Canadian Science Publishing

ARTICLE

Winter road surface condition classification using convolutional neural network (CNN): visible light and thermal image fusion

Ce Zhang, Ehsan Nateghinia, Luis F. Miranda-Moreno, and Lijun Sun

Abstract: During winter, road conditions play a crucial role in traffic flow efficiency and road safety. Icy, snowy, slushy, and wet road conditions reduce tire friction and affect vehicle stability which could lead to dangerous crashes. To keep traffic operations safe, cities spend a significant budget on winter maintenance operations such as snow plowing and spreading of salt and sand. This paper proposes a methodology for automated winter road surface condition classification using convolutional neural network (CNN) and the combination of thermal and visible light cameras. As part of this research, 4244 pairs of visible light and thermal images are captured from pavement surfaces and classified into snowy, icy, wet, and slushy surface conditions. Two single-stream CNN models (visible light and thermal streams) and one dual-stream CNN model are developed. The average F1-Score of dual-stream model is 0.866, 0.935, 0.985, and 0.888 on snowy, icy, wet, and slushy, respectively. The weighted average F1-Score is 0.94.

Key words: winter road surface condition monitoring, winter road maintenance, convolutional neural network, sensor fusion, thermal camera.

Résumé : En hiver, les conditions routières ont un impact crucial sur l'efficacité de la circulation et la sécurité routière. Les conditions glacées, enneigées, boueuses et mouillées sur la route réduisent le frottement des pneus et ont une incidence sur la stabilité du véhicule, ce qui pourrait entraîner des accidents dangereux. Pour assurer la sécurité de la circulation, les villes consacrent un budget important aux opérations d'entretien hivernal, comme le déneigement et l'épandage de sel et de sable. Cet article propose une méthodologie de classification automatisée des conditions de la surface de roulement d'hiver à l'aide d'un réseau neuronal convolutif (RNC) et de la combinaison de caméras thermiques et de caméras à lumière visible. Dans le cadre de cette recherche, 4244 paires d'images de lumière visible et d'images thermiques sont capturées à partir des surfaces de chaussée et classées en conditions de surface neigeuse, glacée, humide et boueuse. Deux modèles RNC à flux unique (lumière visible et flux thermiques) et un modèle RNC à double flux sont développés. Le score F1 moyen du modèle à double flux est de 0,866, 0,935, 0,985 et 0,888 sur neige, glace, mouillée et neige fondante, respectivement. Le score F1 moyen pondéré est de 0,94. [Traduit par la Rédaction]

Mots-clés : surveillance de l'état de la surface des routes l'hiver, entretien des routes l'hiver, réseau neuronal convolutif, fusion de capteurs, caméra thermique.

1. Introduction

In winter, road surface condition greatly affects traffic efficiency and safety, especially in countries that experience long winters and harsh climates. The literature discusses the relationship between road surface condition and crash frequency. Driving conditions often deteriorate during snowfall and ice formation due to a significant decrease in pavement friction and diminished vehicle traction (Norrman et al. 2000).

In cold regions, roads can freeze for a long period each winter, which reduces the friction between the road surface and the vehicle, thereby increasing the stopping distance of the vehicles (Kietzig et al. 2010). Moreover, the reflected sun's glare from snow covered can cause snow blindness, impairing the driver's vision. In the United States, almost 26% of traffic accidents occur on snowy, icy, wet, or slushy roads, and 18% of crashes involving fatalities occur during wintry weather (FHWA 2020).

Winter road maintenance operations are crucial to and the safety of roads. However, these operations incur high monetary costs and cause adverse environmental effects. Each year, governments spend a considerable amount of money on snow plowing and spreading salt and sand on roads to increase the freezing point of the road surface. The cost of winter maintenance, for example, in Ontario, Canada, has been estimated to exceed $100 million per year (Ontario Ministry of Transportation 2016).

Winter in cities such as Montreal, Canada, is extremely long and cold, making road conditions an essential issue for traffic operation. According to weather reports, in Montreal, winter lasts for five months, with an average high of −2.3 °C (daytime) and an average low of −8.9 °C (nighttime). Due to the continental climate, of Montreal, the city experiences several snowy days with more than 1 cm of snow, while in the middle of winter, the average snow cover is 13 cm. Cities like Montreal spend significant resources in maintaining appropriate surface conditions on their roads during winter, to ensure road user safety and traffic operations. In the winter of 2018–2019, for instance, the City of Montreal spent 192 million dollars in winter maintenance, according to the government's annual financial report (City of Montreal 2019).

As an indication of the importance of winter road safety, automated road surface monitoring and data collection systems have emerged in recent years. These systems are designed to reduce potential accidents caused by frozen and slippery roads by gathering

Received 20 September 2020. Accepted 22 June 2021.

C. Zhang, E. Nateghinia, L.F. Miranda-Moreno, and L. Sun. Department of Civil Engineering, McGill University, Room 492, Macdonald Engineering Building, 817 Sherbrooke Street West, Montreal, QC H3A 0C3, Canada.

Corresponding author: Ehsan Nateghinia (email: ehsan.nateghinia@mail.mcgill.ca).

real-time information about road surface conditions and providing feedback to winter road maintenance operators and drivers via on-road warning systems. Using data from automatic monitoring systems, for example, decision-makers in winter operation programs can enhance the efficiency of winter road maintenance by better planning and implementing anti-freezing (anti-icing) procedures and controlling the dosage of chemical solutions.

In recent years, researchers proposed using automated camera-based systems for road surface monitoring. Current systems face the challenge of detecting ice and snow on the road surface. Zhang et al.'s integration of a video monitoring system with Support Vector Machine (SVM) and K-Nearest Neighbors (KNN) performed well in classifying snowy and dry surface conditions (Zhang et al. 2012). Nevertheless, their results indicated the inefficiency of their system in distinguishing wet road surfaces from icy conditions.

Studies have been conducted on the relationship between air temperature and road temperature during the wintery weather conditions, as well as their relationship with pavement surface conditions. Wood and Clark studied the patterns of road surface temperature during wintery weather conditions (Wood and Clark 1999). By analyzing historical and forecast weather information, Kangas et al. attempted to predict the road temperature and to classify road surface conditions as weather conditions change (Kangas et al. 2015). Different studies have developed models for estimating road temperature based on several factors such as air temperature and have used the results of these models to predict road surface conditions (Karsisto and Lovén 2019; Yang et al. 2020).

Junhui and Jianqiang (2010) developed a neural network model that considers road temperature, air temperature, air humidity, season, location, and time as independent variables, and predicts road surface conditions with 90% accuracy. Jonsson et al. (2015) studied utilizing a near-infrared camera and halogen searchlights for surface condition classification, and they demonstrated the accuracy of their system as 80% for dry, 100% for wet, 70% for icy, and 90% for snowy surface conditions.

