts of available training data. In this paper, we present a novel method of applying Möbius transformations to augment input images during training. Möbius transformations are bijective conformal maps that generalize image translation to operate over complex inversion in pixel space. As a result, Möbius transformations can operate on the sample level and preserve data labels. We show that the inclusion of Möbius transformations during training enables improved generalization over prior sample-level data augmentation techniques such as cutout and standard crop-and-flip transformations, most notably in low data regimes.

1. Introduction

Data augmentation has significantly improved the generalization of deep neural networks on a variety of image tasks, including image classification [1, 2], object detection [3, 4], and instance segmentation [5]. Prior work has shown that data augmentation on its own can perform better than, or on par with, highly regularized models using other regularization techniques such as dropout [6]. This effectiveness is especially prominent in low data regimes, where models often fail to capture the full variance of the data in the training set [7].

Many data augmentation techniques rely on priors that are present in the natural world. Standard operations, such as translation, crop, and rotation, in addition to more recent methods, such as cutout [2], have improved generalization by encouraging equivariance to the transformation. For example, an image of a horse remains a horse in its vertical reflection or with its body partially occluded. As a result, these transformations are able to preserve the original label on the augmented sample, enabling straightforward and easy incorporation into the growing number of data augmentation algorithms for both fully supervised and semi-supervised learning. In a nutshell, these sample-level methods have been not only effective, but also interpretable, easy to implement, and flexible to incorporate.

Following the success of these methods, we focus this paper on augmentations that exploit natural patterns to preserve labels after transformation and that operate on the sample level. These transformations easily complement other methods, thus leveraged in a wide variety of data augmentation algorithms. In contrast, multi-sample augmentations, which have had comparably strong empirical results [8], unfortunately connect less clearly to natural priors that would support equivariance to the augmentation. While performant on their own, these methods have had less success with integration into data augmentation algorithms and policies [9–11], except for those tailored to them [12].

In this paper, we propose a novel data augmentation technique, inspired by biological patterns, using bijective conformal maps known as *Möbius transformations*. Möbius transformations perform complex inversion in pixel space, extending standard translation to include divisibility. These transformations enable perspective projection—or transforming perceived distance of objects in an image—and are found naturally in the anatomy and biology of humans and other animals.

Figure 1. Examples of Möbius transformations (original on left), resulting in variations in perspective, orientation, and scale, while still preserving local angles and anharmonic ratios.

We define a class of $\mathcal{M}$-admissible Möbius transformations that preserves image-level labels by minimizing local distortions in an image. We show empirically that the inclusion of $\mathcal{M}$-admissible Möbius transformations can improve performance on the CIFAR-10, CIFAR-100, and Tiny ImageNet benchmarks over prior sample-level data augmentation techniques, such as cutout [2] and standard crop-and-flip baselines. We additionally show that Möbius transformations successfully complement other transformations.

Our key contributions can be summarized as follows.

- **Method:** We introduce a class of $\mathcal{M}$-admissible Möbius transformations for data augmentation in training neural networks. This Möbius class allows for a wide range of sample-level mappings that preserve local angles and can be found in the anatomy of animals.
- **Performance:** Empirically, the inclusion of Möbius data augmentation improves model generalization over prior methods that use sample-level augmentation techniques, such as cutout [2] and standard crop-and-flip transformations. We also show that Möbius transformations, which have been studied and examined in the anatomy and biology of animals, consistently improves on animate classes over inanimate classes.
- **Low data:** Möbius is especially effective in low data settings, where the data quantity is on the order of hundreds of samples per class.

2. Möbius transformations

Möbius transformations are bijective conformal mappings that operate over complex inversion and preserve local angles. They are also known as bilinear or linear fractional transformations. We discuss their biological and perceptual underpinnings, and follow with a formal definition. Finally, we describe their application to data augmentation to improve generalization of convolutional neural networks on image classification tasks.

2.1. Motivation
Möbius transformations have been studied in biology as 2D projections of specimens—such as humans, fungi, and fish—from their 3D configurations [13–15]. Mathematically, most of these examples leverage Liouville's theorem [16], which states that smooth conformal mappings are Möbius transformations on a domain of $\mathbb{R}^n$ where $n > 2$. These biological patterns motivate our application of Möbius transformations to natural images, particularly those that include the relevant species.

Beyond biological underpinnings, Möbius transformations preserve the *anharmonic ratio* [17, 18], or the extent to which four collinear points on a projective line deviate from the harmonic ratio[3]. This invariance is a property that Möbius transformations share with projective transformations, which are used widely in metrology [19]. In the context of transforming natural images, such a transformation can be particularly useful for perspective projection. That is, an image can be transformed to an alternate perceived distance. This effect is visually apparent across examples in figure 1.

2.2. Definition
Existing data augmentation techniques for image data belong to the class of affine mappings, i.e. the group of translation, scaling, and rotation, which can be generally described using a complex function $z \to az + b$, where the variable z and the two parameters a and b are complex numbers. Möbius transformations

---

[3] The anharmonic ratio, also denoted *cross-ratio*, stems from projective geometry and has been studied in biology with respect to Möbius transformations [14, 15].

represent the next level of abstraction by introducing division to the operation [14, 20]. The group of Möbius transformations can be described as all functions f from $\mathbb{C} \to \mathbb{C}$ with the form

$$f(z) = \frac{az+b}{cz+d}, \tag{1}$$

where $a, b, c, d \in \mathbb{C}$ such that $ad - bd \neq 0$. As a result, the set of all Möbius transformations is a superset of several basic transformations, including translation, rotation, inversion, and an even number of reflections over lines.

