Deep Learning and Spatial Statistics for Determining Road Surface Condition

by

Juan Manuel Carrillo Garcia

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Science
in
Electrical and Computer Engineering

Waterloo, Ontario, Canada, 2019

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Machine Learning (ML), and especially Deep Learning (DL) methods, have evolved rapidly over last years and showed remarkable advances in research areas such as computer vision and natural language processing; however, there are still engineering applications in industries such as transportation where DL methods have not been applied yet or that can be benefited from an integrated approach using DL in addition to other methods. For countries in Northern latitudes one of such applications is Monitoring Road Surface Condition (RSC) during the Winter season for improving road safety and road maintenance operations.

In this study we introduce a novel approach for monitoring of RSC that integrates DL methods and Spatial Statistics (SS) to simultaneously process data from roadside cameras and weather stations to determine automatically the category of snow coverage at sample locations across a region of interest. Our approach integrates the advantages of SS for interpolating spatial variables and the strengths of DL for Computer Vision tasks, particularly for image classification. On one hand SS models serve to understand the spatial autocorrelation of random variables and to determine their expected values in unsampled locations based on a number of near observations. On the other hand, DL models extract relevant patterns from a large number of training images and learn a mapping from input images to a set of predefined labels.

We implement and evaluate our approach using data collected in the province of Ontario during the 2017-2018 Winter season. Specifically, we included data from three separate sources, Environment Canada (EC) Weather stations, Road Weather Information System (RWIS) stations, and roadside cameras from the Ministry of Transportation of Ontario (MTO). To the best of our knowledge, this is the first study that integrates both DL and SS techniques for processing the three data sources with the goal of monitoring RSC.

The DL models we implement and compared are Inception, Inception-Resnet, Xception, DenseNet, MobileNetv2, and NASNet. All of these models have achieved remarkable results for image classification in well-known benchmarks. The SS models we evaluate are Ordinary Kriging (OK), Radial Basis Functions (RBF), and Inverse Distance Weigthed (IDW). The first provides a comprehensive understanding of the spatial autocorrelation for each particular variable, while the second and third allow a faster implementation. Our integrated approach works by combining the output feature vector from the DL model with the interpolated values from the SS model to output a more robust prediction of RSC for the locations of interest.

Acknowledgements

I would like to thank my advisor, Prof. Mark Crowley for his willingness to supervise my work as a graduate student and the continuous attention he brings to all aspects related to the success of the students in the Machine Learning Lab.

I also acknowledge the valuable collaboration of Prof. Liping Fu and Dr. Guangyuan Pan from the Innovative Transportation System Solutions (iTSS) Lab at the University of Waterloo. They provided access to the data used in this thesis along with important comments about the Monitoring of Road Surface Condition that certainly helped define the scope and outcome of this work.

Thanks also to Prof. Sebastian Fischmeister for attending my thesis seminar and for his insightful feedback.

Dedication

This thesis is dedicated to Katherine and Esperanza, no words can express my gratitude for their love and support.

Table of Contents

List of Tables

List of Figures

Abbreviations

CAD Computer-aided design 9

DL Deep Learning iii, 23–28

EC Environment Canada iii, ix, 12, 13, 15, 19, 22

GIS Geographic Information Systems 8, 9

IDW Inverse Distance Weigthed iii, ix, 7–9, 11, 20, 21

IoT Internet of things 27

ML Machine Learning iii

MTO Ministry of Transportation of Ontario iii, 12–14, 21–23

OK Ordinary Kriging iii, 7, 10, 11, 20, 21

RBF Radial Basis Functions iii, ix, 7, 9, 11, 20–22

RMSE Root Mean Square Error viii, 20, 21

RSC Road Surface Condition iii, 3, 23, 24

RWIS Road Weather Information System iii, 12, 13, 19, 21, 22

SS Spatial Statistics iii, 5–8, 11

Chapter 1

Introduction

1.1 Road Surface Monitoring and Road Maintenance Operations

Road maintenance during the winter is a safety critical operation that requires a significant amount of resources, as well as proper coordination and best practices for sustainability [55]. In countries located in Northern latitudes, such as Canada, The United States, and Finland, winter road maintenance is a priority for Government Offices at multiple levels. While these countries have made substantial efforts to ensure that the roads are suitable for the transportation of passengers and goods during the winter, achieving a trade off between safety and resources expended is a challenge they face every year.

Driving when there is snow or ice on the roads is not only more difficult but also more dangerous. The harsh weather conditions during the winter are recognized as key factors associated with a higher probability of collisions [11] [47] due to circumstances such as poor road surface friction [28]. Moreover, the risk of fatality also increases in accidents occurring in periods of snow falls or even after snow falls if the road surface has not been cleaned timely and thoroughly [51]. Transportation offices are aware of these numbers and plan accordingly, with specific measures to prioritize road monitoring and cleaning as part of their short-term plans and long-term road safety strategies.

Vision Zero is an innovative approach targeted to improve road safety to the point where zero fatalities happen in the roads [8]. It was created by the Swedish Government in 1997 and encompasses strategies where multiple stakeholders, such as Transportation Offices, Policy makers, car makers, and road users, just to name a few; work simultaneously toward

reducing the factors related to road accidents. International and national organizations across the world have been promoting the Vision Zero approach over the past few years, with significant and quantified benefits in cities of developed and developing countries [58]. The main differences between conventional road safety plans and this new approach is a strong focus on engaging all the stakeholders in the planning process and a data-driven foundation that leverages innovative technologies toward embedding safety on every stage of the planning and operation of road infrastructure [62].