Despite recent advancements, to the best of our knowledge, very few studies have addressed the performance of thermal cameras for monitoring winter surface conditions. The performance of thermal cameras alone and in combination with visible light (RGB) cameras has not been studied. Furthermore, the use of convolutional neural networks (CNN) for monitoring winter road surfaces is a relatively new approach. Despite the Pan et al. (2018) integration of a CNN model into their camera-based system, the limited set of images has restricted the performance of their model.

In this study, a novel solution for the automatic monitoring and classification of winter road conditions is presented using machine learning techniques applied to visible light and thermal images of the road surface. Specifically, CNN models are trained with thermal and regular video cameras to automatically detect and classify the presence of ice or snow in images.

The proposed solution includes a regular camera to capture visible light range images as well as a thermal camera to capture infrared range or thermal images. The cameras are mounted on a vehicle for data collection, and the images are collected by driving through Montreal's streets in winter. This work involves tuning different CNN structures on a training dataset to discover a promising classifier for winter road conditions and to analyze the impact of each image source such as visible light or infrared.

2. Literature review

Knowing the amount of chemical solution that needs to be applied for maintaining a safe road surface in the winter is essential for road maintenance operations. At first, the tire friction coefficient was used to assess the road surface condition. For instance, Erdogan et al. (2009) proposed a method for determining tire-road friction coefficient based on lateral tire forces.

Due to recent developments in the Intelligent Transportation System (ITS), it became more convenient to determine highway surface conditions using camera-based monitoring systems. Takeuchi et al. (2012) collected 91 images of a highway surface using road surveillance cameras, then they used the pixel intensity histogram to determine five texture features, including mean, contrast, variance, energy, and entropy. They used the K-Means algorithm for clustering those images and achieved an accuracy rate of 84.8% in daytime conditions.

Omer and Fu (2010) presented a road surface monitoring system using GPS-tagged images captured by low-cost cameras mounted on non-dedicated vehicles such as public transportation or police vehicles. They categorize their image dataset into three road surface classes: bare, snow-covered, and tracks, where the tracks represent straight lines caused by vehicle wheels rolling over the bare road surface. They reported an accuracy of 86% for their proposed system using the image edge detection operation and SVM classification. Their finding concluded that there is a significant color difference between snow-covered and bare areas, and that image resolution, camera angles, and lighting conditions can affect the accuracy.

Using smartphones, Linton and Fu (2015) designed a system for monitoring winter road surface conditions that can be mounted inside vehicles in such a way as to provide a clear view of the road ahead. The system captured and sent time-stamped GPS-tagged images to a server for classification into three road surfaces including bare, partly snow-covered, and fully snow-covered. Over 16 000 collected images, the average classification accuracy was 73%. Zhang et al. (2012) developed a video-based method for detecting snow cover using an edge background model. They applied neural network, SVM, and KNN classifiers on a set of image features extracted from the co-occurrence matrix of adjacent pixel values. The KNN model outperformed the other two classifiers by achieving an average accuracy of 93% when classifying images into heavy snow, mild snow, and dry covers. However, their system did not distinguish well between wet and icy road conditions.

Jokela et al. (2009) constructed a surface monitoring system that uses a stereo camera, an imaging spectrometer, and remote surface temperature and state sensors. The feature extraction was based on light polarization changes of road surface reflections and granularity analysis. They reported 90% accuracy in detecting icy, wet, snowy, and dry road conditions. In a study conducted by Pan et al. (2018), the accuracy of a CNN model embedded in an image-based road condition monitoring system was assessed at 76.7%. They examined bare, less than half snowy, half snowy, more than half snowy, and fully snowy road surfaces in their dataset.

The Road Weather Information System (RWIS) has been used for road condition prediction by combining weather data such as temperature, humidity, wind speed, and wind direction. However, it could be misleading to rely solely on this information. As an example, it is impractical to use the road temperature to differentiate between wet and icy conditions because icing condition may occur at road temperatures ranging from −20 °C to 0 °C instead of precisely at 0 °C since de-icing chemicals decrease the freezing point of the surface fluid (Omer and Fu 2010; Tabuchi et al. 2003).

Jonsson (2011) evaluate whether a dataset obtained from an RWIS field station equipped with a near-infrared camera and an infrared searchlight is sufficient for an accurate road condition classification. Their system achieved 91% accuracy for dry, 100% for icy, 100% for snowy, 74% for wheel track, and 100% for wet class.

McFall and Niitula (2002) introduced an audiovisual system consisting of an analog high-resolution grayscale camera and a ring buffer for audiovisual synchronization. Using an audio signal spectrogram and an image's edge map as the feature vector, a KNN classifier achieved an accuracy of 95% for icy conditions, 81% for snowy, and 97% for wet, and low accuracy of 23% for dry conditions.

Casselgren (2011) used polarized short-wave infrared light (SWIR) sensors for classifying road conditions. Near-infrared devices emit light of different wavelengths to the road surface.

Fig. 1. ThermiCam wide-thermal camera by FLIR.

Based on how the surface absorbs, scatters, and polarizes the emitted light, they can detect different road conditions. Despite the high average accuracy rate of 93%, the setup was stationary, with a limited, fixed field of view.

Jonsson et al. (2015) classified road surface conditions based on spectral-based image analysis into four types: dry, wet, icy, and snowy. They built a mobile imaging system equipped with a near-infrared camera and two halogen searchlights. With the embedded with an SVM classifier, their system achieved an accuracy of 94% for dry conditions, 94% for wet, 97% for icy, and 98% for snowy road surface conditions.

3. Methodology

There are various steps in this research methodology: (1) establishing a data collection system; (2) collecting and labeling data; (3) implementing and validating CNN models.

3.1. Data collection system

In the data collection system, an infrared camera, ThermiCam wide model built by FLIR (shown in Fig. 1), was used for collecting thermal images from the road surface. With an output resolution of 368 × 296 pixels and a temperature range from −34 °C to 74 °C, this thermal camera can capture 15 frames per second (fps). Additionally, a visible light camera, GoPro Hero 7, is integrated to collect visible light images. GoPro Hero 7 can reduce the effects of vibrations caused by vehicle movements when capturing images. The visible light image data was collected at 30 fps with an output resolution of 1920 × 1080 pixels.

Both cameras, visible light and infrared, were mounted on the front of the vehicle, rather than the back, so that generated heat by vehicle exhaust would not affect the thermal images. As shown in Fig. 2, the thermal camera is installed on the left side of the vehicle's front from the driver's perspective, and the visible light camera is installed next to it without blocking its field of view. The angles of view of both cameras were aligned with one another to capture images from the same part of the road surface.

3.2. Dataset collection and preparation

The video data collection took place on February 28, 2020, when the weather was cloudy with no significant precipitation, and the temperature ranging between −9 °C and −6 °C. Nevertheless, Montreal experienced widespread snowfall the day before data collection. The data were collected for 3 h in several areas of Montreal that had various road surface conditions. After the video data was collected, visible light and thermal footages were sampled at one frame per second, and a total of 4244 images were selected manually from all the sampled images and added to the database.