One method for programatically implementing such a transformation with complex values a, b, c, d in equation (1) is to use the fact that there exists a unique Möbius transformation sending any three points to any three other points in the extended complex plane [18, p 150]. That is, equivalent to specifying a, b, c and d directly in (1), we can define three separate points $z_1, z_2, z_3 \in \mathbb{C}$ in the image and then select three separate target points $w_1, w_2, w_3 \in \mathbb{C}$, to which those initial points will be mapped in the resulting transformation. From these two sets of points, we can then compute the values of the transformation using the knowledge that anharmonic ratios—adding the points z_i and w_i where $i = \{1, 2, 3\}$ completes the two quartets—are Möbius invariant [18, p 154], resulting in the following equality:

$$\frac{(w-w_1)(w_2-w_3)}{(w-w_3)(w_2-w_1)} = \frac{(z-z_1)(z_2-z_3)}{(z-z_3)(z_2-z_1)}. \tag{2}$$

We can rearrange this expression by solving for w:

$$\frac{w-w_1}{w-w_3} = \frac{(z-z_1)(z_2-z_3)(w_2-w_1)}{(z-z_3)(z_2-z_1)(w_2-w_3)}$$

$$w = \frac{Aw_3 - w_1}{A-1}$$

where $A = \frac{(z-z_1)(z_2-z_3)(w_2-w_1)}{(z-z_3)(z_2-z_1)(w_2-w_3)}$. This final expression for w is in the form of equation (1):

$$f(z) = w = \frac{Aw_3 - w_1}{A-1} = \frac{az+b}{cz+d},$$

from which we can compute the following values for $a, b, c,$ and d using basic algebraic operations:

$$a = w_1 w_2 z_1 - w_1 w_3 z_1 - w_1 w_2 z_2 + w_2 w_3 z_2 + w_1 w_3 z_3 - w_2 w_3 z_3,$$

$$b = w_1 w_3 z_1 z_2 - w_2 w_3 z_1 z_2 - w_1 w_2 z_1 z_3 + w_2 w_3 z_1 z_3 + w_1 w_2 z_2 z_3 - w_1 w_3 z_2 z_3,$$

$$c = w_2 z_1 - w_3 z_1 - w_1 z_2 + w_3 z_2 + w_1 z_3 - w_2 z_3,$$

$$d = w_1 z_1 z_2 - w_2 z_1 z_2 - w_1 z_1 z_3 + w_3 z_1 z_3 + w_2 z_2 z_3 - w_3 z_2 z_3.$$

Alternatively, by solving equation (2) using linear algebra, i.e. evaluating a determinant from this construction using the Laplace expansion, one can elegantly express these algebraic expressions above as determinants:

$$a = \begin{vmatrix} z_1 w_1 & w_1 & 1 \\ z_2 w_2 & w_2 & 1 \\ z_3 w_3 & w_3 & 1 \end{vmatrix}, \quad b = \begin{vmatrix} z_1 w_1 & z_1 & w_1 \\ z_2 w_2 & z_2 & w_2 \\ z_3 w_3 & z_3 & w_3 \end{vmatrix}, \quad c = \begin{vmatrix} z_1 & w_1 & 1 \\ z_2 & w_2 & 1 \\ z_3 & w_3 & 1 \end{vmatrix}, \quad d = \begin{vmatrix} z_1 w_1 & z_1 & 1 \\ z_2 w_2 & z_2 & 1 \\ z_3 w_3 & z_3 & 1 \end{vmatrix}.$$

This pointwise method is used in our work to construct valid image augmentations using Möbius transformations. Ultimately, this method can be leveraged to define specific types of Möbius transformations programmatically for needs within and beyond data augmentation.

2.2.1. Equivalent framing: circle reflection

We introduce an equivalent formulation of Möbius transformations on images in $\mathbb{R}^2$. The goal of this section is to lend intuition on constraints that we apply to Möbius data augmentation in section 3 that follows.

Möbius mappings in the plane can also be defined as the set of transformations with an even number of reflections over circles and lines (i.e. circles with infinite radii) on the plane. A reflection, or inversion, in the unit circle is the complex transformation [18, p 124]:

$$z \in \mathbb{C} \to \frac{z}{|z|^2}.$$

Thus, a Möbius transformation on an image is simply a reflection over the unit circle, with pixels inside of the circle projected outwards and pixels on the outside projected inwards. As such, Möbius transformations often reflect a different amount of pixels inwards as opposed to outwards, and this imbalance enables the scale distortions seen in figure 1. Note that a circular shape can be left as an artifact after the transformation, if the reflection occurs at an edge without any pixels to project inwards.

3. Class of $\mathcal{M}$-admissible Möbius transformations

In order to use Möbius transformations for data augmentation, we need to constrain the set of possible transformations. When taken to the limit, Möbius transformations do not necessarily preserve the image label. This is similar to constraining translation in order to ensure that pixels remain afterwards, or to keeping cutout to lengths judiciously less than the size of the image so that it is not fully occluded. Because Möbius transformations inherently reflect more pixels in one direction (into or out of the circle), we will often see two main effects: (1) incongruent sizes of the output from the initial input and (2) gaps between pixels in the result transformation, sometimes significant depending on the location of the circle. For example, if the circle is placed at the edge of the image, there is little to project from the edge inwards. To address both of these effects, we enforce equal sizing after the transformation and cubic spline interpolation during reflection to fill gaps.

In tandem, we introduce a class of $\mathcal{M}$-admissible Möbius transformations that control the local distortion in an image, in order to avoid explosive and implosive mappings, by bounding the modulus of the derivative above and below. If we view each pixel as a circle to be mapped at another circle (pixel) and use the analogue of the $f^{\#}$-function, as defined in an authoritative text on Möbius [21], to the modulus of the derivative of the Möbius transformation, $|f'|$, we can bound it above and below by two constants, or for simplicity, one real constant $M > 1$ such that

$$\frac{1}{M} < |f'| < M. \tag{3}$$

As an approximation, we will only check this condition for only five points on an image of size $[0, p] \times [0, pi]$: the four inverse images of the corner points in the square $[0, p] \times [0, pi]$ and the center point: $\frac{1}{2}p(1 + i)$. Furthermore, in order to only consider transformations that will keep enough information from the original picture, we add the condition that the pre-image of the center point, $f^{-1}(\frac{1}{2}p(1 + i))$, should be inside the centered circle half-way to the sides, i.e.

$$|f^{-1}\left(\frac{1}{2}p(1 + i)\right) - \frac{1}{2}p(1 + i)| < \frac{p}{4}]. \tag{4}$$

To give more concrete and computable conditions (3) and (4), we start with a general Möbius transformation from definition (1) with the condition that $ad \neq bc$ to obtain the inverse f^{-1} as

$$f^{-1}(z) = -\frac{dz - b}{cz - a}, \tag{5}$$

For the condition in (4), we compute (5)

$$f^{-1}\left(\frac{1}{2}p(1 + i)\right) = \frac{\frac{1}{2}p(1 + i)d - b}{a - \frac{1}{2}p(1 + i)c}.$$