Several cities in Canada have already launched their Vision Zero plans and are working actively on promoting citizen engagement and encouraging the use of state-of-the-art technologies to materialize the goal of zero fatalities on the road. For instance, the Vision Zero plan in Toronto identifies areas with high priority to improve road safety, divided into five road user segments: motorcyclists, cyclists, pedestrians, older adults, and school children. They also identify two dangerous practices: aggressive driving and driver distraction [14]. In terms of the promotion of data-driven technologies, the City of Toronto ran a public Challenge in 2018 where citizens, universities, and the general public were invited to design data-driven innovations to help materialize the Vision Zero plan. While public engagement and education are both contributing factors toward improving road safety, engineering innovations in the planning and operation of road infrastructure are also crucial for implementing road safety into transportation infrastructure. In particular, the use of Big Data and Artificial Intelligence technologies within transportation systems is becoming more and more frequent. Specific applications such as video analytics are instrumental in enhancing traffic operations, monitoring road conditions, and producing valuable insights to better understand risk factors [36].

Being one of the largest countries in the Northern hemisphere, Canada deals with winter road maintenance operations over thousands of kilometers of urban and rural roads every year. Even though it is a country with experience on how to keep its roads operative in the winter, there is still room for improvement in terms of road safety. For instance, in the year 2013, more highway fatalities due to snow or ice conditions were reported in the province of Ontario in Canada compared to the previous year [38]. On the positive side, also in Ontario the benefits of a timely winter road maintenance have been quantified [65], highlighting the importance of data-driven methods for improving operations.

The cost of winter road maintenance is significantly higher than that spent in other seasons of the year. For instance, according to the Ministry of Transportation in Ontario (MTO), about 50% of the total budget disbursed for highway maintenance corresponds to snow and ice cleaning during the winter [10]. In Montreal, the budget for snow and ice removal is approximately 150 million CAD annually [29], with similar figures in Toronto (90 million CAD) [4] and Ottawa (68 million CAD) [3].

In Canada, every province has autonomy in how they manage winter road maintenance operations. Depending on the strategy established by the Transportation offices at the provincial or municipal level, the maintenance is done directly by the Transportation offices or by private contractors. Moreover, the standards for maintaining road surface condition during the winter also differ across provinces according to their particular needs and environmental conditions [29]. However, in a recent report published by the Office of the Auditor General of Ontario, road safety concerns were raised due to the outsourcing of snow and ice removal operations to private contractors, including observations about the lack of a proper system to oversee snow removal operations and tools to monitor road surface cover during snow falls [38].

Multiple activities need to be scheduled and coordinated by the Transportation offices in order to keep roads with the minimal possible amount of snow and ice so the drivers can use the road network safely. For this purpose, the transportation offices commonly use RSC as a measure to identify the current state of the road regarding snow or ice coverage, serving also as a communication mechanism between stakeholders across the whole operation. More specifically, RSC introduces several categories to quantify the current amount of snow cover over a road. In Ontario, the Ministry of Transportation (MTO) uses three main categories to identify whether the road is bare, partly covered, or fully covered [20]. Examples of images for each category are shown in Figure 1.1.

Figure 1.1: Example images from a roadside camera near Otter lake in Ontario. Left: Bare pavement. Center: Partial snow cover. Right: Full snow cover.

In terms of logistics, the overall winter road maintenance operation commonly involves two major stages. First, a road monitoring stage continuously checks road surface coverage when snowfall is highly likely to occur, then, once the snowfall has stopped a cleaning stage begins and the necessary resources, such as plowing and salting, are assigned to promptly

remove snow or ice accordingly. The road monitoring stage commonly involves patrolling by trained inspectors as well as observation through roadside cameras and weather stations. Even though patrolling is the most accurate method to identify RSC, it is also costly, and its coverage is limited to only those roads traveled by the inspectors.

1.2 Machine Learning Processing of Roadside Camera Images

Real-time monitoring of road surface conditions using cameras is critical for the safety of drivers and road maintenance operations. A study conducted by the Iowa Department of Transportation, USA, confirmed that the use of roadside cameras for snow removal operations can reduce the number of road patrol expeditions by 33% and the overall cost of those expeditions by 14% [52]. Additionally, a study led by the Minnesota Department of Transportation, USA, found that dash and ceiling-mounted cameras on snowplows allow for streamlined operations and better communication of road conditions across Transportation offices, contractors, and the general public [23].

In that regard, advanced image classification methods for classifying road snow coverage by processing images from roadside cameras have been introduced in recent research [44] [46]. However, it is still a challenging task due to limitations such as image resolution, camera angle, and illumination. Two common approaches to improve the accuracy of image classification methods are: adding more input features to the model and increasing the number of samples in the training dataset.

Deep Learning (DL) methods are a subset of Machine Learning (ML) methods resulting from a combination of advanced algorithms and math [19]. The remarkable achievements of DL methods across multiple applications are also attributed to the recent availability of powerful computing hardware, mainly Graphic Processing Units (GPUs) to train DL models. One of the fields where DL methods have achieved remarkable results is in Computer Vision, showing state-of-the-art accuracy in tasks such as image classification, object detection, and semantic segmentation. In most of the major Computer Vision research conferences, such as CVPR, ICCV, and IGARSS, DL methods have consistently shown better performance across applications like Medical Imaging, Autonomous Vehicles, Remote Sensing, Face Recognition, among other fields.

Recently, a handful of studies have evaluated the use of DL to automate the classification of images with the purpose of determining RSC during the winter, with remarkable results over images from dash cameras [46] [35] [34] [43]. However, fewer studies have

tested DL methods for determining RSC from images collected by roadside cameras. One example is the work of Pan et. al. [45], in which they adapt DL models trained over the ImageNet dataset to the task of RSC classification using the fine-tuning technique, achieving an accuracy of more than 90%.

1.3 Spatial Statistics for Interpolation of Weather Data

Spatial Statistics SS is an area of statistics focused on the analysis of spatial data. While most of the variables we commonly find in SS examples are related to environmental studies, such as weather variables or air pollution, any application that uses georeferenced data can benefit from these methods. Other use cases of SS are common in retail [48], transportation [21], health [9], among other areas.