Despite both cameras being in alignment, a pixel-to-pixel matching is performed between the visible light image and the thermal image to determine the exact overlapping area. In Figs. 3*a* and 3*b*, the red boxes show the overlapping areas on the original visible light and thermal images. The unique coordinates of the four corners of the two red boxes have been determined by manually matching multiple pairs of images. Figures 3*c* and 3*d* show the results after determining the pixel coordinates of the four corners of each box and cropping the original images into these boxes. Thermal images were cropped to 188 × 368 pixels, and cropped visible light images were resized to the same size as the thermal images.

Based on the details of both visible light and thermal images, the pairs of images were manually labeled into four classes: *snowy*, *icy*, *wet*, and *slushy*. Figure 4 shows the visible light and thermal samples for each labeled class. The *snowy* class (Figs. 4*a* and 4*b*) implies that the road surface is snow-covered, the grayscale image is mostly white, and the thermal image is brighter than the icy class. In the *icy* class (Figs. 4*c* and 4*d*), the road surface is transparent (ice-covered) in the grayscale image and dark in the thermal image. An image in the *wet* class (Figs. 4*e* and 4*f*) shows a road surface with no clear evidence of ice or snow but only water. In the *slushy* class (Figs. 4*g* and 4*h*), the road is covered with a mixture of water and ice or water and snow, the visible light image shows melting snow or ice, and the thermal image is brighter than *snowy* and *icy* samples and darker than *wet* samples.

The base dataset was built on collected images of these four classes. However, three alternative datasets were created to address some of the shortcomings of the base dataset. A summary of the base dataset and its three alternatives is presented in Table 1.

1. The "*Multiple*" dataset: in addition to the four aforementioned classes, an additional class called "*multiple*" is inserted (Figs. 4*i* and 4*j*). This class contains images with more than one but separate surface conditions such as snowy and icy, snowy and slushy, wet and slushy, or wet and icy conditions.
2. The "*Artificial*" dataset: to deal with the uneven distribution of samples per class, especially the insufficiency of "*snowy*" images, an image generator from the OpenCV library is used (Bradski and Kaehler 2008). Accordingly, an additional 241 artificial "*snowy*" images are generated by flipping, rotating, masking, or cropping the original snowy images.
3. The "*Split*" dataset: to reduce the number of parameters and to increase the sample size, the original images of the "*Base*" dataset are split horizontally into two images with the same dimension of 188×184 pixels each.

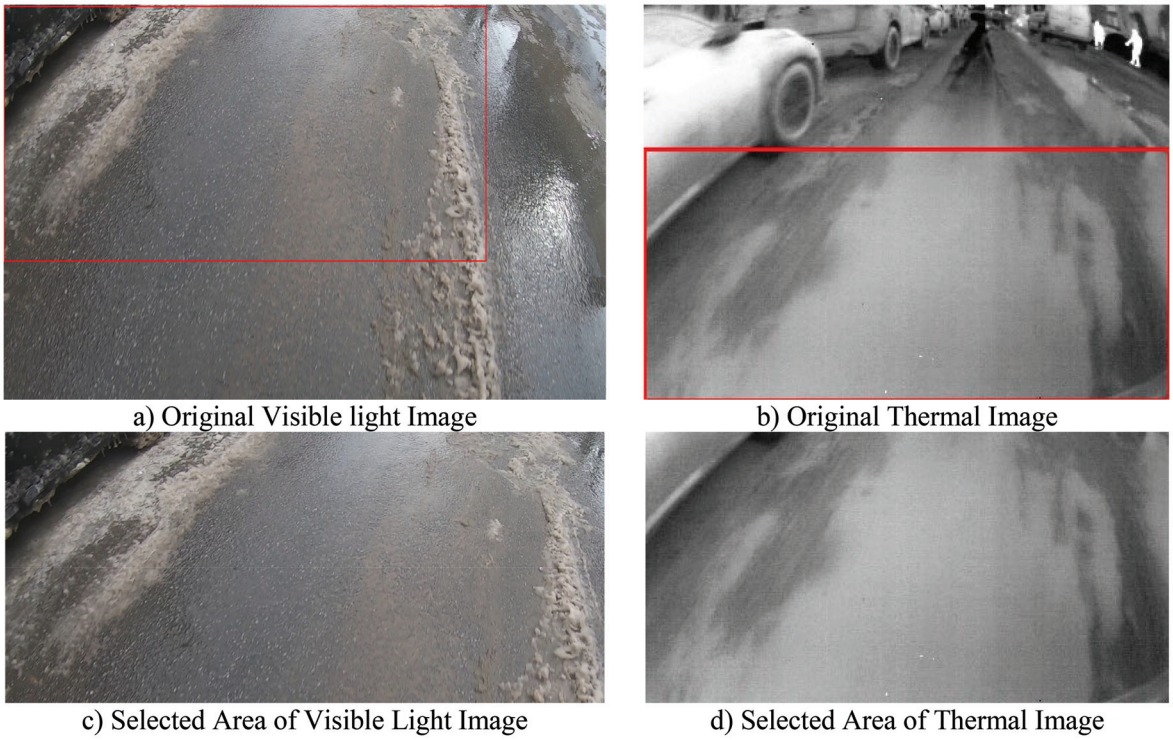
3.3. CNN models implementation

The convolutional neural network (CNN) has demonstrated promising performance in various applications including image classification. This paper presents an automated system for winter road surface condition detection and classification using CNN models constructed from a benchmark model called the VGG-16 (Simonyan and Zisserman 2014). The VGG-16 has 16 layers and more than 138 million parameters, and it is designed for big datasets with many classes such as the ImageNet dataset with 15 million images and 10 thousand classes. Since the dataset for this study contains 4244 labeled images from four classes, the proposed CNN is restructured into seven layers to increase processing speed while maintaining performance.

There are four types of layers in a CNN model: convolutional, max pooling, flattening, and fully connected layers. Each convolutional layer includes feature maps, kernels, and padding. The first feature map is the original image. The kernel is a digital filter (matrix) that operates only across its receptive field and its size and stride are hyperparameters. The padding is used to preserve the input size in cases that some part of the receptive field of the kernel is outside the

Fig. 2. Thermal camera setup (front-view and side-view). [Colour online.]

Fig. 3. The original visible light (RGB) and thermal images with matching boxes: (*a*) original visible light image, (*b*) original thermal image, (*c*) selected area of visible light image, and (*d*) selected area of thermal image. [Colour online.]



boundary of the feature map. Some of the convolutional layers are followed by a max-pooling layer that reduces the size of the feature map by taking the maximum value of a square window (3 × 3 or 2 × 2 pixels) from the current feature map and inserting it in the reduced feature map. After the last pooling layer, there is a flattening layer that transforms the last two- or three-dimensional feature map into a one-dimensional array (feature vector) and passes the feature vector to the fully connected layers. The fully connected layers abstract a feature vector into a decision vector of the same size as the number of classes.

