Machine Learning Models for Road Surface and Friction Estimation using Front-Camera Images

N.B. When citing this work, cite the original published paper.

(article starts on next page)

Machine Learning Models for Road Surface and Friction Estimation using Front-Camera Images

Sohini Roychowdhury[1,3], Minming Zhao[1], Andreas Wallin[1], Niklas Ohlsson[2], Mats Jonasson[2]

[1]Volvo Cars Technology USA, Mountain View, CA-94043, [2]Vehicle Motion and Control, Volvo Cars, 405 31 Mölndal, Sweden
[3]Affiliated to University of Washington Bothell, WA-98011

Abstract—**Automotive active safety systems can significantly benefit from real-time road friction estimates (RFE) by adapting driving styles, specific to the road conditions. This work presents a 2-stage approach for indirect RFE estimation using front-view camera images captured from vehicles. In stage-1, convolutional neural network model architectures are implemented to learn region-specific features for road surface condition (RSC) classification. Texture-based features from the drivable surface, sky and surroundings are found to be separate regions of interest for dry, wet/water, slush and snow/ice RSC classification. In stage-2, a rule-based model that relies on domain-specific guidelines is implemented to segment the ego-lane drivable surface into [5x3] patches, followed by patch classification and quantization to separate images with high, medium and low RFE. The proposed method achieves average accuracy of 97% for RSC classification in stage-1 and 89% for RFE classification in stage-2, respectively. The 2-stage models are trained using publicly available data sets to enable benchmarking for future methodologies in the autonomous driving domain.**

Index Terms—**Deep learning, convolutional neural network, classification, drivable surface, features.**

I. INTRODUCTION

Active safety systems and self-driving cars can significantly benefit from real-time prediction of drivable surface conditions, since driving styles can be adapted to corresponding road conditions. In 2017, the US department of transportation Federal Highway Administration reported that about 22% of vehicle crashes per year, with about 16% life casualties, are weather-related [1]. Also, majority of accidents were reported to happen during wet road conditions, where 73% of accidents occurred on wet pavements, and 17% on snow or sleet. Further, the National Safety Council in USA estimated a 6% rise in motor vehicle crashes in 2016 over the year 2015 leading to motor-vehicle deaths, injuries, and property damage in 2016 that amounted to $432 billion [2]. Such events agglomerated with the recent surge in vehicles on roads, motivate active monitoring of drivable surface/road conditions that can lead to prediction and prevention of motor vehicle crashes on slippery roads. One of the most important measures of drivable surface/road condition is the road-tire friction coefficient that indicates vehicle grip. Active safety functions, such as collision avoidance by braking, can significantly benefit from low friction predictions by avoiding any imminent threat early.

Road friction estimation methods are dived into direct and indirect categories [3]. Direct methods, such as [4], use slip measurements from the tires for friction estimation. However, such methods have low data availability and large levels of friction utilization i.e., the vehicle must approach the current grip limit to estimate friction with sufficient confidence. To increase the data availability, studies such as [5] have estimated friction independently from high friction utilization by using specific optical sensor-based indirect methods. Some prior works have used front-view camera sensors to infer road surface conditions, such as [6], that transformed the road surface image into spatial frequency spectra and analyzed the frequency distribution corresponding to road surface types such as: dry/wet asphalt, fresh/trampled snow, and black ice. Another work in [7] fed color-transformed road surface images from static cameras to multi-layered perceptrons for surface classification. Further, the works in [8] [9] used polarization filter and wavelet transformations to classify dry, snow and ice surfaces. All such existing methods lack consistency in predicting road surface condition (RSC) and inferring road friction estimates (RFEs). In this work, we investigate a variety of supervised methods in a 2-stage approach to analyze the importance of image-based features towards RSC and RFE classification tasks, as shown in Fig. 1.

This paper makes two key contributions. First, we analyze deep learning and feature-based models for drivable RSC classification. We observe that the texture of the drivable surface in conjunction with the sky and surroundings are significant features for RSC detection with 94-97% average classification accuracy per class. Second, we propose a novel approach for spatially segmenting the drivable surface to identify the extent of vehicle ego-lane patchiness that can indicate high, medium or low RFE. The proposed approach is capable of separating icy roads (low RFE) from well-ploughed snow roads (high RFE) with 89% accuracy in real time using only front camera images.

The rest of the paper is organized as follows. Related prior works are discussed in Section II and materials and methods are presented in Section III. In Section IV the experimental setups and results are presented. Finally, conclusions and discussion are given in Section V.

II. RELATED WORK

Most prior works on RFE detection implement experiment-based or model-based approaches to find a correlation between the vehicle dynamics sensor data and road-tire friction related parameters [10]. The recent work in [11] used 6 vehicle
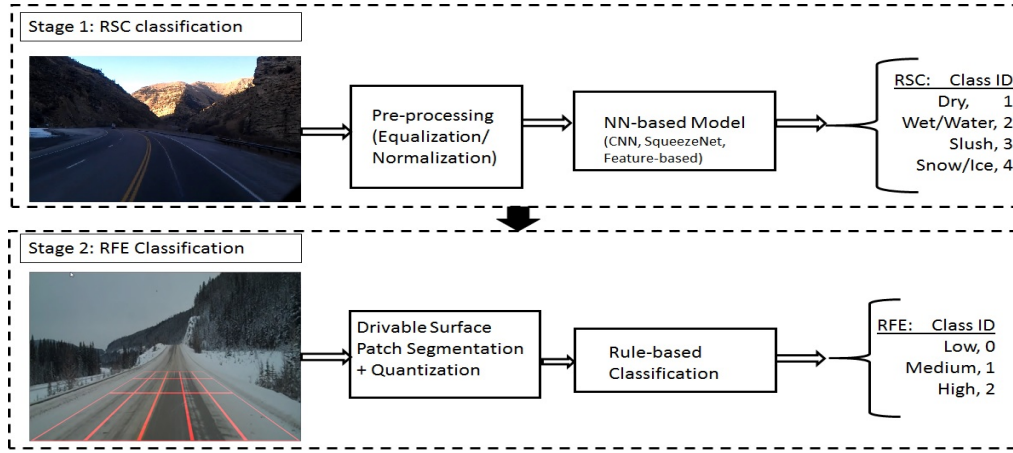
Fig. 1. The proposed 2-stage method. Stage-1 includes pre-processing each image followed by the implementation of deep learning and feature-based models for RSC classification. Stage-2 includes spatial segmentation of drivable surface into patches followed by patch quantization and classification for low, medium and high RFE. The 2-stage method enables hierarchical classification for RFE. Images with 'dry' RSC can be assumed to have high RFE, hence, they are not subjected to Stage-2. Stage-2 is designed specifically for combination surfaces with dry asphalt road and snow patches in the ego-lane drivable surface.