In the context of weather data, SS is mostly used for interpolation of variables in places located between weather stations. Since the installation and maintenance of weather stations require significant resources, Environment Agencies usually install a limited number of stations and use data collected in those those stations to estimate weather variables over the area of interest.

The definition of both the number and locations of weather stations over a region of interest requires expertise in SS as well as in-depth knowledge of the characteristics of the area. Multiple considerations such as mountainous terrains or proximity to water bodies can significantly increase variability in weather patterns; therefore, requiring the installation of more stations to reduce uncertainty.

The use of SS for the interpolation of weather data is a common practice in geographies across the world, with examples in Canada [26], Colombia [54], South Africa [15], and Switzerland [40], to name just a few.

Chapter 2

Selected Spatial Statistics Models For Interpolation

In section 1.3 we present an overview of SS concepts and main applications, in this section we describe in more detail the main underlying ideas as well as some of the commonly used methods for interpolation of weather variables.

All SS methods for interpolation implement the idea of spatial vicinity or neighborhood, in other words, the methods work under the assumption that sample values that are closer to our location of interest provide more accurate information than farther samples. Even though this idea comes to our minds intuitively, interpolation methods in SS provide formal approaches on how to put it into practice.

In order to run the interpolation procedure there must be a number of sample values observed across the area of interest, with a *reasonable* number of sample points available for estimation of the variable of interest at unseen locations. What is a *reasonable* number of points is a question dependent on the specific characteristics of the variable of interest as well as the particular region and accuracy requirements.

To better illustrate these ideas we use the graphical example shown in Figure 2.1. Let's imagine our variable of interest is the height above sea level, we have already measured this variable at multiple locations (red and black dots) and our goal is to infer the height in the yellow dot. The most trivial approach would be to calculate an average of all the sample points, including the red and black dots; however, we know that the height measured at the red dots is more similar to the one at the yellow point than that of the black points.

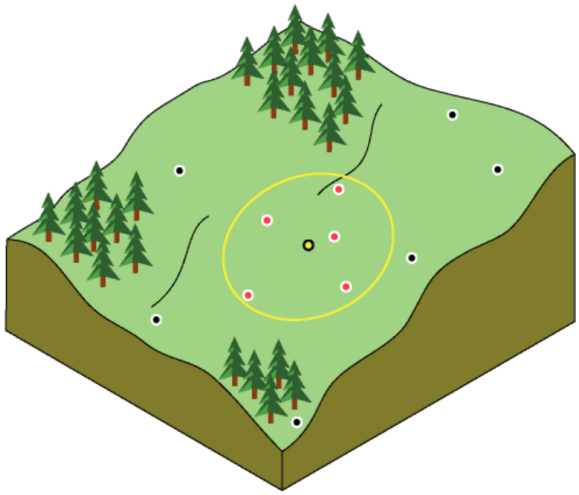


Figure 2.1: Spatial vicinity around a location of interest. *Image source: arcgis.com*

The estimation of the height at the yellow point is an example of the main goal of interpolation methods in SS. We know is better to give more relevance to the red points than the black points, but how could we use the values of the red dots to obtain the best estimate of the height in the yellow point? The approach we use is what differentiates the available interpolation methods, in other words, each interpolation method offers it own mathematical formulas and statistical assumptions to obtain the desired value at the yellow point, and those are the details we explore in the following sections.

More formally, the spatial variable we aim to interpolate is defined as v, and we interpolate the variable v in the location l_o as a function of the sample values v_i observed in the vicinity of l_o, as seen in Equation 2.1. Those sample values are in turn dependent of the location in which they where sampled, in other words $v_i = v(l_i)$ for $i = 1, 2, 3, ..., n$ where n is the number of observed samples.

$$v(l_o) = f(v_1, v_2, v_3, ..., v_n) \tag{2.1}$$

In the context of spatial interpolation, the literature describes two main categories of methods, deterministic and geostatistical methods [16]. IDW and RBF are examples of deterministic methods because they assign weights to sample points l_i in the vicinity of the target point l_o without making any special assumption about the spatial variable v. On the other hand, geostatistical methods such as OK treat the spatial variable as a stochastic variable and include a robust statistical analysis to better understand its behaviour.

2.1 Inverse Distance Weighted

The IDW interpolation method gives more relevance to sample values that are much closer to the location of interest by assigning a particular weight to each observation. We previously introduced the concept of spatial vicinity as the region that contains sample observations relevant for the interpolation of the variable v in the location l_o; however, in practice that vicinity is usually not defined as a boundary, instead, it is implemented by setting the k nearest number of sample observations to consider and by giving less and less relevance to sample points that are farther away from l_o.

The relevance we assign to an individual observation is measured by its weight, and in the IDW method this weight is calculated as the inverse of the power function. The two parameters for the IDW method are k number of nearest sample points to consider and the exponent p of the power function. Detailed descriptions of the IDW method can be found in [37] and [5].

Equation 2.2 shows how the weight of each sample observation is calculated based on the distance d_{io} between the sample point l_i and the target point l_o. It is defined only when $d_io > 0$ for all sample points.

$$w_i = \frac{\frac{1}{d_{io}^p}}{\sum_{i=1}^{k} \frac{1}{d_{io}^p}} \tag{2.2}$$

And Equation 2.3 makes use of the values from the k nearest samples as well as their corresponding weights w_i to interpolate the value of v at the target location l_o.

$$v(l_o) = \sum_{i=1}^{k} w_i v(l_i) \tag{2.3}$$

The values for the k nearest number of samples to consider as well as the exponent p of the power function are commonly determined by cross validation [42]. However, $p = 2$ and $k = 5$ are common values used as default for many software implementations of the method.

The IDW method only requires the definition of two parameters and the computational resources it demands are low; therefore, it is well suited for cases in which researchers do not have extensive information about the spatial variable of interest v or do not have access to high-end computers. The method is also known for its ease of use and is offered by most Geographic Information Systems (GIS) and SS software packages.