The proposed CNN model contains seven layers in three blocks. The first block includes a convolutional layer with 16 kernels (3 × 3 receptive field) and a max-pooling layer (3 × 3 pool size). The second block includes a convolutional layer with 32 kernels (3 × 3 receptive field) and a max-pooling layer (3 × 3 pool size). The third block includes 2 convolutional layers with 32 kernels (3 × 3 receptive field) and a max-pooling layer (3 × 3 pool size). The Rectified Linear Unit (ReLU) function, formulated in eq. 1, has been used as the activation function of the convolutional layers.

$$f(x_i) = \max(0, x_i) \quad (1)$$

where x_i is the output of the feature map and input to the ith neuron of the activation map, and $f(x_i)$ is the output of the same activation neuron.

The *SoftMax* function (Jang et al. 2016), formulated in eq. 2, is used in the neurons of the last fully connected layer. The number of neurons in the last layer is equal to the number of classes and each *SoftMax* function calculates the probability of its corresponding class. Then the class with the maximum probability will be chosen as the decision or label.

Fig. 4. Visible light and thermal images of *snowy*, *icy*, *wet*, *slushy*, and *multiple* classes: (*a*) snowy-visible light, (*b*) snowy-thermal, (*c*) icy-visible light, (*d*) icy-thermal, (*e*) wet-visible light, (*f*) wet-thermal, (*g*) slushy-visible light, (*h*) slushy-thermal, (*i*) multiple-visible light, and (*j*) multiple-thermal. [Colour online.]

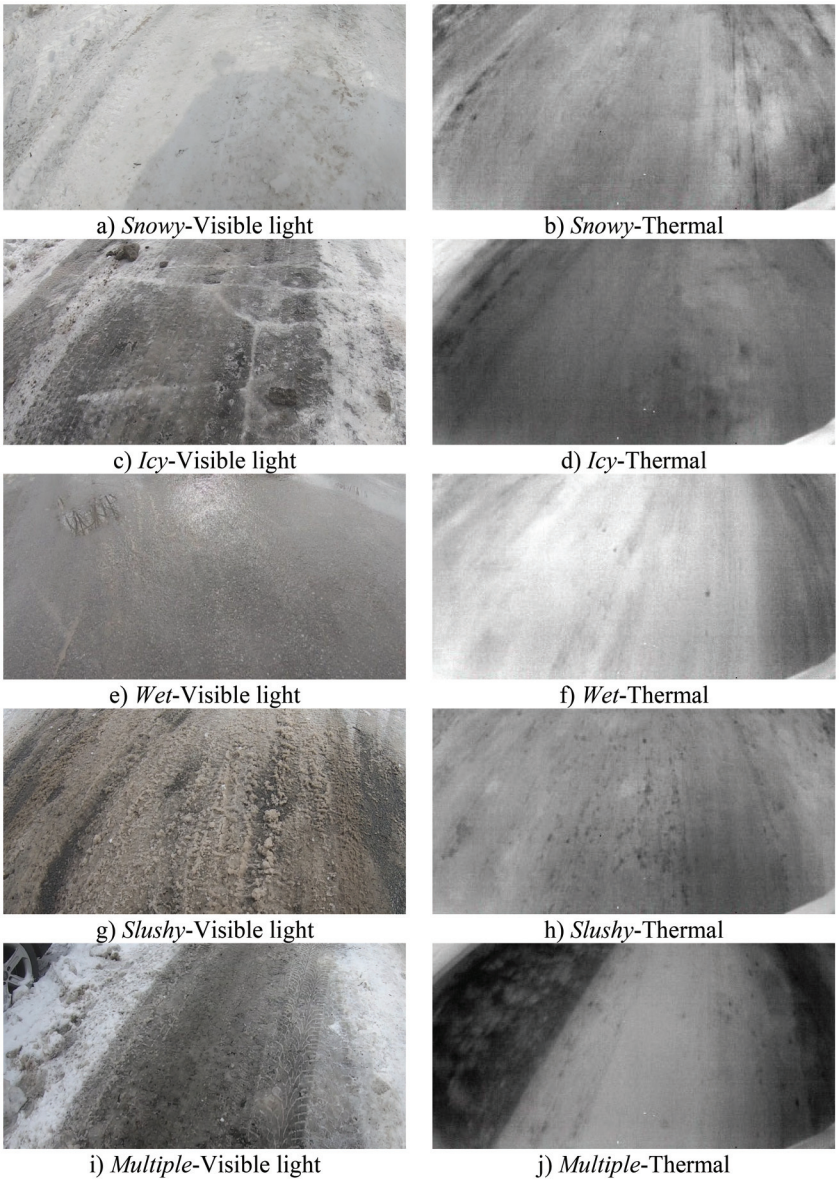


Table 1. Summary of sample distributions per class of the four datasets.

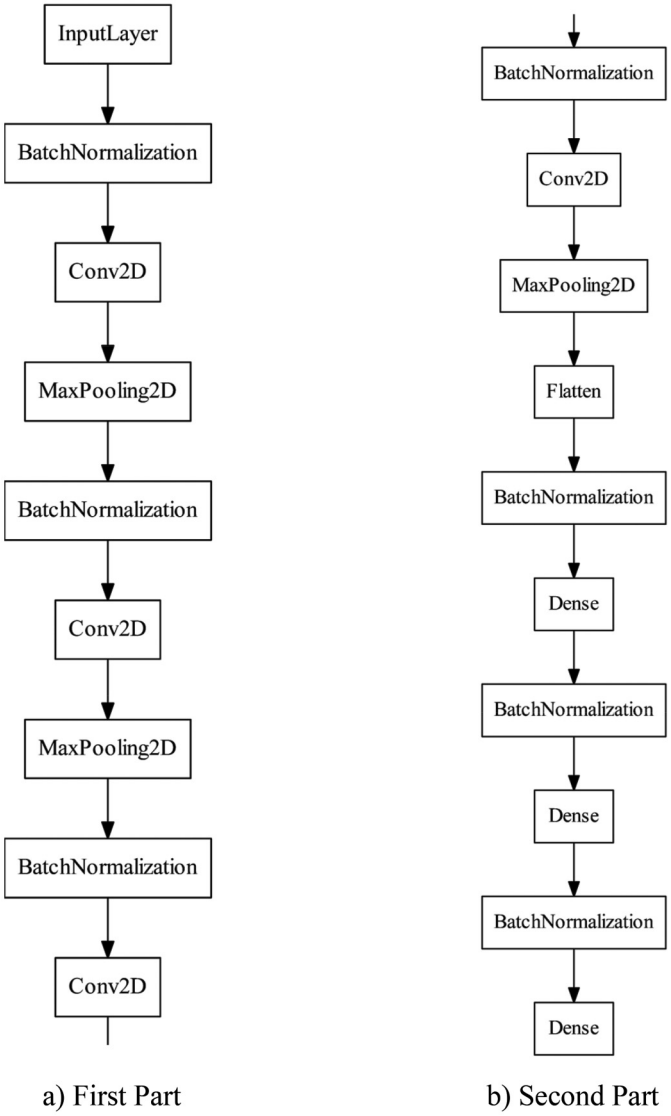
	Base		Artificial		Multiple		Split	
Classes	Training	Test	Training	Test	Training	Test	Training	Test
Snowy	238	50	408	121	223	65	459	117
Icy	1302	334	1314	322	1319	317	2629	643
Wet	1204	301	1208	297	1228	277	2407	603
Slushy	651	164	658	157	634	181	1295	335
Multiple	—	—	—	—	432	119	—	—
Sum	3395	849	3588	897	3836	959	6790	1698

$$P(Y=i|x_j)=g(x_i)=\frac{\exp(x_i)}{\sum_{j=0}^{K}\exp(x_j)} \quad (2)$$

where x_i is the input to the activation unit of ith neuron, and $g(x_i)$ is the probability of choosing ith class.