sensors as inputs to a neural network (NN) model for prediction of slippery/non-slippery RSC, which controlled the driver maneuvering actions for left or right lane departures. Some other vision-based works have implemented classification of the road texture immediately ahead of vehicles using rule-based models or texture-based features in [12] and [13], respectively. Till date, several patents have addressed the problem of drivable RSC prediction. The apparatus in the patent [14] included a stereo camera and memory to store road surface conditions and predict dry, wet, snow and icy states. This work depended on color space transformations, wavelet transforms and classifiers such as support vector machines (SVM) and AdaBoost for RSC classification. The patent in [15] included apparatus for focusing circular lights ahead of the vehicle and measuring 4 intensity-based features that include normalized mean, variance and maximum of pixel intensities within a certain ROI of the focused light. The patent in [16] implemented several probabilistic binary classifiers for classification of dry, wet, snow and ice RSC followed by a fusion unit, to generate control parameters for braking, steering and acceleration sensors.

Among state-of-the-art methods that implement vehicle front-facing camera image processing for RSC classification, the work in [9] color transformed image planes and used Naive Bayesian, Random Forest, NN and SVM classifiers, that resulted in a recognition rate of 83% for dry and wet roads. Another work in [17], investigated hybrid road conditions by partitioning of the road surface images to [9x5], [8x6] etc. sub-regions/patches followed by classification of each patch as dry, wet, snow, ice or water. This method reported up to 90% accuracy on image patch classification. All these existing works were analyzed on limited local data sets, which raise bench-marking issues for further research. This work is the first contribution towards methodological bench-marking on public domain data sets and inferring image-based features indicative of RFE.

III. MATERIALS AND METHODS

The proposed method involves two stages that classify RSC using image-based features, followed by segmentation of the drivable surface for RFE classification. The 2-stage process is designed for hierarchical decision making for active safety functions. While images with 'dry' RSC imply high RFE, images with slush or wet, unploughed snow and ice roads may display medium to low RFE. Thus, images classified with RSC=[2,3,4] are subjected to stage-2 classification. The RSC definitions are provided in Fig. 1.

In stage-1, three variants of NN-based models are implemented that include two deep learning models and one feature-based classification model. So far, deep learning models, optimally parameterized for few to several deep layers, have proven effective for extracting abstract features from groups of image pixels [18] [19]. While deep learning models provide generalizability with minimal pre-processing, they require large numbers of training samples to avoid over-fitting. Further, we analyze the hidden layer kernels of the trained deep learning models, also known as the activated feature maps (AFMs), and learn the region-based features that can then be extracted directly from the images for feature-based classification. One deep learning model under analysis is the convolutional neural network (CNN) architecture defined in [19]. Training this model on the RSC data set from scratch results in AFMs shown in Fig. 2, where we observe that texture of road surface, sky and remaining regions in the image that constitute 'surroundings' are significant for overall drivable condition classification. The second deep learning model includes the SqueezeNet model [20], where the parameters from Alexnet [21] are pruned to significantly reduce the number of training parameters. The third feature-based model used in stage-1 segments each image into regions of interest (ROIs) followed by regional feature extraction and classification using single hidden layer NN classifier as shown in Fig. 3.
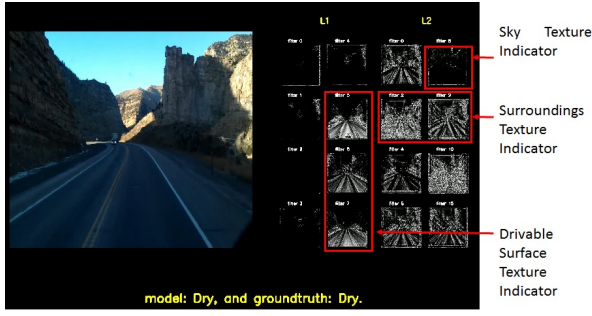
Fig. 2. Feature learning from AFMs of a CNN model. The AFMs from the first 2 layers are shown. Drivable surface, sky and surroundings texture are learned as significant features from different kernels for this image.
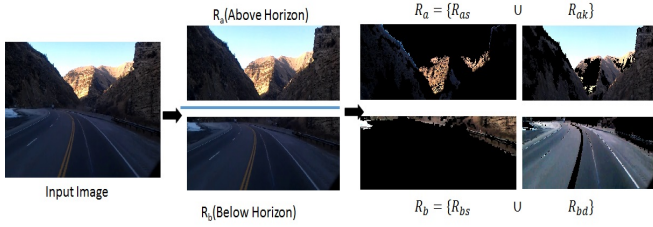

Fig. 3. Image processing followed by NN-classifier for RSC classification. The automated horizon detection enables separation of the sky above and drivable surface below it.