For the condition (3), we compute the derivative f' :

$$f'(z) = \frac{a}{cz + d} - \frac{c(az + b)}{(cz + d)^2}. \tag{6}$$

Combining (5) and (6) and simplifying, we obtain the following simple expression:

$$f'(f^{-1}(z)) = \frac{(a - cz)^2}{ad - bc}. \tag{7}$$

By using (7) and (5), we can give a reformulation of the conditions in (3) and (4) to define a subclass of all Möbius transformations,

$$f(z) = \frac{az + b}{cz + d}, \text{ where } ad - bc \neq 0,$$

which we call the class of $\mathcal{M}$-**admissible Möbius transformations** as long as the function f fulfills the following list of inequalities by checking the points $0, p, pi, p(1 + i), \frac{1}{2}p(1 + i)$:

$$\frac{1}{M} < \frac{|a|^2}{|ad - bc|} < M, \qquad\qquad \frac{1}{M} < \frac{|a - pc|^2}{|ad - bc|} < M,$$

$$\frac{1}{M} < \frac{|pc + ai|^2}{|ad - bc|} < M, \qquad\qquad \frac{1}{M} < \frac{|a - p(1 + i)c|^2}{|ad - bc|} < M,$$

$$\frac{1}{M} < \frac{|a - \frac{1}{2}p(1 + i)c|^2}{|ad - bc|} < M, \qquad\qquad \left| \frac{\frac{1}{2}p(1 + i)d - b}{a - \frac{1}{2}p(1 + i)c} - \frac{1}{2}p(1 + i) \right| < \frac{p}{4}.$$

Sampling from the $\mathcal{M}$-admissible class, we can incorporate label-preserving Möbius transformations into classical data augmentation methods of the form $(x, y) = (f(x), y)$, where f here is a Möbius transformation on an image x, preserving label y.

4. Related work

A large number of data augmentation techniques have recently emerged for effectively regularizing neural networks, including both sample-level augmentations, such as ours, as well as multi-sample augmentations that mix multiple images. We discuss these, as well as data augmentation algorithms that leverage multiple augmentations. Finally, we examine ways in which Möbius transformations have been applied to deep learning. To our knowledge, this is the first work using Möbius transformations for data augmentation in deep neural networks.

4.1. Data augmentation
4.1.1. Sample-level augmentation
Möbius transformations generalize standard translation to include inversion as an operation under conformity, demonstrating outputs that appear to have gone through crop, rotation, and/or scaling, while preserving local angles from the original image. We recognize that the list of image transformations is extensive: crop, rotation, warp, skew, shear, distortion, Gaussian noise, among many others. Additional sample-level data augmentation methods use occlusion such as cutout [2] and random erasing [3], which apply random binary masks across image regions. Finally, there is adjacent work that follows the aim of data augmentation methods to learn invariant representations, and directly learns network architectures that are invariant to select transformations, such as rotation invariance [22, 23].

4.1.2. Multi-sample augmentation
Data augmentation on images also consists of operations applied to multiple input images. In such cases, original labels are often mixed. For example, MixUp [8] performs a weighted average of two images (over pixels) and their corresponding labels in varying proportions to perform soft multi-label classification. Between-class learning [24] and SamplePairing [25] are similar techniques, though the latter differs in using a single label. Comparably, RICAP [26], VH-Mixup and VH-BC+ [27] form composites of several images into one. While these methods have performed well, we focus this paper on comparisons to sample-level augmentations that preserve original labels and that can be more readily incorporated into data augmentation policies.

4.1.3. Algorithms and policies for data augmentation
Various strategies have emerged to incorporate multiple data augmentation techniques for improved performance. AutoAugment [9], Adatransform [28], RandAugment [10], and Population Based Augmentation [29] offer ways to select optimal transformations (and their intensities) during training. In semi-supervised learning, unsupervised data augmentation [11], MixMatch [12], and FixMatch [30] have shown to effectively incorporate unlabeled data by exploiting label preservation and consistency training. Tanda [31] composes sequences of augmentation methods, such as crop followed by cutout then flip, that are tuned to a certain domain. DADA [7] frames data augmentation as an adversarial learning problem and applies this method in low data settings. We do not test all of these augmentation schemes: our results suggest that Möbius could add value as an addition to the search space of possible augmentations, e.g. in AutoAugment, or as a transformation that helps enforce consistency between original and augmented data, e.g. in unsupervised data augmentation.

4.2. Möbius transformations in deep learning

Möbius transformations have been previously studied across a handful of topics in deep learning. Specifically, they have been used as building blocks in new activation functions [32] and as operations in hidden layers [33]. Coupled with the theory of gyrovector spaces, Möbius transformations have inspired hyperbolic neural networks [34]. They also play an important component in deep fuzzy neural networks for approximating the Choquet integral [35]. Finally, model activations and input–output relationships have been theoretically related to Möbius transformations [36]. While prior work has primarily leveraged them for architectural contributions, our work is the first to our knowledge to introduce Möbius transformations for data augmentation and their empirical success on image classification benchmarks.

5. Experiments

We experiment on CIFAR-10, CIFAR-100, and Tiny ImageNet. The CIFAR-10 and CIFAR-100 image classification benchmarks use standard data splits of 50k training and 10k test [37]. CIFAR-10 has 10 classes with 10k images per class, while CIFAR-100 has 100 classes with 500 images per class, in their training sets. Finally, we experiment on Tiny ImageNet [38], a subset of ImageNet that still includes ImageNet's variability and higher resolution imagery, while needing fewer resources and infrastructure than running the full ImageNet dataset. The training set constitutes 100k images across 200 classes, and the test set contains 10k images.

Thus, we explore three dataset settings: (1) CIFAR-10, (2) CIFAR-100, and (3) Tiny ImageNet. The goal of these experiments is to assess the fundamental concept of including Möbius data augmentation across data settings.

5.1. Evaluation of benchmarks

Following prior work on introducing novel data augmentation methods [2, 9], we use standard crop-and-flip transformations as the baseline across all experimental conditions. We design our experiments to both compare to, and complement, cutout [2], the previous state-of-the-art image transformation that operates on the sample level, preserves labels, and thus has been easy to incorporate into data augmentation policies. Cutout and standard crop-and-flip also remain the default augmentation choices in recent work [9]. Thus, we compare the following conditions: (1) baseline with only crop and flip, (2) cutout, (3) Möbius, and (4) Möbius with cutout. Note that all conditions incorporate crop and flip transformations, following the original cutout paper [2]. Because all augmentation techniques are sample-level and preserve labels, they are complementary and can be layered on each other. We further explore these effects by combining Möbius with cutout in our experiments.