2.2 Radial Basis Functions

RBF are regarded as efficient interpolation methods that are more flexible than the IDW method but at the same time require the setting of more parameters; therefore, more decisions need to be make during implementation. Due to its deterministic nature, RBF methods do not allow for further analysis about aspects such as spatial correlation or interpolation errors.

RBF methods use splines as the underlying functions to interpolate the spatial variable of interest v. Splines are mathematical functions defined with piecewise polynomials that have proven to be useful for applications in Computer-aided design (CAD) and computer graphics. Splines can take many forms depending on the mathematical formulation and parameters, more details about splines can be found in [50].

As opposed to IDW, the values estimated by RBF methods can be higher than those observed in the sample locations l_i, as illustrated in Figure 2.2. For many cases this is a desired characteristic that makes RBF methods an attractive alternative for researchers.

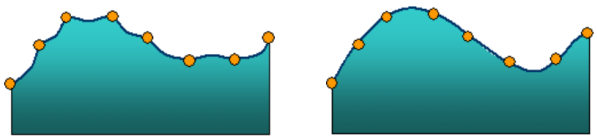


Figure 2.2: Profile of interpolation by IDW(left) and RBF(right) methods. *Image source: arcgis.com*

While multiple types of RBF have been described in the literature, modern GIS and CAD software packages offer ready to use implementions of those that offer reasonable tradeoffs between complexity and accuracy. Some of the most common ones are listed below.

- Completely regularized spline
- Multilog
- Multiquadric
- Natural cubic
- Spline with tension
- Thin plate spline

The advantages of the Completely regularized spline have been well studied in the literature [41] [24]; therefore, it is the one we selected for the interpolation of weather variables in Chapter 3.

2.3 Ordinary Kriging

OK belongs to the category of geostatistical methods along with other Kriging based interpolation techniques. It models the variable of interest v as a stochastic variable and uses robust statistical inference to make predictions at target locations l_o. The use of more complex statistical ideas in OK such as spatial autocorrelation and anisotropy allow for more flexibility and a better understanding of the spatial variable of study.

To use the OK method we start by calculating an experimental semivariogram to identify three characteristics that describe the spatial autocorrelation between sample observations located at multiple distances between each other. These three characteristics are the range, sill, and nugget. The range is the distance at which two observations seem to be not correlated anymore. The sill is the average squared difference between points that are beyond the range distance and the nugget describes the background noise that causes slight differences in v even for points that are very close to each other.

Once we have the experimental semivariogram, we use a mathematical function to model the three characteristics we mentioned before in order to assign suitable weights to sample points l_i in the vicinity of the target locations l_o. Common functions are Circular, Exponential, Gaussian, and Spherical. In practice, researchers select the function that better models the shape of the experimental semivariogram, Figure 2.3 shows an example of a Gaussian function along with the experimental semivariogram as averaged points.

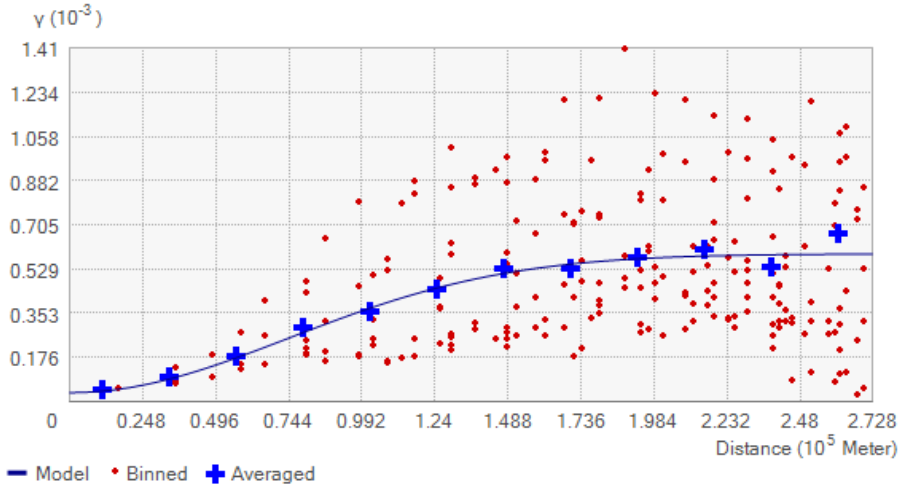


Figure 2.3: Example of an experimental semivariogram along with a Gaussian function. *Image source: arcgis.com*

The selected function provides an estimate of the spatial autocorrelation between points separated at certain distance, and is the basis to determine the weights that will be assigned to each one of the sample observations l_i in the vicinity of the target location l_o. The statistical formulation of the OK method is especially designed to minimize the variance of the estimates with respect to the mean value of v, in other words, it is an unbiased estimator.

While other variations of Kriging methods such as Simple Kriging, Universal Kriging, and Cokriging allow for more complex analysis, OK is a method that offers a good trafeoff between robust statistical inference and ease of use. OK is one of the most accurate interpolation methods for spatial variables; however, it requires the setting of multiple parameters and could be time consuming for cases where efficiency and ease of use are important.

2.4 Summary

SS offers a variety of methods to accomplish spatial interpolation in locations that were not sampled within an area of interest. In practice, deterministic methods such as IDW and RBF require the definition of few parameters and therefore are best suited for cases in which researchers do not have extensive knowledge of the spatial variable. On the other hand, geostatistical methods such as OK are appropriate for applications in which having an in-depth understanding of the spatial autocorrelation of the variable of interest is fundamental for the analysis.

All of the three methods described are used extensively in both research and industry applications, and the choice between them commonly involves the consideration of available data, expertise, and computational resources.