For all the aforementioned NN-based models, model training is performed by stratified sampling using 70% of the data samples and tested on the remaining 30% samples. Stratified sampling ensures similar sample class frequency distributions in the training and test data sets [18]. For the stage-2, the drivable surface is spatially segmented into a [5x3] matrix of patches followed by color and texture-based quantification for each patch. Finally the drivable surface matrix is subjected to classification for high, medium and low RFE indicative of non-slippery, moderately slippery and highly slippery surfaces, respectively. The image data sets and the data models under analysis are explained in the following subsections.

A. Data, Notation and Metrics

For this work, public domain data was collected from 48 youtube videos. Each video was sampled to isolate frames that are greater than 1 second apart, to ensure variability across images. Each image, treated as an independent sample, has size in the range of [640x360] to [1280x720] pixels. Examples of the 4 classes of images, thus extracted from the videos are shown in Fig. 4. Each image is manually annotated to assign one RSC and RFE groundtruth (GT) label per image by two expert annotators. The data set is available for benchmarking purposes[1].

For model training and testing in stage-1, 40-100 frames are extracted per video, thereby resulting in 3750 training images and 1550 test images, respectively. The numbers of images

belonging to class [1,2,3,4] in the training and test data sets are [1047,893,574,1196] and [430,376,238,546], respectively. The pre-processing module for the deep learning models extracts 6 color channels per image that represent the red (R), green (G), blue (B), hue (H), saturation (S) and intensity (V) planes, respectively. Also, for the deep learning models, each image color channel is resized to [256x256] to control the model complexity, thereby resulting in input size per image of [256x256x6]. Additionally, for the deep learning models, training data is augmented to six-times the number of samples by applying histogram-based image equalization ('clahe' function) and contrast enhancement ('imadjust' function) followed by horizontal flip applied to each image plane.

The stage-2 of the proposed method is designed to analyze images with varying combinations of dry and snow patches to accurately estimate high, medium and low RFE conditions. Since classification of RFE is the tire-to-road friction assessment that can be expected when driving with similar tires, the stage-2 model is trained and tested with data that appear to satisfy this assumption. A drivable surface patch classifier is trained with [5x3=15] patches per image on perspective transformed images of vehicle ego-lane from dry asphalt and fresh snow images gathered from public and local data sets, respectively. Next, the classified and quantized patches are analyzed to predict RFE as class 0: low RFE, class 1: medium RFE, class 2: high RFE. The stage-2 patch classifier is tested against a homogeneous set of images that contain 355 and 128 perspective transformed images of dry asphalt and fresh snow, respectively. Next, a test set of images containing 114 images with mixed snow and dry RSC are analyzed for RFE classification task, with [8, 6 and 100] images with high, medium and low RFE, respectively.

The stage-1 and stage-2 models are analyzed using input vector X_i for each image 'i', such that manually annotated GT label assigned per image is target label vector t_i and RFE_i for stage-1 and stage-2, respectively. The proposed methods are analyzed on the respective test data sets in terms of classification accuracy (Acc), precision (PR) and recall (RE). Here, Acc analyzes the percentage of correct classifications, PR analyzes the false positive rate, while RE analyzes the false negative rate, respectively. The definition of notations used by the models are presented in Table I.

B. Step 1: RSC Classification

1) CNN Model: This category of deep learning models are well known for hybrid feature abstraction from images [22]. While the convolution (C) layers extract neighborhood pixel contributions, the sub-sampling (S) /pooling layers followed by a non-linear activation function enable multi-resolution feature learning. A sample architecture of a CNN model with 3 convolution/sub-sampling layers (CS^3) followed by a dense (D) flattening layer is shown in Fig. 5.

Optimal parameterization for the CNN model involves selection of sufficient C-S layers, and optimal filter bank sizes to ensure low trainable parameters, thereby preventing overfitting. Often, deep CNN models with many hidden layers

| Dry Asphalt: Class 1 | Wet/Water: Class 2 | Slush: Class 3 | Snow/Ice: Class 4 |

Fig. 4. Examples of images under analysis for Stage-1 of the proposed method.

TABLE I
NOTATION TABLE FOR STAGE-1 AND STAGE-2 MODELS PER IMAGE/PATCH.

| Notation	Definition
X_i	Input features for image i
x_i^j	j th feature for image i, $x_i^j \in X_i$
$y_{i'}$	i' th output from previous NN-layer
$w_{ji'}$	NN-layer weights between input j and output i'
t_i	GT class label vector for stage-1 models and image i
p_{sl}^i	Predicted probability that image/patch i belongs to class l.
s_i	Predicted class label for image/patch i.
$\bar{p}^i_{s1}\|col$	Average probability of image patch i per column col in the patch quantized matrix for stage-2.
R_a	Segmented ROI above detected horizon for stage-1 feature-based model.
R_{as}	Segmented ROI above horizon depicting the surroundings.
R_{ak}	Segmented ROI above horizon depicting the sky.$R_a = R_{as} \cup R_{ak}$
R_b	Segmented ROI below detected horizon for stage-1 feature-based model.
R_{bs}	Segmented ROI below horizon depicting the surroundings.
R_{bd}	Segmented ROI below horizon depicting drivable surface.$R_b = R_{bs} \cup R_{bd}$
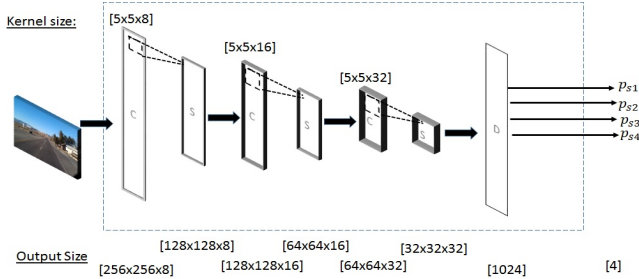
Fig. 5. Example of CNN architecture with 3 consecutive C-S layers (CS^3) followed by a dense D layer for 4 output classes.