To avoid overfitting toward training set in early iterations of the learning process, two regularization methods, Dropout (DO) and Batch-Normalization (BN) have been used in the structure of the proposed CNN. Dropout operation helps to avoid overfitting by temporarily and randomly removing some of the learned parameters

Fig. 5. The single-stream CNN model: (*a*) first part and (*b*) second part.



from the network (Srivastava et al. 2014). Batch-Normalization (Ioffe and Szegedy 2015) is an alternative regularization method that prevents the training process from getting stuck in the saturated regimes of non-linearities (Shimodaira 2000).

Three CNN models have been implemented using the collected datasets. First, a single-stream CNN is built by using only visible light images. Second, another single-stream CNN is trained with only thermal images. Third, a dual-stream CNN is designed to capture the combination of both sources of visible light and thermal images. The dual-stream CNN has two independent inputs, and each input stream has one convolutional block fed by one of the image sources.

Both single-stream CNN models, whether using visible light images or thermal images, have the same structure and input size (displayed in Fig. 5). The image on the right shows the continuation of CNN's part on the left. Conv2D and MaxPooling2D are two-dimensional convolutional and max-pooling layers. Since the thermal images have only one channel, the visible light images are also converted to grayscale images. Therefore, both input sources are two-dimensional matrices with 188 × 368 pixels. Each dense box is one layer of the fully connected layers, and the last *Dense* box has only four neurons producing the probability of choosing a class.

The structure of the dual-stream CNN is shown in Fig. 6. A dataset composed of visible light and thermal images is used to tune the dual-stream CNN model. The last feature map of each stream is flattened and merged into a one-dimensional feature vector of 4992 elements, which is twice the size of the flattened vector of each single-stream CNN. Afterward, the flattened vector is given to a fully connected layer to generate the probability of choosing each class.

A performance comparison between two single-stream networks and the dual-steam network is presented in Section 4. Additionally, the dual-stream network with different weights for each stream, thermal and visible light, is built and analyzed. To alter the weight of a particular input in the dual-stream network, the filter size of the last max-pooling layer can be adjusted to the required level. For example, if the filter size of the last max-pooling layer in the visible light stream is set to 5 × 4, the size of the last feature map of that stream changes to 32 × 4 × 10 instead of 32 × 6×13. Therefore, the size of the flattened feature vector is reduced by 50%, from 2496 to 1280. This means that the visible light stream weights half as much as the thermal input.

4. Results

This work implements CNN with different structures using the Keras backend with TensorFlow (Géron 2019). The coefficients of the deep neural network are learned by optimizing the cost function using the Adadelta method (Zeiler 2012). The average training time for a single-stream CNN model, developed using the base dataset, is 44 seconds per iteration on an NVIDIA GeForce GPU (GTX1060 3GB). The average training time for a dual-stream CNN model is 48 seconds per iteration on this setup. The labels of 4244 images were predicted in 30 s, which implies that each prediction took place in under 10 ms.

4.1. Performance measures

To evaluate the performance of the proposed system, different error measures such as multiple-class average and weighted average of Precision, Recall, and F1-Score are used. Table 2 shows the confusion matrix obtained by the dual-stream CNN model on the test set of the base dataset. The rows of the confusion matrix are the observed labels of each pair of visible light and thermal images, and the columns are the predicted labels by the CNN model.

The Precision measure of a particular class is the ratio of the correctly classified samples of that class to the total number of predicted samples of the same class. For example, the Precision of the snowy class is 42 divided by 47. The Recall measure of a particular class is the ratio of correctly classified samples of that class to the total number of observed samples of the same class. For example, the Recall of snowy class is 42 divided by 50. The Precision and Recall can be calculated for each class individually. Finally, the F1-Score (eq. 3) measure is estimated by calculating the harmonic mean of Precision and Recall.

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{3}$$

To compare different CNN models, each measure is aggregated in two ways. First, the average of each measure over the four classes is calculated (eqs. 4–6).

$$\text{Precision}_{\text{average}} = \frac{\sum_{i=1}^{4} \text{Precision}_i}{\sum_{i=1}^{4} 1 = 4} \tag{4}$$

$$\text{Recall}_{\text{average}} = \frac{\sum_{i=1}^{4} \text{Recall}_i}{\sum_{i=1}^{4} 1 = 4} \tag{5}$$

Fig. 6. The dual-stream CNN model: (*a*) first part and (*b*) second part.

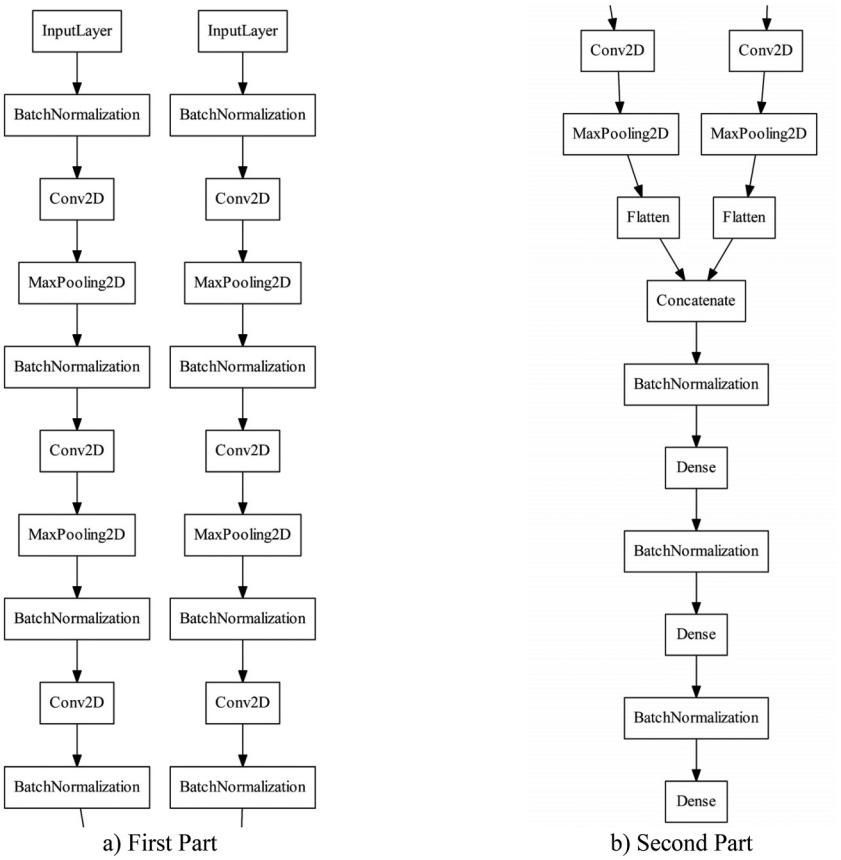


Table 2. Confusion matrix of the dual-stream CNN over the test set of the base dataset.