We draw from prior work on cutout [2] to set the training procedure across all experiments. Specifically, we use 0.1 learning rate, cosine annealing learning rate scheduler, stochastic gradient descent for optimization, 200 epochs and a standard wide residual network [39] on CIFAR, and 100 epochs and a standard residual network on Tiny ImageNet. For cutout, we tune hyperparameter values for each dataset based on prior work [2]. Note that we select this setup to optimize for cutout and compare directly to their work; it is possible that better hyperparameters exist for Möbius. We incorporate Möbius augmentation 20% of the time in all experiments, and show further improvements varying its inclusion on different data settings in the appendix. On the Tiny ImageNet dataset, for which cutout did not present a baseline, we use a standard residual network, average across three runs, and train on two NVIDIA Tesla V100 GPUs. Training time is 9.565 GPU hours. Finally, we compute significance using independent t-tests between sample performances of pairwise conditions across runs.

As shown in table 1, these experiments highlight several key observations.

Table 1. Experimental results on several dataset settings, including CIFAR-10, CIFAR-100, and Tiny ImageNet. Möbius performs best empirically on low data settings, such as Tiny ImageNet and CIFAR-100, where the number of images per class is on the order of hundreds. Statistical significance indicated in bold.

Augmentation method	Dataset	Images per class	# Training images	Accuracy
Crop-and-flip	CIFAR-10	5000	50k	$96.47\% \pm 0.04$
Cutout$_{l=16}$	CIFAR-10	5000	50k	$\mathbf{97.13\% \pm 0.03}$
Möbius	CIFAR-10	5000	50k	$96.67\% \pm 0.13$
Möbius + Cutout$_{l=16}$	CIFAR-10	5000	50k	$\mathbf{97.10\% \pm 0.16}$
Crop-and-flip	CIFAR-100	600	50k	$81.91\% \pm 0.20$
Cutout$_{l=8}$	CIFAR-100	600	50k	$82.35\% \pm 0.19$
Möbius	CIFAR-100	600	50k	$82.48\% \pm 0.38$
Möbius + Cutout$_{l=8}$	CIFAR-100	600	50k	$\mathbf{82.67\% \pm 0.21}$
Crop-and-flip	Tiny ImageNet	500	100k	$68.40\% \pm 0.33$
Cutout$_{l=16}$	Tiny ImageNet	500	100k	$68.64\% \pm 0.40$
Möbius	Tiny ImageNet	500	100k	$69.04\% \pm 0.42$
Möbius + Cutout$_{l=16}$	Tiny ImageNet	500	100k	$\mathbf{69.51\% \pm 0.25}$

- Empirically, Möbius is able to achieve a higher accuracy than cutout on average.
- Möbius with cutout significantly outperforms all other conditions on CIFAR-100 and Tiny ImageNet, suggesting that the two techniques are complementary. This is important, as we designed Möbius to combine easily with other augmentation methods.
- In particular, this improvement is apparent where the number of images per class is small (on the order of hundreds). On CIFAR-10, however, where the number of images per class is an order of magnitude larger, cutout shows significant improvement over Möbius.
- We speculate that Mobius does best when there is a lack of variance within classes, as it provides considerable regularization to the dataset. CIFAR-10 naturally contains generous variance within each class, while CIFAR-100, which is the same dataset but with an order of magnitude more classes, contains naturally more variance per class. Based on significance tests, cutout and Mobius+cutout performed on par, without significant differences.

The results of this experiment suggest that Möbius data augmentation can improve over cutout, the state-of-the-art performance on sample-level and label-preserving augmentation strategy. This effect is especially prominent in low data regimes, where there are fewer (on the order of hundreds) of samples per class. Provided that the distortions generated by Möbius are highly varied, we expect, and observe empirically, that Möbius data augmentation performs significantly better on the larger image resolutions of Tiny ImageNet over CIFAR10 or CIFAR100. The increased number of pixels permits a greater diversity of available $\mathcal{M}$-admissible Möbius transformations.

5.2. Analysis on animate classes

We analyze the predictions from Möbius data augmentation by superclass in each dataset (table 2). Specifically, we compare two best model checkpoints trained on CIFAR-100: one with Möbius and the other with the standard crop-and-flip baseline. In CIFAR-100, there are 20 superclasses with higher-level aggregations of five classes each [37]. In our analysis, we find that the Möbius-trained model improves performance on the 10 animate superclasses {aquatic mammals, fish, insects, large omnivores and herbivores, large carnivores, non-insect invertebrates, medium-sized mammals, people, small mammals, reptiles}. This contrasts with inconsistent performance differences among the inanimate superclasses.
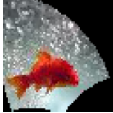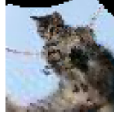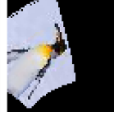
We perform a similar analysis by averaging the results of five baseline and Möbius models on Tiny ImageNet, whose superclasses are derived from the ImageNet category tree[4]. We compare animate superclasses {vertebrate, arthropod, coelenterate, shellfish, echinoderm} with inanimate superclasses {artifact, geological formation, natural object, miscellaneous}. We find that Möbius improves performance significantly more on animate classes than on inanimate classes compared to the standard baseline.

These results suggest that Möbius transformations, which have been studied in animals in prior literature [13–15], are especially effective on animate classes in image classification. While this finding is particularly interesting and consistent with scholarship, we heed that this observation remains empirical to this study and requires additional examination in order to be conclusive.

---

[4] http://www.image-net.org/api/xml/structure_released.xml.

Table 2. Analysis of animate and inanimate superclasses, shown with animate {goldfish, swan, cat, penguin} and inanimate {car, pizza, teapot, lighthouse} image samples. Möbius data augmentation consistently improves classification accuracy on animate superclasses (✓ on ↑), as opposed to inanimate superclasses on CIFAR (upper rows) and Tiny ImageNet (lower). On the Tiny ImageNet, Möbius transformations improve performance on animate classes significantly more than inanimate classes. This empirical observation suggests that Möbius transformations, having been studied in animals, are particularly attuned to improving generalization in these classes. Note that this finding is interesting, though by no means definitive or theoretically potent.