are shown to overfit to training data, when huge disparities between validation and test data accuracies are observed [18]. In this work, we investigate several CNN architectures with consecutive C-S layers, such as CS^1, CS^2, CS^3, where, single C-S configuration, 2 consecutive C-S layers and 3 consecutive C-S layers, respectively, are applied between the input and the D-layer. Also, the optimal filter bank sizes $[w, h, v]$, representing the width, height and number of filter banks per 'C' layer are varied in the range [3:17] with stride of 2. The initial filter bank dimension that results in maximum test accuracy following this grid search is [5x5x8].

Each CNN model is designed to process images of size $[256 \times 256 \times 6]$, where 6 refer to the color channel features. The max-pooling (S) layer further reduces the resolution of the convolved image exactly by half, followed by a Rectified Linear Unit (ReLU) activation function and batch normalization operation along with dropout probability of 0.8. The addition

of batch normalization [19] makes the network train faster while ensuring low training losses. The number of filters in each convolutional layer double in each consecutive stage. Thus, for n_T training images, and n_c number of output classes, the cross-entropy loss function used for training NN-based models by back-propagation is given in (1),

$$L = -\sum_{i=1}^{n_T} \sum_{l=1}^{n_c} \mathbb{1}, [t_i = l] log(p_{sl}^i) \tag{1}$$

where, $\mathbb{1}[t_i = l]$ is the indicator function that GT vector t_i is equal to class label l, while p_{sl}^i represents the model predicted probability that image i belongs to class label l. The ReLU activation function applied on weighted sum of outputs from the previous layer i' are given in (2).

$$y_{i'} = \sum_{k=1}^{n} x_k w_{ki'}, \; p_{si'}^i = max(0, y_{i'}). \tag{2}$$

Here, n signifies the number of neurons present in the layer just before the activation layer. The final predicted class label is the class with maximum predicted probability in (3).

$$s_i = \arg_l \max(p_{sl}^i) \quad \forall l \text{ classes.} \tag{3}$$

The predicted class (s_i) is then compared with the actual target class label, and with the aid of the loss function L, model weights are updated such that the loss is minimized for subsequent epochs. Once trained, convolution of the learned kernel weights demonstrate the structure and textural features/ROIs that aid the classification process, as shown in Fig. 2. By analyzing AFMs from different drivable surface conditions, we observe that texture of drivable surface, surroundings and the sky are collectively useful for RSC classification.

2) SqueezeNet Model: This deep learning model proposed in [20] involves replacing [3x3] convolution filters in the AlexNet model [21] by [1x1] filters, thereby decreasing the number of input channels by using 'squeeze' layers. The process of reducing the number of model parameters is through a *Fire* Module. Here, a *Fire* Module contains the following: a [1x1] convolution module that reduces the number of channels, and an expand Module, where a [1x1] convolution and a [3x3] convolution, both are applied to the output of the previous layer, and their results are concatenated. The use of several [1x1] convolutions is found to be advantageous, since they need less parameters than [3x3] filters. Additionally, the activation layer uses ReLUs [20], only convolutions and

dropout (0.5, before the last convolution). This model uses late max-pooling to improve accuracy while not needing additional parameters. Finally, residual connections are applied between layers of same dimensionality followed by parameter pruning from '*Deep Compression*' to reduce the parameters further.

For our data set, each input image channel [256x256x6] is subjected to convolution, fire and subsampling/pooling layers (max-pooling with filter size [3x3] and stride 2) based on the network structure in [20]. The input channel size for each layer are as follows: Input:[256x256x6], C^1:[127x127x96], S^1:[127x127x96], $Fire^2$:[63x63x128], $Fire^3$:[63x63x128], $Fire^4$:[63x63x256], S^2:[31x31x256], $Fire^5$:[31x31x256], $Fire^6$:[31x31x384], $Fire^7$:[31x31x384], $Fire^8$:[63x63x512], S^3:[15x15x512], $Fire^9$:[15x15x512], D-layer (flatten):[15x15x4], Output:[1x1x4].

3) Feature-based Model: In this model, the pre-processing steps involve a horizon detector to separate the sky and drivable surface components. For this purpose, GIST feature-based horizon detector described in [23][24] is invoked. Here, the relative horizontal location of the horizon is evaluated as a regression problem using GIST features from the training set of images on the LabelMe data set [25]. Next, the sub-image under the horizon (R_b) is segmented into the drivable surface (R_{bd}) and surroundings (R_{bs}). The sub-image above the horizon (R_a) is similarly segmented into sky (R_{ak}) and surroundings (R_{as}), as shown in Fig. 3. Further, the RGB image plane is transformed to HSV plane, and HOG descriptors are computed for each of the 4 segmented ROIs (R_{bd}, R_{bs}, R_{ak}, R_{as}), to extract color and texture-based features, as shown in Fig. 6. Next, the following 15 color and texture based regional features are extracted per segmented ROI: mean and standard deviation of pixel intensities segmented ROI RGB, HSV plane, and HOG feature planes, and the fraction of pixels belonging to each segmented ROI. The HOG features are extracted using an empirically selected block size of [16x16]. These 15 features corresponding to 4 segmented ROIs result in a 60 feature-vector corresponding to X_i for each image i, that are then subjected to classification using a single hidden layer NN-model. The number of hidden neurons are varied between [10:500], to identify the optimal number of hidden neurons that maximize test accuracy.
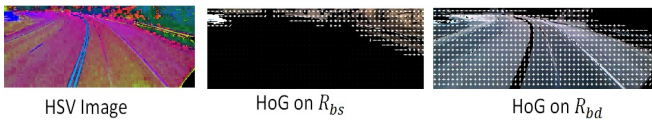

Fig. 6. Examples of color and texture-based features extracted for each segmented ROI.