		Classified (predicted) label					Error measure		
		Snowy	Icy	Wet	Slushy	Total	Precision	Recall	F1-score
Observed label	Snowy	42	8	0	0	50	0.894	0.840	0.866
	Icy	5	317	0	12	334	0.922	0.949	0.935
	Wet	0	2	297	2	301	0.983	0.987	0.985
	Slushy	0	17	5	142	164	0.910	0.866	0.888
	Total	47	344	302	156	849	—	—	—
Average	—	—	—	—	—	—	0.927	0.910	0.918
Weighted average	—	—	—	—	—	—	0.940	0.940	0.940

$$\text{F1-Score}_{\text{average}} = \frac{\sum_{i=1}^{4} \text{F1} - \text{Score}_i}{\sum_{i=1}^{4} 1 = 4} \quad (6)$$

Additionally, the weighted average of each measure is calculated (eqs. 7–9), where the weights are proportional to the number of samples in each class.

$$\text{Precision}_{\text{weighted}} = \frac{\sum_{i=1}^{4} \text{Precision}_i \times n_i}{\sum_{i=1}^{4} n_i} \quad (7)$$

$$\text{Recall}_{\text{weighted}} = \frac{\sum_{i=1}^{4} \text{Recall}_i \times n_i}{\sum_{i=1}^{4} n_i} \quad (8)$$

$$\text{F1-Score}_{\text{weighted}} = \frac{\sum_{i=1}^{4} \text{F1} - \text{Score}_i \times n_i}{\sum_{i=1}^{4} n_i} \quad (9)$$

where i is the class ID that corresponds to the snowy, icy, wet, and slushy classes, and n_i is the number of samples in the ith class.

The average Precision, Recall, and F1-Score of the dual-stream CNN model are 92.7%, 91.0%, and 91.8%, respectively. Considering the number of samples in each class, the weighted average of these three measures is 94.0%. In the following sections, in-depth sensitivity analysis and comparison of different models and datasets have been discussed. First, the dual-stream model is compared with two single-stream models. Second, the ratio of the combination of each stream in the dual-stream model is reduced by 50%. Third, the performance of the dual-stream model on alternative datasets (discussed in Table 1) is assessed.

4.2. Input configuration evaluations

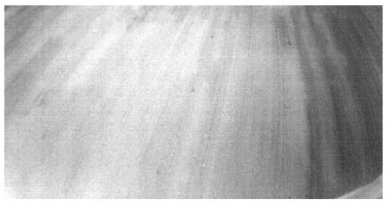
The two single-stream and one dual-stream CNN models were fine-tuned and optimized with the training set of the base dataset, and their performances are evaluated on the test data and are presented in Table 3. In addition to the classification results

Table 3. Classification results of single-stream and dual-stream CNN models.

		Correct predictions (TP)				Average performance			Weighted average performance		
Set	CNN model	Snowy	Icy	Wet	Slushy	Precision	Recall	F1-Score	Precision	Recall	F1-Score
Training	Visible Light	236	1302	1204	651	1.00	1.00	1.00	1.00	1.00	1.00
	Thermal	235	1288	1198	641	0.99	0.99	0.99	0.99	0.99	0.99
	Dual-stream	238	1302	1203	650	1.00	1.00	1.00	1.00	1.00	1.00
	Ground-truth	238	1302	1204	651	—	—	—	—	—	—
Test	Visible Light	35	299	296	144	0.89	0.86	0.87	0.91	0.91	0.91
	Thermal	34	307	271	112	0.85	0.80	0.82	0.86	0.85	0.85
	Dual-stream	42	317	297	142	0.93	0.91	0.92	0.93	0.94	0.94
	Ground-truth	50	334	301	164	—	—	—	—	—	—

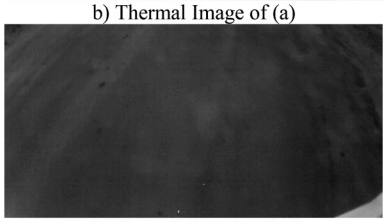
Fig. 7. Two examples of winter road condition classification: (*a*) visible light image of a snowy condition, (*b*) thermal image of (*a*), (*c*) visible light image of an icy condition, and (*d*) thermal sample of (*c*).

a) Visible Light Image of a S*nowy* Condition



b) Thermal Image of (a)

c) Visible Light Image of an I*cy* Condition



d) Thermal Sample of (c)

of each configuration, the ground truth (the number of observed samples in each category) of training and test sets are presented in the 4th and 8th rows of Table 3.

Amongst the two single-stream CNN models, the model built by the visible light images performs better than the one using the thermal images. Using only visible light images, the correct predictions of the snowy, icy, wet, and slushy samples are 35, 299, 296, and 144, respectively. The thermal camera is used as a complementary source of information, and CNN built by thermal images provides better results on icy class, correctly predicting 317 of the test samples compared to 299 correct predictions by the visible light CNN model. The results are similar in the snowy class, but on wet and slushy classes, using the visible light CNN provides better results than thermal CNN because the images of these classes contain more color information.

The dual-stream CNN model improves the classification rate on snowy and icy classes by 14% and 5.3%. These two classes are crucial for the objective of such a system. Moreover, the dual-stream CNN maintains the same performance as the visible light CNN on *wet* and *slushy* classes. This means that the combination of the visible light and thermal image sources provides more information for the CNN model regarding the classification of winter road surface conditions.

The simple and weighted average values of Precision, Recall, and F1-Score on the test set also indicate the superiority of the dual-stream model. The average Precision, Recall, and F1-Score have been improved by 4%, 5%, and 5% compared to the visible light model; and improved by 8%, 11%, and 10% compared to the thermal model.

Figures 7*a* and 7*b* illustrate the visible light and thermal images of a snowy sample. The predictions by visible light, thermal, and dual-stream CNN models are snowy, slushy, and snowy conditions. Figures 7*c* and 7*d* illustrate the visible light and thermal images of an icy sample. The predictions by visible light, thermal, and dual-stream CNN models are snowy, snowy, and icy conditions. In these two examples, only the dual-stream model has correctly predicted the winter road condition.