Animate	Baseline	Möbius	↑	Inanimate	Baseline	Möbius	↑
Aquatic Mammals	73.0%	74.4%	✓	Large Natural Outdoor Scenes	85.2%	87.2%	✓
Fish	81.8%	82.4%	✓	Food Containers	79.6%	80.6%	✓
Insects	85.0%	85.8%	✓	Fruit and Vegetables	86.2%	89.6%	✓
Large Omnivores and Herbivores	84.4%	84.8%	✓	Household Electrical Device	86.4%	87.0%	✓
Large Carnivores	83.8%	84.4%	✓	Vehicles 1	88.6%	90.8%	✓
Non-insect Invertebrates	81.4%	82.0%	✓	Large Man-made Outdoor Things	89.8%	89.6%	
Medium-sized mammals	82.6%	85.2%	✓	Household Furniture	86.%.	84.8%	
People	64.6%	66.8%	✓	Trees	75.8%	74.4%	
Small Mammals	74.8%	79.0%	✓	Flowers	86.0%	85.6%	
Reptiles	73.8%	77.2%	✓	Vehicles 2	91.8%	90.2%	
Arthropod	73.65%	76.26%	✓	Artifact	66.10%	66.51%	✓
Coelenterate	77.60%	78.00%	✓	Geological Form	68.16%	68.24%	✓
Echinoderm	65.20%	68.80%	✓	Miscellaneous	69.08%	69.57%	✓
Shellfish	60.80%	65.60%	✓	Natural Object	74.40%	75.12%	✓
Vertebrate	72.64%	72.79%	✓				

6. Conclusion

In this paper, we introduce Möbius data augmentation, a method that applies $\mathcal{M}$-admissible Möbius transformations to images during training to improve model generalization. Empirically, Möbius performs best when applied to a small subset of data per class, e.g. in low data settings. Möbius transformations are complementary to other sample-level augmentations that preserve labels, such as cutout or standard affine transformations. In fact, across experiments on CIFAR-10, CIFAR-100, and Tiny ImageNet, we find that cutout and Möbius can be combined for superior performance over either alone. In future work, we plan to examine integrating them into the many successful data augmentation policies for fully supervised and semi-supervised learning. Ultimately, this work presents the first foray into successfully employing Möbius transformations—the next level of mathematical abstraction from affine transformations—for data augmentation in neural networks and demonstrating the efficacy of this biologically motivated augmentation on image classification benchmarks.

Data availability statement

The data that support the findings of this study are openly available. The official repository containing experimental code, analysis, and demos are at https://github.com/stanfordmlgroup/mobius.

Broader impact

Those who may benefit from this research largely include computer vision researchers and practitioners wishing to improve the generalization of their models without changing their model architectures or requiring additional large computational resources. This may specifically and especially benefit those who work on animal datasets, including the camera trap community who identify (rare) species in camera trap footage. As for biases, Möbius has shown improvements on animate classes over inanimate ones. Consequently, this may result in imbalanced performance improvements, particularly on inanimate classes where this type of equivariance may not make sense to enforce.

Appendix A. Unconstrained Möbius

Table A1. Juxtaposition of model performance using randomly parameterized Möbius transformations with those using defined ones. Möbius transformations with random parameters suffers in performance, though still better than the crop-and-flip baseline.

Augmentation method	Dataset	Accuracy
Crop-and-flip	C10	96.47% $\pm$ 0.04
Möbius	C10	**96.72% $\pm$ 0.06**
Random Möbius	C10	96.54% $\pm$ 0.06
Crop-and-flip	C100	81.91% $\pm$ 0.20
Möbius	C100	**82.85% $\pm$ 0.31**
Random Möbius	C100	82.30% $\pm$ 0.11
Crop-and-flip	R C10	83.98% $\pm$ 0.16
Möbius	R C10	**86.07% $\pm$ 0.24**
Random Möbius	R C10	85.54% $\pm$ 0.26

Möbius data augmentation operates by constraining the group of Möbius transformations that preserve image labels to the class of $\mathcal{M}$-admissible Möbius transformations, discussed in greater detail in section 3. We run experiments comparing the performance of unconstrained Möbius transformations in $\mathbb{R}^2$ against our method on all dataset settings. Figure E2 displays a visual comparison.

As shown in table A1 below, we observe that unconstrained Möbius still outperforms the baseline of crop-and-flip transformations, even though it performs worse than our proposed method. Recall that the goal of the $\mathcal{M}$-admissible class is to prevent disruptive transformations, for example, that would cause only a single pixel to remain (similar to allowing 'crop' to crop the image down to 1 pixel).

Our speculation is that most of the time, randomly parameterized Möbius transformations are relevant to the image's invariance, which would improve the model's ability to generalize. Under fully unconstrained Möbius, we would expect such transformations, that would unlikely improve regularization and may even hurt generalization, to occur more frequently.
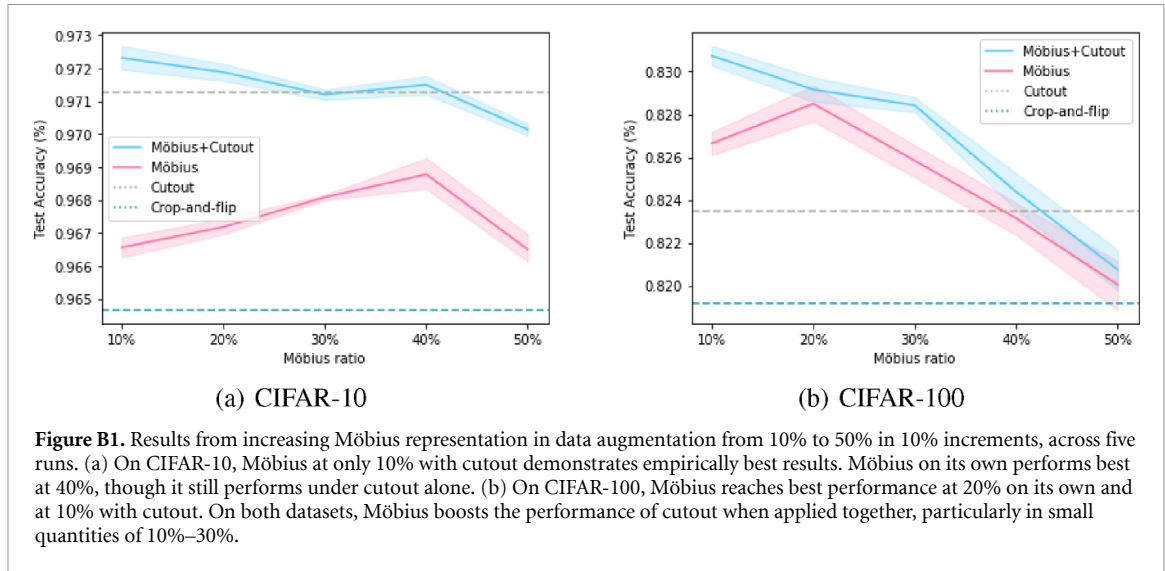
Appendix B. Modulating the inclusion of Möbius

Given the inherent complexity of Möbius transformations, we additionally explore the effects of incorporating an increasing amount of Möbius transformations into the data augmentation process. We evaluate Möbius representations of 10%–50%, at increments of 10% in between, on CIFAR-10 and CIFAR-100. The goal of this experiment is to examine the effects of modulating Möbius representation during the training process. Note that the experiments in section 5.1 only focused on a stationary amount (20%) of Möbius.