C. Step 2: RFE Classification

Once RSC is detected, the next step involves estimation of patchiness in the vehicle's ego-lane, to separate low RFE conditions from iced roads from high to medium RFE conditions on ploughed snow-roads. For this purpose, the ego-lane ahead of the vehicle can be segmented into 5 spatial sub-regions/patches across the width and 3 patches across the height, as shown in Fig. 7. The patch sizes are motivated by domain knowledge to accommodate the two wheel-tracks and neighboring lanes for an automobile. The steps for patch segmentation, quantization and RFE classification are explained below.
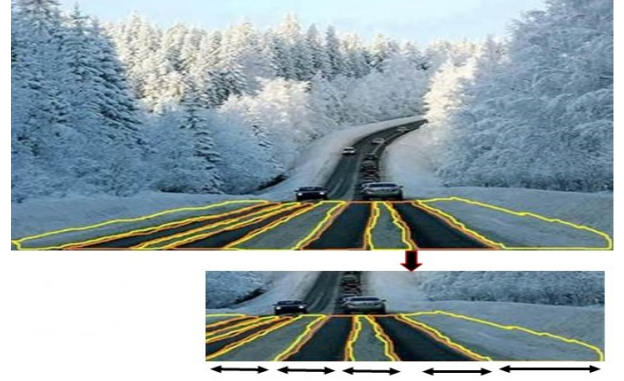

Fig. 7. Rationale behind [5x3] patch segmentation of the ego-lane drivable surface. The 5 patches across width accommodate vehicle wheels and their neighborhood. The 3 patches across height accommodate 30-50 meters of drivable surface ahead of the vehicle. Note that the opposite lane and ego-lane share one column of patch segments.

1) Patch Segmentation: To ensure similarly sized patches, the ego-lane drivable surface below horizon is subjected to perspective transformation to obtain the birds-eye view of the drivable surface. The resulting [5x3] image patches in the driving direction are shown in Fig. 8.
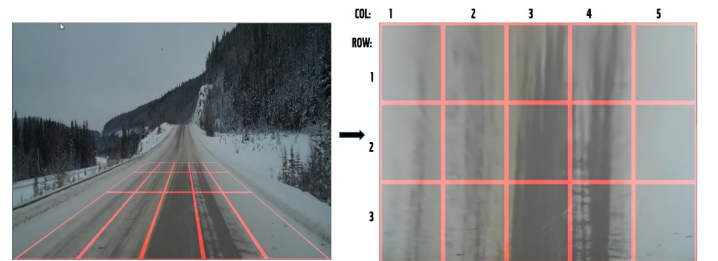

Fig. 8. Segmentation of the [5x3] patches for the drivable surface. (Left:) Perspective transform limits ROI up to 50 meters ahead of vehicle in driving direction. (Right:) Segmented uniformly sized patches.

2) Patch Quantization: Each image patch is classified for its composition as 'dry road' (class 0) or 'snow' (class 1) using a logistic regression model, resulting in a probabilistic patch matrix shown in Fig. 9. This patch classifier is trained on 400 dry asphalt and 400 fresh snow image patches, respectively. 26 color and textural features are extracted per patch for this classification purpose. These features include mean and standard deviation of pixel intensities in [R,G,B,H,S,V] planes with and without contrast enhancement ('imadjust' function), and mean and standard deviation of HOG features extracted using [16x16] block size. This step outputs a probability score (p_{s1}^i) corresponding to per-patch composition. A low
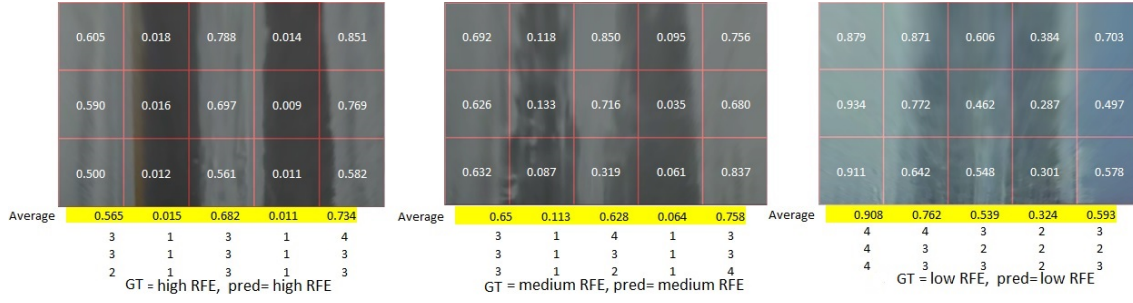
Fig. 9. Example of road patch quantization and RFE classification. (Top:) Probability score 'p_{s1}^i' for patch composition. (Bottom:) Quantized Patches and their GT. (Left:) High RFE for well-ploughed road. (Center:) Medium RFE for partially patchy road. (Right:) Low RFE for patchy snow and dry surfaces that can lead to slippery roads.

probability indicates a 'dry' patch while a high score indicates a snow patch.