4.3. Tuning input ratio for the dual-stream CNN model

In the dual data stream CNN model, the ratio of the combination of visible light images and thermal images can be adjusted based upon the concept discussed in Section 3.3. Table 4 presents the classification results of the same CNN configuration using different weight adjustments for each stream. First, the weight of the thermal stream is set to 0.5, while the weight of the visible light stream weight remains 1; second, the weight of the visible light stream is set to 0.5, while the weight of the thermal stream remains 1; third, the weights of both streams are set to 0.5. The fourth row of Table 4 shows the results when both weights remain 1 and no change is applied.

Results from the test set show that reducing the weight of the thermal image stream by 50%, slightly increases the number of correct predictions of snowy and slushy classes by 1 and 5 samples, respectively. Besides, decreasing the weights of both image streams increases the weighted Precision, Recall, and F1-score by no more than 1%. This suggests that a CNN model with a smaller fully connected block, (i.e., halving the first layer of the fully connected block) provides almost the same promising results. In general, the performance measures are not significantly impacted by reducing the number of neurons of one data stream, suggesting that the other data stream compensates for any errors.

4.4. Performance evaluation of alternative datasets

The dual-stream CNN model is chosen as the best model for further analysis on different configurations of the database. In this model, the weights of each image stream are set to 1. In addition to the base dataset, three variations, outlined in Table 1, are built and used for training and evaluating the selected CNN model. Table 5 compares the performance of the dual-stream model on the four

Table 4. Classification results of the dual-stream CNN with different input ratio.

		Correct prediction (TP)				Average performance			Weighted average performance		
Set	Stream with halved neurons	Snowy	Icy	Wet	Slushy	Precision	Recall	F1-score	Precision	Recall	F1-score
Training	Thermal	238	1301	1204	651	1.00	1.00	1.00	1.00	1.00	1.00
	Visible light	238	1302	1204	651	1.00	1.00	1.00	1.00	1.00	1.00
	Both	238	1300	1204	650	1.00	1.00	1.00	1.00	1.00	1.00
	None	238	1302	1203	650	1.00	1.00	1.00	1.00	1.00	1.00
	Ground-truth	238	1302	1204	651	—	—	—	—	—	—
Test	Thermal	43	315	296	147	0.92	0.92	0.92	0.94	0.94	0.94
	Visible light	36	316	297	149	0.92	0.89	0.91	0.94	0.94	0.94
	Both	45	318	298	143	0.94	0.93	0.93	0.95	0.95	0.95
	None	42	317	297	142	0.93	0.91	0.92	0.94	0.94	0.94
	Ground-truth	50	334	301	164	—	—	—	—	—	—

Table 5. Classification results of the dual-stream CNN on different datasets.

		Average performance			Weighted average performance		
Set	Scenarios	Precision	Recall	F1-Score	Precision	Recall	F1-Score
Training	Base	0.999	0.999	0.999	0.999	0.999	0.999
	Artificial	1.000	1.000	1.000	1.000	1.000	1.000
	Multiple	0.998	0.995	0.997	0.998	0.998	0.998
	Split	0.998	0.999	0.998	0.999	0.999	0.999
Test	Base	0.927	0.910	0.918	0.940	0.940	0.940
	Artificial	0.936	0.939	0.937	0.944	0.944	0.944
	Multiple	0.858	0.859	0.858	0.869	0.891	0.879
	Split	0.873	0.872	0.873	0.906	0.906	0.906

different scenarios. The F1-score of the CNN model on the test set of the base dataset is 94.0%. When artificial snowy images are created and added to the database, the F1-score increases to 94.4%, which mostly increases the correct predictions of the snowy class.

Adding an extra class of multiple patterns to the database decreases the F1-score on the test set by 6.1%. This occurs because the samples of the multiple class have a mixture of snowy, icy, wet, and slushy patterns, which makes the learning and optimization of the model's coefficients difficult. However, the F1-Score of 87.9% implies that the dual-stream CNN model performs well when used to classify images with more than one surface condition.

Finally, the performance of the model is evaluated by splitting the original image into two smaller images of 188 $\times$ 184 pixels. In this configuration, the weighted average F1-score of the CNN model is decreased by 3.4% which can be explained by the reduced number of parameters in the model, and by the fact that the smaller images may have less informative pixels. Although the performance of the dual-stream model is degraded, the capacity of the system is increased because it can classify more images at the same time. Besides, since an image is split into two images, the CNN model can classify two surface conditions per image with an F1-Score of 90.6%.

5. Conclusion

Automatic winter road condition monitoring and classification is crucial for winter maintenance operations, especially in countries with a long and cold winter like Canada. This research develops an image-based road monitoring method based on the fusion of the thermal and visible light cameras and evaluates its performance by using data collected from the road surface in Montreal during the winter. Several CNN models are implemented and tested, including two single-stream models that use either visible light or thermal images as their input source, and a dual-stream model that uses both visible light and thermal image images.

The F1-Score results show that the dual-stream CNN model outperforms the two single-stream CNN models. The F1-Score of the dual-stream model is 0.866 for snowy, 0.935 for icy, 0.985 for wet, and 0.888 for slushy surface conditions. Furthermore, the comparison between the two single-stream models reveals that the classification of snowy, wet, and slushy images relies more on color information from the visible light camera, but *icy* images are detected and classified better by the thermal camera because they show sharper temperature maps. Moreover, the comparison of different input weight adjustments on the dual-stream CNN setup indicates that reducing the weight of each data stream or both have a negligible impact on the system performance.

As part of this research, four dataset variations derived from the collected images are used for the sensitivity analysis. Adding artificially created *snowy* images to the dataset improves significantly the performance of the dual-stream CNN on *snowy* class; however, it encountered an overfitting issue on the training set. In another scenario, splitting original images into small sub-images degrades the classification performance because the split sub-images contain less relevant information, However, the system becomes capable of classifying two surface conditions in each image with an F1-Score of 90.6%.

There are some limitations to this research. First, a fixed setup of the thermal and visible light cameras is required because the difference in the field of view between visible light and thermal cameras may lead to an unwanted error in the pixel-to-pixel matching of visible light and thermal images. Second, in this work, a few hours of winter surface condition data have been collected. As for future work, the CNN models can benefit from a longer period of data collection in different street types under diverse weather conditions during the winter season. Additional classes such as black ice could also be added. Moreover, instead of labeling the entire image with a single category, every single image could be annotated and labeled by the cover type. Moreover, a 3D imaging system such as LiDAR or stereo camera with depth information can also be integrated into the system.

Data availability

The Montreal Winter Road Surface Dataset used to support the findings of this study is available from the corresponding author upon request.

Conflicts of interest

The authors declare that there are no known conflicts of interest associated with this publication.

Funding statement

Funding for this project was provided in part by the Natural Sciences and Engineering Research Council.