We compare these increments of Möbius both with and without cutout. We then juxtapose these results with the baseline of cutout alone and that of standard crop-and-flip. We again report average performance and standard deviations across five runs on all experimental conditions. The results presented in figure B1 emphasize the following findings.

- Too much Möbius data augmentation can result in disruptive training and poorer generalization.
- Möbius augmentation nevertheless outperforms both cutout and standard crop-and-flip baselines, across several values of representation particularly at 10% and 20% representation.
- Möbius augmentation alone experiences a local optimum at 40% inclusion on CIFAR-10 and 20% on CIFAR-100.
- Möbius with cutout performs best with a very modest amount (10%) of Möbius. This is expected, as cutout provides additional regularization.

Though not shown in the graph, we also experiment with an even lower representation (5%) of Möbius in the Möbius with cutout condition, in order to observe local optima and a bottoming out effect. We find

(a) CIFAR-10 (b) CIFAR-100

Figure B1. Results from increasing Möbius representation in data augmentation from 10% to 50% in 10% increments, across five runs. (a) On CIFAR-10, Möbius at only 10% with cutout demonstrates empirically best results. Möbius on its own performs best at 40%, though it still performs under cutout alone. (b) On CIFAR-100, Möbius reaches best performance at 20% on its own and at 10% with cutout. On both datasets, Möbius boosts the performance of cutout when applied together, particularly in small quantities of 10%–30%.

that 10% still shows superior performance to 5% representation on both datasets. Specifically, Möbius at 5% with cutout performs $97.18\% \pm 0.14$ on CIFAR-10 and $82.97\% \pm 0.17$ on CIFAR-100.

Appendix C. Defined Möbius parameters

Given the inherent variability of Möbius transformations, we additionally explore the effects of predefining a set of fixed Möbius transformations as a way of decreasing variation and constraining the transformations to be human interpretable. Specifically, we define eight highly variable parameterizations that we visually verify to maintain their respective class labels. Based on their appearance, we describe them each as follows: (1) clockwise twist, (2) clockwise half-twist, (3) spread, (4) spread twist, (5) counter clockwise twist, (6) counter clockwise half-twist, (7) inverse, and (8) inverse spread. Concretely, these parameters are presented below, where $\Re(p)$ and $\Im(p)$ denote the respective real and imaginary components of a point p, and height x and width y are dimensions of the original image.

Across all data settings, we found that the defined set of parameterizations performed better on average in experiments than our proposed class of $\mathcal{M}$-admissible Möbius transformations, though the difference was not significant. This suggests that restraining variability to human interpretable transformations could improve model regularization and lead to improved generalization, though the difference is not significant. This is not extremely surprising, because Möbius transformations can take on highly variable forms, for which some we may not expect or desire invariance. Nevertheless, this fixed method trades off the method's generalizability and ease of implementation.

Here is the precise parameterization:

(a) Clockwise twist:

$$\Re(z) = \{1, 0.5x, 0.6x\},$$
$$\Im(z) = \{0.5y, 0.8y, 0.5y\},$$
$$\Re(w) = \{.5x, 0.5x + 0.3\sin(0.4\pi)y, 0.5x + 0.1\cos(0.1\pi)y\},$$
$$\Im(w) = \{y - 1, 0.5y + 0.3\cos(0.4\pi)y, 0.5y - 0.1\sin(0.1\pi)x\}.$$

(b) Clockwise half-twist:

$$\Re(z) = \{1, 0.5x, 0.6x\},$$
$$\Im(z) = \{0.5y, 0.8y, 0.5y\},$$
$$\Re(w) = \{.5x, 0.5x + 0.4y, 0.5x\},$$
$$\Im(w) = \{y - 1, 0.5y, 0.5y - 0.1x\}.$$

(c) Spread:

$$\Re(z) = \{.3x, 0.5x, 0.7x\},$$
$$\Im(z) = \{0.5y, 0.7y, 0.5y\},$$

$$\Re(w) = \{0.2x, 0.5x, 0.8x\},$$
$$\Im(w) = \{0.5y, 0.8y, 0.5y\}.$$

(d) Spread twist:

$$\Re(z) = \{.3x, 0.6x, 0.7x\},$$
$$\Im(z) = \{0.3y, 0.8y, 0.3y\},$$
$$\Re(w) = \{0.2x, 0.6x, 0.8x\},$$
$$\Im(w) = \{0.3y, 0.9y, 0.2y\}.$$

(e) Counter clockwise twist:

$$\Re(z) = \{1, 0.5x, 0.6x\},$$
$$\Im(z) = \{0.5y, 0.8y, 0.5y\},$$
$$\Re(w) = \{0.5x, 0.5x + 0.4y, 0.5x\},$$
$$\Im(w) = \{y - 1, 0.5y, 0.5y - 0.1x\}.$$

(f) Counter clockwise half-twist:

$$\Re(z) = \{1, 0.5x, 0.6x\},$$
$$\Im(z) = \{0.5y, 0.8y, 0.5y\},$$
$$\Re(w) = \{0.5x, 0.5x + 0.3\sin(.4\pi)y, 0.5x + 0.1\cos(.1\pi)x\},$$
$$\Im(w) = \{y - 1, 0.5y + 0.3\cos(.4\pi)y, 0.5y - 0.1\sin(.1\pi)x\}.$$

(g) Inverse:

$$\Re(z) = \{1, 0.5x, x - 1\},$$
$$\Im(z) = \{0.5y, 0.9y, 0.5y\},$$
$$\Re(w) = \{x - 1, 0.5x, 1\},$$
$$\Im(w) = \{0.5y, 0.1y, 0.5y\}.$$

(h) Inverse spread:

$$\Re(z) = \{0.1x, 0.5x, 0.9x\},$$
$$\Im(z) = \{0.5y, 0.8y, 0.5y\},$$
$$\Re(w) = \{x - 1, 0.5x, 1\},$$
$$\Im(w) = \{0.5y, 0.1y, 0.5y\}.$$

Appendix D. Möbius points mapping with and without interpolation

We include visual representations of mapping Möbius transformations from three points $\{w_1, w_2, w_3\}$ on the original image to three separate target points $\{z_1, z_2, z_3\}$ on the plane. In each example, the red, green, and blue points demonstrate various mappings between the two sets of three corresponding points. We also illustrate the effects of interpolation in filling in the gaps created by the Möbius transformations.