Next, the individual patch scores are subjected to four levels of patch quantization using (4). Here, an 'ice' patch would typically be identified as $s_i = [2, 3]$. The quantization thresholds are empirically determined to maximize patch classification accuracy. The resultant quantized drivable surface patches are shown in Fig. 9.

$$s_i = \begin{cases} 1, & 0 \le p_{s1}^i < 0.05, => \text{dry road}, \\ 2, & 0.05 \le p_{s1}^i < 0.5, => \text{more dry, less snow}, \\ 3, & 0.5 \le p_{s1}^i < 0.8, => \text{more snow, less dry}, \\ 4, & 0.8 \le p_{s1}^i \le 1, => \text{snow}. \end{cases} \quad (4)$$

3) RFE Classification: The quantized and probabilistic patch matrices are analyzed for patterns suggestive of well-ploughed/uniform patches in high RFE roads versus irregularly placed dry road and snow patches in unploughed roads, or black ice formations. Due to data set size and class imbalance constraints, a rule-based method is utilized for RFE classification from the patch matrices. We initially applied structural similarly index metric (SSIM) [26] and coefficient of variation to quantify regular vs. irregular patchiness. However, we observe that the domain specific rule-based model in (5) results in the most accurate prediction of RFE_i. The conditions for (5) are domain-motivated, such that if at least two out of five columns exist with average patch probability score per column less than 0.05, it would indicate at least two well-ploughed surfaces under the vehicle wheels, implying a high RFE surface. However, if at least two out of five columns have medium average patch probability score (denoted by indicator function: $\mathbb{1}[0.05 < \bar{p}^i{}_{s1}|col < 0.12]$), then the road surface would most likely have medium RFE, or low RFE otherwise. (5) constitutes the rule-based classification module for stage-2 in Fig. 1.

$$RFE_i = \begin{cases} 2, & \sum_{col=1}^{5} \mathbb{1}[\bar{p}^i{}_{s1}|col \le 0.05] \ge 2, \\ 1, & \sum_{col=1}^{5} \mathbb{1}[0.05 < \bar{p}^i{}_{s1}|col \le 0.12] \ge 2, \\ 0, & otherwise. \end{cases} \quad (5)$$

IV. EXPERIMENTS AND RESULTS

The performances of stage-1 and stage-2 models in comparison with state-of-the-art methods are given below.

A. Step 1: RSC Classification Performance

The average performance metrics per RSC class for various CNN architectures, SqueezeNet and the feature-based model are presented in Table II. Here, we observe that the SqueezeNet model is the fastest and most accurate model. Also, the CS^2, CS^3 architectures of the CNN model have comparable performances. To further analyze the sensitivity of RSC classification to training data, a limited subset of data is created with 400 training and 400 test images, such that all images in this data are high quality with similar fields of view and similar camera calibrations. The average classification performance metrics per class on this limited data set is presented in Table III. Here, we observe that CS^3, CS^2 CNN model architectures have best classification Acc, followed by the feature-based model. Thus, the SqueezeNet model, although fast, is more sensitive to training data when compared to the proposed CNN models.

It is noteworthy that images that are falsely classified as 'dry' RSC are not further analyzed in stage-2, which can lead to erroneous RFE estimates. Also, we observe that images with slush and wet/water RSC have significantly lower numbers of samples, thereby resulting in 94% and 97.34% Acc, respectively. Future work with high resolution class-balanced data can be directed to reduce false positives for 'dry' RSC conditions and to improve classification Acc for slush and wet RSC classes. Also, the feature-based model utilizes color and textural features for individual images, and it is robust to data set variabilities with lowest computational time complexity due to few parameters. Some examples of the limiting conditions for RSC classification using the feature-based model are shown in Fig. 10. Here, we qualitatively observe that structural similarities between certain ROIs of each test image with the ROIs from images in the training data set cause the misclassifications. Such misclassifications can be significantly reduced by training the models with high quality and similar field of view images.

B. Step 2: RFE Classification Performance

The average performance metrics per class for the stage-2 model are shown in Table IV. Here, we observe that for patch classification into dry road and snow patches, 60 images

TABLE II
AVERAGE PERFORMANCE METRICS PER CLASS FOR STAGE-1 MODELS PERFORMING RSC CLASSIFICATION. THE TEST TIME PER IMAGE IS PRESENTED IN MILISECONDS.

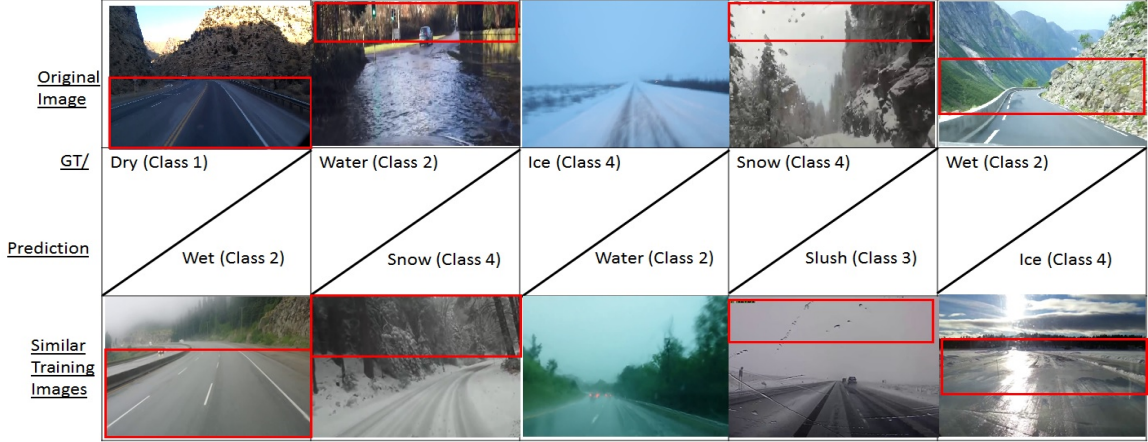| NN-based Model	Train Time (hrs)	Train Acc (%)	Test time (ms)	Acc (%)	PR (%)	RE (%)	# parameters
CNN (CS^1, 10 epochs)	4.6	99.10	1.1	93.96	94	94	134224060
CNN (CS^2, 10 epochs)	4.9	99.66	1.15	95.53	96	96	67118412
CNN (CS^3, 10 epochs)	5.1	99.54	1.26	95.60	96	96	33576812
SqueezeNet (20 epochs)	**0.69**	**98.5**	**4.0**	**97.36**	**97**	**97**	**727428**
Feature-based Model	0.51	93.35	250	91.76	92	92	7200