Chapter 3

Integration of Roadside Camera Images and Weather Data

3.1 Problem Formulation

In Ontario, the MTO monitors winter road surface conditions through the RWIS, a network of stations that include roadside cameras and specialized instruments to measure weather variables [64]. However, there are a limited number of RWIS stations across Ontario; therefore, the network has reduced spatial coverage. In order to enhance the range of the RWIS system, more roadside cameras and weather stations would be required. With that requirement in mind, the first objective of this study is to introduce a novel data integration approach that makes use of all the other MTO roadside cameras as well as EC weather stations to have both more images and more weather records for the purpose of road surface monitoring during the winter.

We highlight the potential of integrating additional datasets to improve the spatial coverage and accuracy for monitoring winter road surface conditions. Consequently, we examine the main characteristics of the input datasets, with a focus on the location and spatial configuration of the observing stations. Moreover, we also quantify the benefits of integrating the input datasets into a larger one that offers wider spatial coverage and more input features.

3.2 Area of Study and Data

The area of interest for this study is the Province of Ontario in Canada. More specifically, we calculate the benefits of the data integration approach over the three most densely inhabited Ecoregions in Ontario (Table 3.1). The datasets we consider include RWIS stations, MTO roadside cameras, and EC weather stations.

Table 3.1: The three most densely inhabited ecoregions in Southern Ontario, Statistics Canada 2016.

Ecoregion	Inhabitants/km	Rank across Canada
Lake Erie Lowland	344	2nd
St. Lawrence Lowlands	179	3rd
Manitoulin-Lake Simcoe	66	6th

3.2.1 RWIS Stations

The Road Weather Information System (RWIS) was introduced as a network of advanced stations to continuously monitor roads at certain locations, improving the coverage and complementing the conventional road patrolling [10]. In general, RWIS stations include roadside cameras along with weather sensors to measure variables such as air temperature, wind speed, pressure, and humidity. Depending on the vendor and the configuration, the stations can also have embedded pavement sensors to determine road surface temperature.

Previous research related to RWIS stations has focused primarily on quantifying their overall benefits for road winter maintenance, both in terms of safety and cost savings [33] [63], as well as best practices for determining their optimal locations over an area of interest [32] [57]. The evidence on previous studies confirms the relevance of RWIS stations for winter maintenance operations, with networks of stations installed across North America and Northern Europe. However, with dozens or even hundreds of stations collecting data 24 hours per day, it is difficult for the operators of those sensor networks to process all that data simultaneously and perform decision making in almost real time during snowfalls.

Automated processing of RWIS stations data is potentially beneficial for multiple stakeholders involved in the winter road maintenance operations. On one hand, private companies hired by Transportation offices monitor the roads using the RWIS network and send salt trucks and snowplows to maintain road surface standards when needed. On the other hand, Transportation offices also monitor the roads through the RWIS network to oversee the private contractors work and publish road status information to the general public.

3.2.2 MTO Roadside Cameras

The MTO manages a large number of roadside cameras for traffic monitoring purposes. In total, more than 400 roadside cameras are distributed across Ontario, with most of the cameras located beside major roads such as freeways and highways. These cameras are connected to an information system that helps monitor traffic conditions in real time, additionally, this system is useful for emergency services such as the Police or Fire Departments when quick response is necessary on the road.

The information system includes a Traffic Operation Centre where data coming from multiple sources, such as cameras and road patrols, is continuously processed to assess traffic conditions and determine prompt measures to improve traffic flow. The system was launched in 1991 and since then it has updated multiple times to accommodate new hardware and software technologies. Most of the images collected by the roadside cameras have a size of 640 by 480 pixels, a size also know as standard VGA. Figure 3.1 shows and example of an image captured by an MTO roadside camera beside Highway 401.

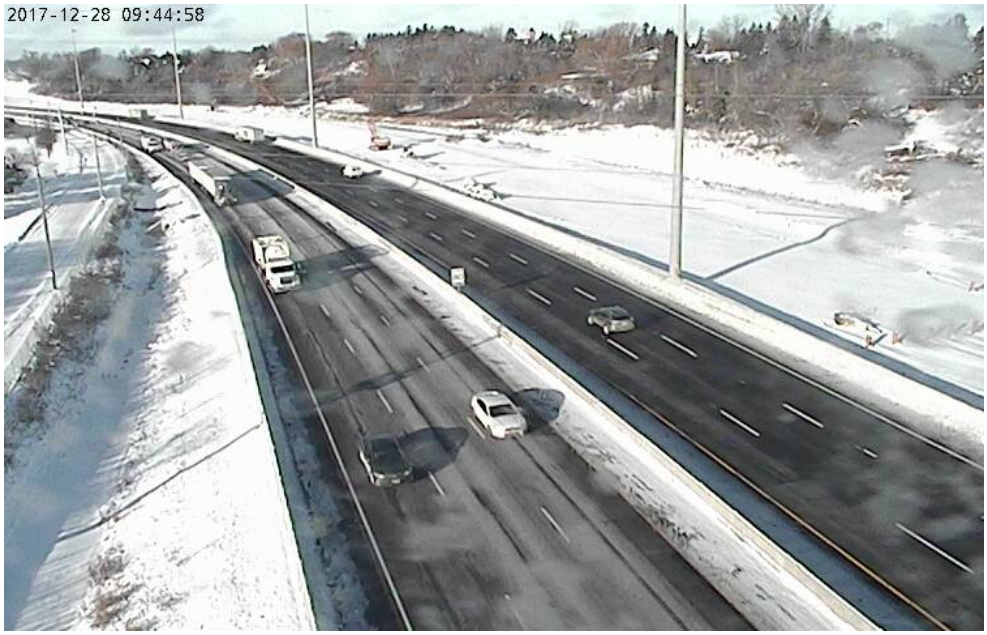


Figure 3.1: Example image captured by an MTO roadside camera looking at a section of the 401 Highway on December 28th, 2017.

3.2.3 Environment Canada Weather Stations

Environment Canada EC is a department of the Government of Canada and its goal is to oversee topics related to the preservation of the Natural Environment and the monitoring of Climate conditions. To perform real time monitoring of weather conditions across the country, EC collects and process data from a variety of sensors, including weather stations, as well as satellite measurements. In Ontario, EC uses a network of about 100 weather stations distributed across the province. Figure 3.2 shows a picture of an EC weather station installed at the University of Waterloo in Southern Ontario.