Note that there is increased scatter of points when the mapped points are closer to the edge of the image and pixels are lost in the transformation, similar to scaling and cropping.
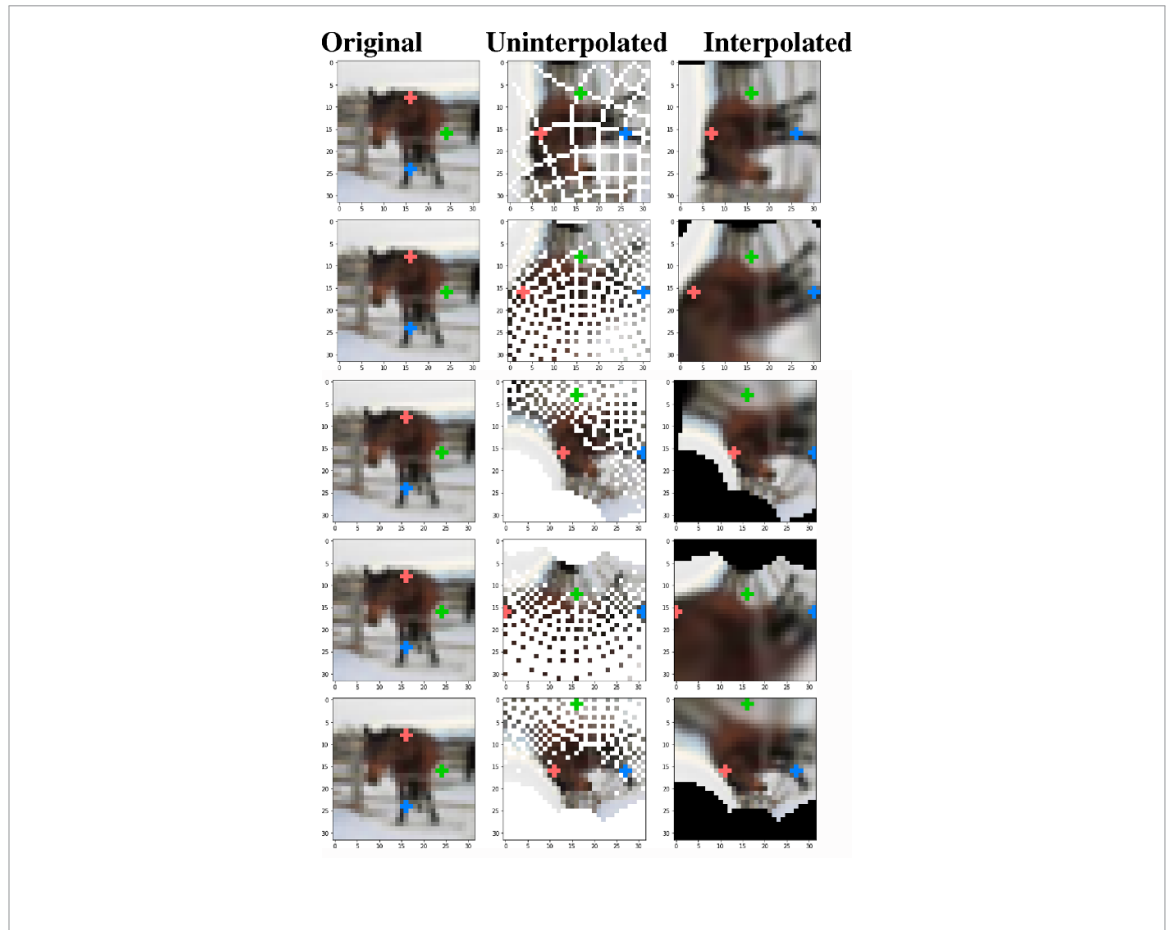
Appendix E. Training logs for different settings

As data augmentation is largely used for regularization and invariance, we also examine training accuracy against test accuracy to gain some indication of overfitting on different settings: Baseline crop-and-flip in figure E2, Cutout in figure E3, Möbius in figure E4, and Möbius with Cutout in figure E5. With Möbius, we find some empirical evidence that we are able to reduce overfitting, while slightly improving the test accuracy. All figures below were trained on Tiny ImageNet.
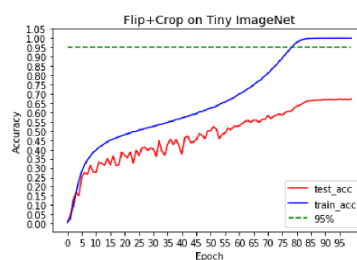
Figure E2. Accuracy curves on Tiny Imagenet under the baseline crop-and-flip setting.
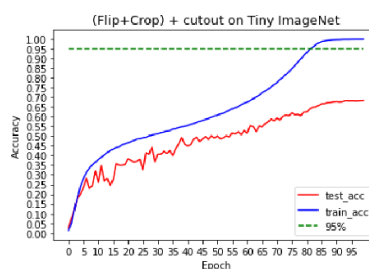

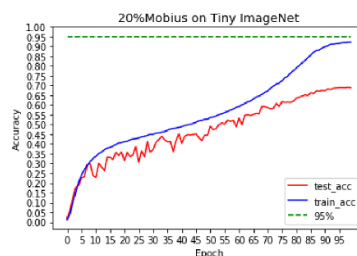**Figure E3.** Accuracy curves on Tiny Imagenet under Cutout.

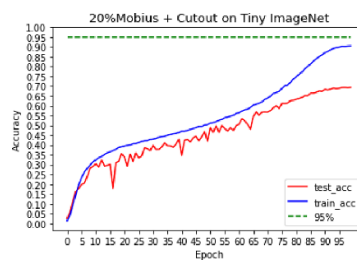**Figure E4.** Accuracy curves on Tiny Imagenet under Möbius.

Figure E5. Accuracy curves on Tiny Imagenet under Möbius + Cutout.