Fig. 10. Analysis of the misclassified RSC images using feature-based model. For the dry test image, the shadowed drivable surface resembles that of a wet day training image. For the water test image, the limited field of view obstructs the sky, thus resulting in misclassification. The ice image resembles a water training image due to image blurriness, while the snow image is misclassified owing to the the snow particles settled on the windshield resembling a slush training image. The surrounding below horizon for the wet image resembles ice texture from a training image, thus causing misclassification.

TABLE III
AVERAGE PERFORMANCE METRICS PER CLASS FOR STAGE-1 MODELS ON LIMITED TRAINING/TEST DATA.

NN-based Model	Train Time (hrs)	Train Acc (%)	Acc (%)	PR (%)	RE (%)
CNN (CS^2, 10 epochs)	0.5	98.2	92.9	94	93
CNN (CS^3, 10 epochs)	**1.0**	**99.82**	**94.6**	**95**	**95**
SqueezeNet (20 epochs)	0.31	93.1	88.3	89	88
Feature-based Model	0.165	95.83	92.1	92	92

(45 dry, 15 snow) suffer misclassifications. Next, on the test images for RFE classification, 12 images set are misclassified, primarily due to class imbalance. Some examples of the rule-

TABLE IV
AVERAGE PERFORMANCE METRICS PER CLASS FOR STAGE-2 MODEL PERFORMING PATCH AND RFE CLASSIFICATIONS.

Test	Data (# images)	Acc(%)	PR (%)	RE (%)
Patch classification	355 dry road/128 snow	87.5	89	88
RFE classification	114 mixed snow and dry	89.5	88	89

based RFE misclassifications are shown in Fig. 11. Here, we observe that the lack of camera calibration information on public data can lead to a mis-aligned birds-eye view image with the patch quantizated matrix, thereby leading to misclassification. Thus, the use of camera calibrated, similar field of view and larger training data sets can significantly enhance the RFE classification performances.
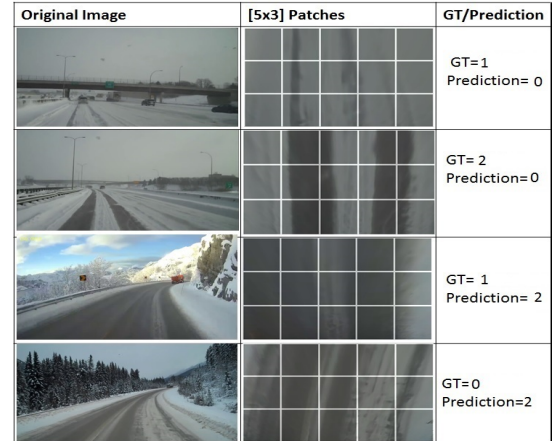

Fig. 11. Analysis of misclassified RFE images. (Row 1:) Low resolution and narrow view of drivable surface leads to a low RFE prediction. Remaining rows suffer from mis-aligned birds-eye view image.

C. Comparison with State-of-the-art Methods

Among existing works, Li et al. [17] applied color transformations to road surface images using an NN-based model and achieved an accuracy rate of about 85% on a local limited data set. Further, in the recent work [9], out of 376 test images, an accuracy of 83% was achieved for classification of dry and wet roads. The comparative evaluation of the proposed models with works that report higher recognition

accuracies are shown in Table V. We observe that for RSC

TABLE V
COMPARATIVE PERFORMANCE ANALYSIS WITH EXISTING WORKS.

Stage-1:	RSC	
Method	# Images	*Acc* (%)
Kim et. al. [13](Dry)	25	88
(Wet)	25	84
(Snow/Ice)	50	78
Average	100	85
Proposed (Dry)	**430**	**96.98**
(Wet/Water)	**376**	**97.34**
(Snow/Ice)	**546**	**99.08**
Average	**1307**	**97.36**
Stage-2:	**RFE**	
Raj et. al. [12]	1180	90
Proposed	114	89.5

classification, the SqueezeNet model significantly improves classification accuracies when compared to existing works. For RFE classification, the proposed method has comparable accuracy to the existing rule-based work in [12].

V. CONCLUSIONS AND DISCUSSION

This work presents a 2-stage hierarchical method for RSC and RFE classification tasks using public domain data sets in stage-1 and 2, respectively. This hierarchical method separates images with 'dry' RSC, that generally incur high RFE, from images with combination drivable surfaces, that can then be further analyzed for patch-based RFE classification. Thus, stage-1 significantly reduces the number of images that require fine-tuned analysis of the ego-lane drivable surface in stage-2. Stage-1 involves feature-learning from trained CNN model kernels, that lead to 94-99% classification accuracies for dry, wet/water, slush and snow/ice conditions. In stage-2, a rule-based model motivated by domain-specific guidelines, is found to achieve up to 89.5% RFE classification accuracy. Future works will be directed towards generating larger annotated data sets that are high quality and camera calibrated to learn the decision boundaries for ego-lane drivable patch quantization and RFE classification tasks. Also, images with multiple RSC labels, e.g., surfaces with snow and water etc., can be analyzed for robust feature learning tasks. Besides, future efforts on application of deep learning models for fusing vehicle dynamics sensors along with the camera data may significantly enhance RFE estimation performances. Since the proposed models have low computational time complexity for processing test images, they are suitable for real-time prediction of drivable conditions for active safety functions and autonomous vehicles.