Figure 3.2: EC weather station located at the University of Waterloo. *Image source: uwaterloo.ca*

These weather stations collect a number of variables that help identify short term and long term climate conditions. Access to those variables is important for weather forecast, especially in the winter season, when snow storms can affect traffic flow significantly. Some of the most commonly recorded weather variables are listed below.

- Air Temperature
- Dew point
- Humidity
- Precipitation
- Pressure
- Wind speed

EC publishes records from the weather stations as open data and includes an historic catalog spanning multiple decades. The availability of this dataset is important for both short term and long term environmental studies, with applications for multiple industries and users across the country.

3.3 Data Integration Analysis

The RWIS system is the one currently used by the Ministry of Transportation of Ontario to monitor road winter surface conditions. Each RWIS station records images from roadside camera and weather variables using specialized measurement instruments. To improve the spatial coverage of the current system we propose to include all the other MTO roadside cameras as well. However, those MTO cameras do not record any weather variables. Therefore, we also propose the use of data from Environment Canada stations to interpolate weather variables for each one of the added MTO cameras. Figure 3.3 shows the spatial configuration of the observing stations in the three input datasets and locates the three most populous ecoregions in Ontario.

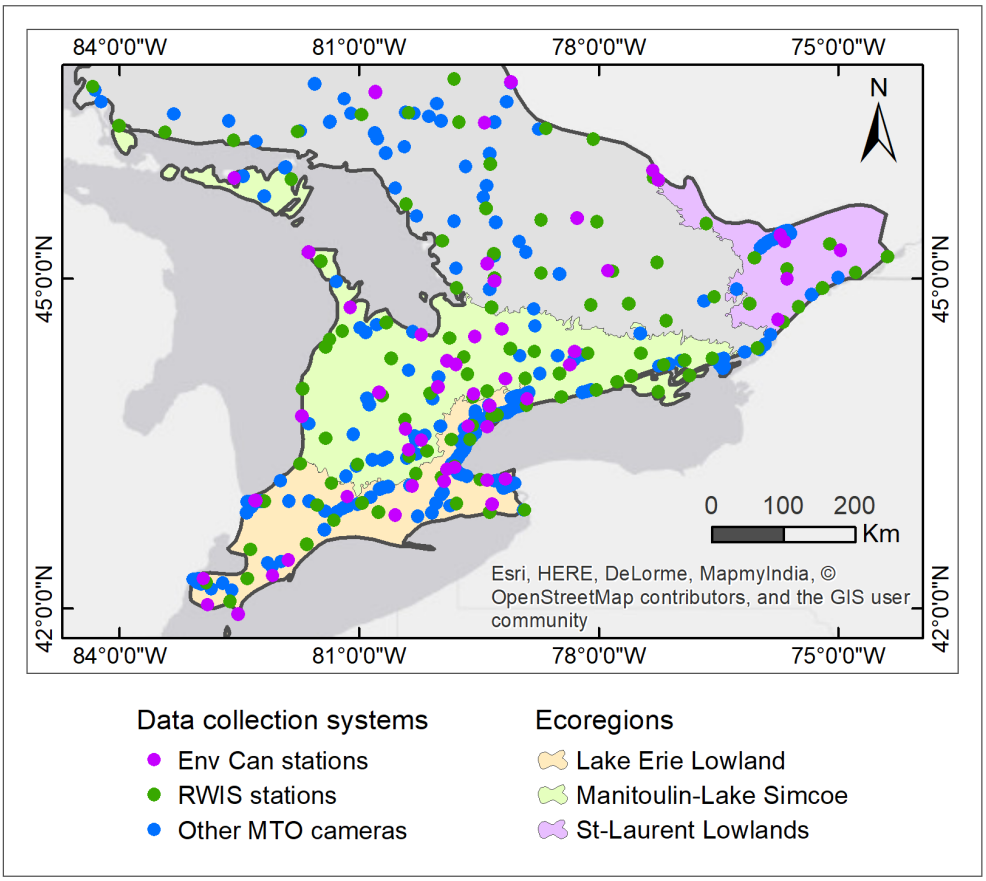


Figure 3.3: Location of observing stations from the three input systems.

The reason why ecoregions are important to the analysis is twofold: on the one hand having more observations within the same ecoregion facilitates the interpolation of weather variables since we could expect similar weather conditions, on the other hand, we focus our analysis in the three most populous ecoregions in Canada; therefore maximizing the benefits of the system integration for a higher number of road users.

Adding other MTO cameras to the RWIS system increases the total number of available roadside cameras in Ontario by more than four times and reduces the average distance to the nearest camera (Nearest neighbor NN) from around 38 km to less than 10 km. It also increases the number of cameras in the three most populous ecoregions in the province increases by more than six times. (Table 3.2)

Table 3.2: Adding other MTO roadside cameras to increase the number of images.

Type	**# of locations in Ontario**	**Avg. distance to NN (km)**	**# of locations selected ecoregions**
RWIS	139	38.4	68
Other MTO	439	7.2	364
RWIS + MTO	578	9.4	432

The size (in pixels) and format of the images (.jpg) collected by RWIS and other MTO cameras is generally the same. In addition, every camera in both systems takes at least one picture of the road every 15 minutes, which is enough for monitoring winter road surface in most cases. A visual inspection (Figure 3.4) of images taken by more than 40 RWIS stations and 30 other MTO cameras confirms that the point of view with respect to the road and the viewing angle of the roadside cameras in both systems are very similar; a fact that facilitates the combined processing of all the images.

Moreover, adding weather stations from Environment Canada to the RWIS system increases the number of weather stations in Ontario by 1.7x and reduces the average distance to the nearest weather station (Nearest neighbor NN) from around 38 km to less than 26 km. More importantly, the number of stations in the three most populous ecoregions in the province increases by 1.7x. (Table 3.3).