Credit author statement

Ce Zhang: conceptualization, data curation, formal analysis, investigation, methodology, software, validation, visualization, writing - original draft. Ehsan Nateghinia: conceptualization, data curation, formal analysis, methodology, validation, writing - review & editing. Luis F. Miranda-Moreno: conceptualization, funding acquisition, methodology, project administration, resources, supervision, writing - review and editing. Lijun Sun: conceptualization, methodology, Supervision, writing - review and editing.

References

Bradski, G., and Kaehler, A. 2008. Learning OpenCV: Computer vision with the OpenCV library. O'Reilly Media, Inc.

Casselgren, J. 2011. Road surface characterization using near infrared spectroscopy. Doctoral thesis, Department of Engineering Sciences and Mathematics, Fluid and Experimental Mechanics, Luleå tekniska universitet, Luleå.

City of Montreal. 2019. Rapport financier annuel. [In French.] Available from http://ville.montreal.qc.ca/portal/page?_pageid=43,143504434&_dad=portal&_schema=PORTAL.

Erdogan, G., Alexander, L., and Rajamani, R. 2009. Friction coefficient measurement for autonomous winter road maintenance. Vehicle System Dynamics, **47**(4): 497–512. doi:10.1080/00423110802220554.

FHWA. 2020. Available from https://ops.fhwa.dot.gov/weather/q1_roadimpact.htm.

Géron, A. 2019. Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems. O'Reilly Media, Inc.

Ioffe, S., and Szegedy, C. 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Proceedings of the 32nd International Conference on Machine Learning, Lille, France, 7-9 PMLR. pp. 448–456.

Jang, E., Gu, S., and Poole, B. 2016. Categorical reparameterization with Gumbel-Softmax. arXiv. reprint arXiv:1611.01144.

Jokela, M., Kutila, M., and Le, L. 2009. Road condition monitoring system based on a stereo camera. In IEEE 5th International Conference on Intelligent Computer Communication and Processing, Cluj-Napoca, Romania, 27–29 August 2009. IEEE. pp. 423–428. doi:10.1109/ICCP.2009.5284724.

Jonsson, P. 2011. Road condition discrimination using weather data and camera images. In 14th International IEEE Conference on Intelligent Transportation Systems (ITSC). Washington, D.C., USA, 5–7 Oct. 2011 IEEE. pp. 1616–1621. doi:10.1109/ITSC.2011.6082921.

Jonsson, P., Casselgren, J., and Thörnberg, B. 2015. Road surface status classification using spectral analysis of NIR camera images. IEEE Sensors Journal, **15**(3): 1641–1656. doi:10.1109/JSEN.2014.2364854.

Junhui, L., and Jianqiang, W. 2010. Road surface condition detection based on road surface temperature and solar radiation. In 2010 International Conference on Computer, Mechatronics, Control and Electronic Engineering. Changchun, China 24–26 Aug. 2010. IEEE. pp. 4–7. doi:10.1109/CMCE.2010.5610255.

Kangas, M., Heikinheimo, M., and Hippi, M. 2015. RoadSurf: a modelling system for predicting road weather and road surface conditions. Meteorological Applications, **22**(3): 544–553. doi:10.1002/met.1486.

Karsisto, V., and Lovén, L. 2019. Verification of road surface temperature forecasts assimilating data from mobile sensors. Weather and Forecasting, **34**(3): 539–558. doi:10.1175/WAF-D-18-0167.1.

Kietzig, A.-M., Hatzikiriakos, S.G., and Englezos, P. 2010. Physics of ice friction. Journal of Applied Physics, **107**(8): 081101. doi:10.1063/1.3340792.

Linton, M.A., and Fu, L. 2015. Winter road surface condition monitoring: field evaluation of a smartphone-based system. Transportation Research Record, **2482**(1): 46–56. doi:10.3141/2482-07.

McFall, K., and Niitula, T. 2002. Results of audio-visual winter road condition sensor prototype. In Proceedings of the 11th Standing International Road Weather Commission, Sapporo, Japan, 26–28 Jan.

Norrman, J., Eriksson, M., and Lindqvist, S. 2000. Relationships between road slipperiness, traffic accident risk and winter road maintenance activity. Climate Research, **15**(3): 185–193. doi:10.3354/cr015185.

Omer, R., and Fu, L. 2010. An automatic image recognition system for winter road surface condition classification. In The 13th International IEEE Conference on Intelligent Transportation Systems. Funchal, Portugal, 19–22 September 2010. IEEE. pp. 1375–1379. doi:10.1109/ITSC.2010.5625290.

Ontario Ministry of Transportation. 2016.

Pan, G., Fu, L., Yu, R., and Muresan, M.I. 2018. Road surface condition recognition using a pre-trained deep convolutional neural network. In Transportation Research Board 97th Annual Meeting, Washington, D.C., USA.

Shimodaira, H. 2000. Improving predictive inference under covariate shift by weighting the log-likelihood function. Journal of Statistical Planning and Inference, **90**(2): 227–244. doi:10.1016/S0378-3758(00)00115-4.

Simonyan, K., and Zisserman, A. 2014. Very deep convolutional networks for large-scale image recognition. arXiv. preprint arXiv:1409.1556.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. 2014. Dropout: a simple way to prevent neural networks from overfitting. The Journal of Machine Learning Research, **15**(1): 1929–1958.

Tabuchi, T., Yamagata, S., and Tamura, T. 2003. Distinguishing the road conditions of dry, aquaplane, and frozen by using a three-color infrared camera. *In* Thermosense XXV, Orlando, Fla., USA, 21–25 April 2003. International Society for Optics and Photonics. pp. 277–283. doi:10.1117/12.502153.

Takeuchi, K., Kawai, S., Shibata, K., and Horita, Y. 2012. Distinction of winter road surface conditions using road surveillance camera. *In* The 12th International Conference on ITS Telecommunications, Taipei, Taiwan, 5–8 November 2012. IEEE. pp. 663–667. doi:10.1109/ITST.2012.6425264.

Wood, N., and Clark, R. 1999. The variation of road-surface temperatures in Devon, UK during cold and occluded front passage. Meteorological Applications, **6**(2): 111–118. doi:10.1017/S1350482799001097.

Yang, C.H., Yun, D.G., Kim, J.G., Lee, G., and Kim, S.B. 2020. Machine learning approaches to estimate road surface temperature variation along road section in real-time for winter operation. International Journal of Intelligent Transportation Systems Research, **18**(2): 343–355. doi:10.1007/s13177-019-00203-3.

Zeiler, M.D. 2012. Adadelta: an adaptive learning rate method. arXiv. preprint arXiv:1212.5701.

Zhang, C., Wang, W., Su, Y., and Bai, X. 2012. Discrimination of highway snow condition with video monitor for safe driving environment. In The 5th International Congress on Image and Signal Processing, Chongqing, China, 16–18 October 2012. IEEE. pp. 1241–1244. doi:10.1109/CISP.2012.6469850.