REFERENCES

[1] U. D. of Tranportation Federal Highway Administration, "How do weather events impact roads ?," Feb 2017. [Online]. Available: https://ops.fhwa.dot.gov/weather/q1_roadimpact.htm

[2] K. Korosec, "2016 was the deadliest year on american roads in nearly a decade," Feb 2017. [Online]. Available: http://fortune.com/2017/02/15/traffic-deadliest-year/

[3] M. Andersson, F. Bruzelius, J. Casselgren, M. Hjort, S. Löfving, G. Olsson, J. Rönnberg, M. Sjödahl, S. Solyom, J. Svendenius *et al.*, "Road friction estimation part ii," 2010.

[4] F. Gustafsson, "Slip-based tire-road friction estimation," *Automatica*, vol. 33, no. 6, pp. 1087–1099, 1997.

[5] J. Casselgren, "Road surface characterization using near infrared spectroscopy," Ph.D. dissertation, Luleå tekniska universitet, 2011.

[6] S. Yoda, H. Okabe, J. Takagi, and T. Yamashita, "Road surface recognition sensor using an optical spatial filter," in *Proceedings of the Intelligent Vehicles' 95 Symposium.* IEEE, 1995, pp. 253–257.

[7] L. Hong, L. Jun, and F. Yanhui, "Road surface condition recognition method based on color models," in *First International Workshop on Database Technology and Applications.* IEEE, 2009, pp. 61–63.

[8] S.-k. Ryu, T. Kim, and E. Bae, "Algorithm and experiment for vision-based recognition of road surface conditions using polarization and wavelet transform," *J Emerg Trends Comput Inf Sci*, vol. 5, no. 10, pp. 739–745, 2014.

[9] C. Panhuber, B. Liu, O. Scheickl, R. Wies, and C. Isert, "Recognition of road surface condition through an on-vehicle camera using multiple classifiers," in *Proceedings of SAE-China Congress 2015: Selected Papers.* Springer, 2016, pp. 267–279.

[10] S. Khaleghian, A. Emami, and S. Taheri, "A technical survey on tire-road friction estimation," *Friction*, pp. 1–24, 2017.

[11] I.-H. Kim, J.-H. Bong, J. Park, and S. Park, "Prediction of drivers intention of lane change by augmenting sensor information using machine learning techniques," *Sensors*, vol. 17, no. 6, p. 1350, 2017.

[12] A. Raj, D. Krishna, K. Shantanu *et al.*, "Vision based road surface detection for automotive systems," in *Applied Electronics (AE), 2012 International Conference on.* IEEE, 2012, pp. 223–228.

[13] J. H. Kim and J. M. Won, "A development of the road surface decision algorithm using svm (support vector machine) clustering methods," *The Journal of The Korea Institute of Intelligent Transport Systems*, vol. 12, no. 5, pp. 1–12, 2013.

[14] L. Hansung, S. Choi, and J. Yang, "Driver assistance apparatus capable of recognizing a road surface state and vehicle including the same," Mar. 21 2017, uS Patent 9,598,088.

[15] Q. Zhao, Q. Zhang, J. Wang, B. Litkouhi, and W. Zhang, "Systems and methods for determining a condition of road surface," Mar. 21 2017, uS Patent 9,598,087 B2.

[16] Q. Zhao, J. Wang, W. Zhang, and B. B. Litkouhi, "Road surface condition detection with recursive adaptive learning and validation," Sep. 22 2015, uS Patent 9,139,204.

[17] J. Zhao, H. Wu, and L. Chen, "Road surface state recognition based on svm optimization and image segmentation processing," *Journal of Advanced Transportation*, vol. 2017, 2017.

[18] S. Roychowdhury and J. Ren, "Non-deep cnn for multi-modal image classification and feature learning: An azure-based model," in *IEEE International Conference on Big Data (Big Data).* IEEE, 2016, pp. 2893–2812.

[19] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[20] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "Squeezenet: Alexnet-level accuracy with 50x fewer parameters and< 0.5 mb model size," *arXiv preprint arXiv:1602.07360*, 2016.

[21] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[22] G. Liu, P. Han, Y. Niu, W. Yuan, Z. Lu, and J.-R. Wen, "Graph-boosted convolutional neural networks for semantic segmentation," in *International Joint Conference on Neural Networks (IJCNN).* IEEE, 2017, pp. 612–618.

[23] A. Torralba and P. Sinha, "Statistical context priming for object detection," in *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, vol. 1. IEEE, 2001, pp. 763–770.

[24] J. Sivic, B. Kaneva, A. Torralba, S. Avidan, and W. T. Freeman, "Creating and exploring a large photorealistic virtual space," in *Computer Vision and Pattern Recognition Workshops, 2008. CVPRW'08. IEEE Computer Society Conference on.* IEEE, 2008, pp. 1–8.

[25] B. C. Russell, A. Torralba, K. P. Murphy, and W. T. Freeman, "Labelme: a database and web-based tool for image annotation," *International journal of computer vision*, vol. 77, no. 1, pp. 157–173, 2008.

[26] M. Menenberg, S. Pathak, H. P. Udyapuram, S. Gavirneni, and S. Roychowdhury, "Topic modeling for management sciences: A network-based approach," in *2016 IEEE International Conference on Big Data (Big Data).* IEEE, 2016, pp. 3509–3518.