Having more weather stations allows us to interpolate weather variables for all other MTO roadside cameras by using observations from both the RWIS system and the Environment Canada weather network. Furthermore, in order to better understand the spatial configuration of all three input datasets and to investigate the performance of some selected spatial interpolation methods, we conduct a more comprehensive analysis supported by spatial statistics.

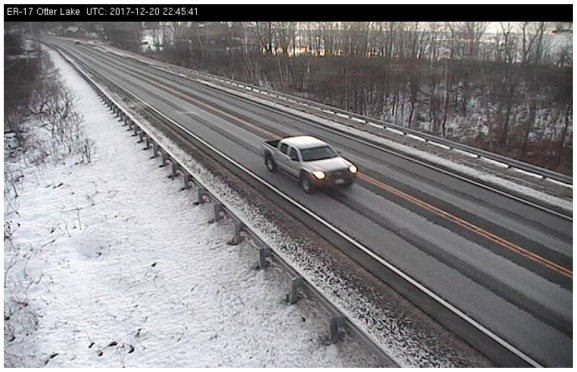


Figure 3.4: On the left, an image from an RWIS station. On the right an image from an MTO camera.

Table 3.3: Adding Environment Canada stations to interpolate weather data.

Type	**# of locations in Ontario**	**Avg. distance to NN (km)**	**# of locations selected ecoregions**
RWIS	139	38.4	68
Env. Canada	99	35.8	45
RWIS + Env. Can	238	25.7	113

In the following step of our analysis, we study the spatial clustering of observing locations in the three systems: RWIS, other MTO cameras, and Environment Canada stations. In other words, we evaluate how clustered the stations are in those three systems under the assumption that weather stations distributed at approximately random locations generally produce better weather interpolation compared to settings where stations are clustered [6].

For this purpose, we select the L-Function [18] to identify whether the spatial distribution of stations in each system corresponds to a completely random spatial point pattern or in contrary if the distribution corresponds to a clustered point pattern. The L-Function works by comparing the spatial distribution of the points in the reference sets (RWIS, MTO, EnvCan) against a set of points generated at random locations by sampling coordinates as uniformly distributed random variables within the area of study. What makes the L-Function advantageous for our analysis is the fact that it repeats the described process for multiple scales; therefore, it outputs a function that allows us to infer the degree of spatial randomness of every input dataset at different distance bands.

Figure 3.5 shows how clustered or randomly distributed are each one of the three datasets when seen across multiple spatial scales, in other words, when compared against

multiple sets of randomly generated points each one having an increasing average distance between the random points. When the red line (reference set) is above the blue line (random sets of points), as we see for the RWIS dataset (a) and the other MTO cameras dataset (b) we infer that both datasets are highly clustered, which makes sense considering those cameras are all installed beside major roads. For Environment Canada stations (c), the red and blue lines are close to each other, which means those stations are distributed randomly across Ontario. In all three cases, we also plot a confidence interval obtained by creating nine permutations of the random points.

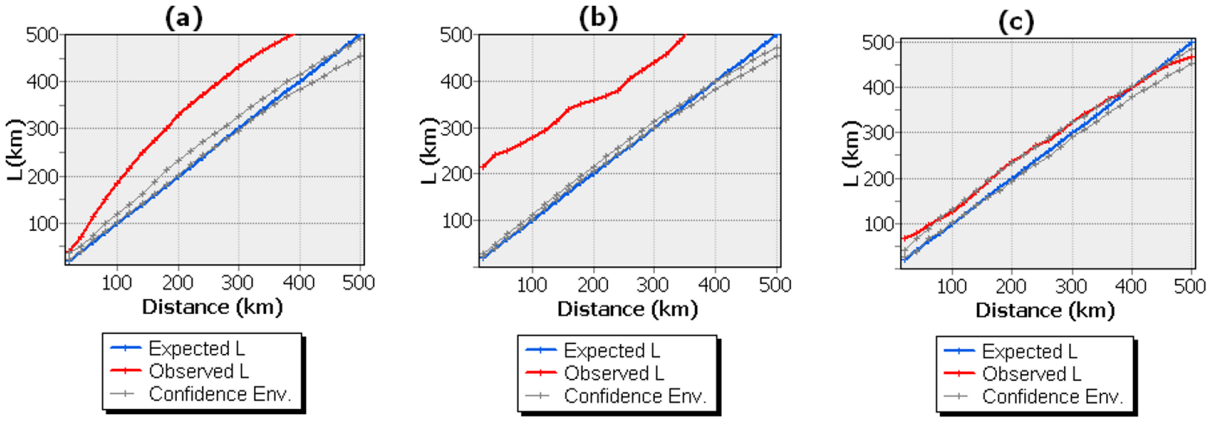


Figure 3.5: Multi-distance spatial cluster (L-Function) plots for: (a) RWIS stations, (b) other MTO cameras, and (c) Environment Canada stations.

Based on previous literature related to the optimal location of weather stations [6] [27] [33] [57], we infer that the random spatial distribution of Environment Canada stations improves the interpolation of weather variables across Ontario compared to the clustered locations of the RWIS stations. Therefore, when combined with RWIS stations we can increase the spatial coverage and have a more distributed set of input stations to interpolate weather variables for MTO cameras.

3.4 Results

We compare the performance of three spatial statistics methods to interpolate three key weather variables, namely air temperature, wind speed, and air pressure. For this purpose, we selected a sample of 40 RWIS stations and 40 EC stations, for those 80 stations we

obtain data values for a no-snow day (November 7th, 2017), and for a snowy day (December 25th, 2017).

Before comparing the interpolation methods, we see the summary statistics for the three weather variables of interest on a no snow day in Table 3.4 and on a snowy day in Table 3.5.

Table 3.4: Summary statistics of three weather variables for a no-snow day.