ORCID iD

Sharon Zhou ☉ https://orcid.org/0000-0003-4991-9336

References

[1] Perez L and Wang J 2017 The effectiveness of data augmentation in image classification using deep learning (arXiv:1712.04621)
[2] DeVries T and Taylor G W 2017 Improved regularization of convolutional neural networks with cutout (arXiv:1708.04552)
[3] Zhong Z, Zheng L, Kang G, Li S and Yang Y 2017 Random erasing data augmentation (arXiv:1708.04896)
[4] Guo J and Gould S 2015 Deep CNN ensemble with data augmentation for object detection (arXiv:1506.07224)
[5] Wang P, Chen P, Yuan Y, Liu D, Huang Z, Hou X and Cottrell G 2018 Understanding convolution for semantic segmentation *2018 IEEE Winter Conf. on Applications of Computer Vision (WACV)* (IEEE) pp 1451–60
[6] Hernández-García A and König P 2018 Data augmentation instead of explicit regularization (arXiv:1806.03852)
[7] Zhang X, Wang Z, Liu D and Ling Q 2019 DADA: deep adversarial data augmentation for extremely low data regime classification *ICASSP 2019-2019 IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)* (IEEE) pp 2807–11
[8] Zhang H, Cisse M, Dauphin Y N and Lopez-Paz D 2017 mixup: beyond empirical risk minimization (arXiv:1710.09412)
[9] Cubuk E D, Zoph B, Mane D, Vasudevan V and Le Q V 2019 Autoaugment: learning augmentation strategies from data *Proc. IEEE Conf. on Computer Vision and Pattern Recognition* pp 113–23
[10] Cubuk E D, Zoph B, Shlens J and Le Q V 2019 Randaugment: practical data augmentation with no separate search (arXiv:1909.13719)
[11] Xie Q, Dai Z, Hovy E, Luong M-T and Le Q V 2019 Unsupervised data augmentation (arXiv:1904.12848)
[12] Berthelot D, Carlini N, Goodfellow I, Papernot N, Oliver A and Raffel C 2019 Mixmatch: a holistic approach to semi-supervised learning (arXiv:1905.02249)
[13] Thompson D W *et al* 1942 On growth and form *On Growth and Form* (Cambridge: Cambridge University Press)
[14] Petukhov S V 1989 Non-Euclidean geometries and algorithms of living bodies *Comput. Math. Appl.* **17** 505–34
[15] Lundh T, Udagawa J, Hänel S-E and Otani H 2011 Cross- and triple-ratios of human body parts during development *Anatomical Rec.* **294**
[16] Liouville J 1850 Extension au cas des trois dimensions de la question du tracé géographique *Note VI* pp 609–17 (in French)
[17] Ahlfors L V 1989 *Möbius Transformations in Several Dimensions* (Minneapolis, MN: University of Minnesota Press)
[18] Needham T 1998 *Visual Complex Analysis* (Oxford: Oxford University Press)
[19] Criminisi A, Reid I and Zisserman A 2000 Single view metrology *Int. J. Comput. Vis.* **40** 123–48
[20] Lie S 1893 Theorie der transformationsgruppen abschn *Theorie der Transformationsgruppen* (Leipzig: Teubner)
[21] Dubejko T and Stephenson K 1995 Circle packing: experiments in discrete analytic function theory *Exp. Math.* **4** 307–48
[22] Cheng G, Zhou P and Han J 2016 Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images *IEEE Trans. Geosci. Remote Sens.* **54** 7405–15
[23] Cheng G, Han J, Zhou P and Xu D 2018 Learning rotation-invariant and Fisher discriminative convolutional neural networks for object detection *IEEE Trans. Image Process.* **28** 265–78
[24] Tokozume Y, Ushiku Y and Harada T 2018 Between-class learning for image classification *Proc. Conf. on Computer Vision and Pattern Recognition* pp 5486–94
[25] Inoue H 2018 Data augmentation by pairing samples for images classification (arXiv:1801.02929)
[26] Takahashi R, Matsubara T and Uehara K 2020 Data augmentation using random image cropping and patching for deep CNNs *IEEE Trans. Circuits Syst. Video Technol.* **30** 2917–31
[27] Summers C and Dinneen M J 2019 Improved mixed-example data augmentation *2019 IEEE Conf. on Applications of Computer Vision (WACV)* (IEEE) pp 1262–70
[28] Tang Z, Peng Xi, Li T, Zhu Y and Metaxas D N 2019 Adatransform: adaptive data transformation *Proc. IEEE Int. Conf. on Computer Vision* pp 2998–3006
[29] Ho D, Liang E, Stoica I, Abbeel P and Chen Xi 2019 Population based augmentation: efficient learning of augmentation policy schedules (arXiv:1905.05393)
[30] Sohn K, Berthelot D, Li C-L, Zhang Z, Carlini N, Cubuk E D, Kurakin A, Zhang H and Raffel C 2020 Fixmatch: simplifying semi-supervised learning with consistency and confidence (arXiv:2001.07685)
[31] Ratner A J, Ehrenberg H, Hussain Z, Dunnmon J and Ré C 2017 Learning to compose domain-specific transformations for data augmentation *Advances in Neural Information Processing Systems* pp 3236–46
[32] Özdemir N, İskender B B and Özgür N Ylmaz 2011 Complex valued neural network with möbius activation function *Commun. Nonlinear Sci. Numer. Simul.* **16** 4698–703
[33] Zammit-Mangion A, Ng T L J, Vu Q and Filippone M 2019 Deep compositional spatial models (arXiv:1906.02840)
[34] Ganea O, Bécigneul G and Hofmann T 2018 Hyperbolic neural networks *Advances in Neural Information Processing Systems* pp 5345–55
[35] Islam M A, Anderson D T, Pinar A, Havens T C, Scott G and Keller J M 2020 Enabling explainable fusion in deep learning with fuzzy integral neural networks *IEEE Trans. Fuzzy Syst.* **28** 1291–300
[36] Mandic D P 2000 The use of Mobius transformations in neural networks and signal processing *Neural Networks for Signal Processing X. Proc. 2000 IEEE Signal Processing Society Workshop (Cat. No. 00TH8501)* vol 1 (IEEE) pp 185–94
[37] Krizhevsky A *et al* 2009 Learning multiple layers of features from tiny images *Proc. British Machine Vision Conference (BMVC)* pp 87.1–87.12 (http://www.bmva.org/bmvc/2016/papers/paper087/index.html)
[38] Wu J, Zhang Q and Xu G 2017 Tiny imagenet challenge *Technical Report* Stanford University
[39] Zagoruyko S and Komodakis N 2016 Wide residual networks (arXiv:1605.07146)