Summary statistics	**Air temp. °C**	**Wind speed km/h**	**Pressure kPa**
Mean	-1.921	4.912	99.950
Std. dev.	5.195	6.419	2.809
CV%	—	131%	3%

Table 3.5: Summary statistics of three weather variables for a snowy day.

Summary statistics	**Air temp. °C**	**Wind speed km/h**	**Pressure kPa**
Mean	-12.186	13.587	98.518
Std. dev.	9.509	11.128	2.782
CV%	—	82%	3%

We see that the mean air temperature decreases by approximately 10 °C and the mean wind speed triples from the no-snow to the snowy day; in contrary, air pressure only shows a slight change of less than 2%. Therefore, we infer a better correlation between air temperature and wind with road snow coverage.

The three methods we compare are Inverse Distance Weighted IDW, Radial Basis Function RBF, and Ordinary Kriging OK. Tables 3.6 and 3.7 summarize the performance of the three interpolation methods on a no-snow and snowy days accordingly.

In general, the RMSE tends to be higher for the snowy day, likely due to high spatial variability not represented by the conventional standard deviation statistic in Table 3.5. OK scores the lowest RMSE in four of the six day-variable pairs; however, it also requires significantly more time to set up due to a greater number of parameters and model design decisions. Consequently, we suggest RBF as the interpolation method that offers the best tradeoff between complexity and accuracy.

In all three interpolation methods, we set up three and six, as the minimum and the maximum number of neighbors to use as input for the calculations. All parameter values are optimized through cross-validation, except for OK where we set up the parameters based

Table 3.6: RMSE of three interpolation methods applied on a no-snow day.

Interpolation method	Air temp. °C	Wind speed km/h	Pressure kPa
IDW	2.054	6.073	3.094
RBF	1.971	6.156	3.001
OK	**1.868**	**5.660**	**2.992**

Table 3.7: RMSE of three interpolation methods applied on a snowy day.

Interpolation method	Air temp. °C	Wind speed km/h	Pressure kPa
IDW	4.139	8.761	3.053
RBF	**3.898**	8.718	**2.963**
OK	3.921	**8.654**	2.999

on findings from a study targeted to determine the optimal location of RWIS stations across multiple regions in North America [2].

Specific parameter configurations for each of the three interpolation methods are:

- IDW: Optimized power parameter
- RBF: Completely regularized spline with optimized kernel parameter.
- OK: First order trend removal with the exponential kernel. For the experimental variogram, we set up a lag size of 10 km with 20 lags. The selected semivariogram is always Gaussian with a fixed range of 100km.

By adding all other MTO cameras as image data sources, the total number of cameras in the combined dataset increases from 139 to 578 across Ontario and the average distance to the nearest camera decreases from 38.4km to 9.4km. Additionally, six times more cameras are available in the three most populated ecoregions in Ontario. Moreover, the experimental evaluation of three spatial interpolation methods for inferring weather variables in unobserved locations shows that the best tradeoff between complexity and accuracy is offered by RBF. Overall, we introduced a novel data integration approach to improve the spatial coverage of winter road surface monitoring stations and provide experimental evidence of the benefits, especially for the most densely populated areas in Southern Ontario.

Generally, RWIS stations also include pavement-embedded sensors to better determine the conditions of the road, especially the sub-surface temperature. Even though our analysis does not consider these sensors, the benefits of our suggested data integration approach

are still significant considering that our goal is to facilitate the maintenance operations by providing a wider coverage for weather and camera data only. Further research can look at optimal ways to integrate data from pavement-embedded sensors as a complementary data source.

From the perspective of government transportation offices, our approach can provide actionable insights which can be used to more selectively perform manual patrolling to better identify road surface conditions. From a broader perspective, integrating these three datasets is feasible and can benefit the design and development of automated image classification methods for monitoring road snow coverage. Which in turn can help materialize the Vision Zero by improving the road maintenance operations and reducing the number of incidents due to poor road conditions during the winter.

3.5 Summary

To summarize, as part of this data integration study we first quantify the benefits of extending the RWIS system with images from other MTO cameras and weather observations from EC. Then we use spatial statistics to better understand the spatial configuration of stations in the three systems and find that Environment Canada stations provide a better spatial coverage than existing RWIS stations. Furthermore, we evaluate three different weather interpolation methods under snowy and no-snow weather conditions and suggest RBF as the one that better balances accuracy and ease of implementation.

Our initial results are promising and demonstrate that additional image and weather datasets can be incorporated to road monitoring systems, leading to measurable improvements in road monitoring tasks.

Chapter 4

Selected Deep Learning Models for Image Classification

DL methods have achieved remarkable accuracy for tackling multiple Computer Vision tasks in both research and industry settings, with applications ranging from automated diagnosis in medical imaging to quick detection of cyclists and pedestrians in autonomous vehicles, just to name a few. One of the factors that have influenced the rise of DL is the current availability of large datasets, such as ImageNet [17] with millions of images labeled according to predefined classes, which are essential for training DL algorithms. However, having such large datasets is not common for industry-specific applications; therefore, researchers have developed a practice called fine-tuning in which DL models trained over large and generic image datasets are adapted for classification of images in other domains.

DL models are an evolution of a simpler type of model called the Multi-Layer Perceptron, which takes an input vector and uses consecutive groups of non-linear functions called layers to produce higher-level representations of the input data. In contrast, DL models use many layers, each one taking as input the output of the previous one. Most DL models fall into the category of supervised learning because their goal is to make the model learn a mapping from input observations into the desired output, in our particular case we expect the DL models to receive images from roadside cameras as input and produce as output the correspondent RSC category as defined by MTO.

Figure 4.1 shows a simplified description of the architecture of DL models for image classification. An input image is represented as a three-dimensional matrix, with the height, width, and Red Green Blue (RGB) channels of the image as its initial dimensions. This input image moves through the layers in the model and in the end, the model outputs a

vector of probabilities, with the highest probability corresponding to the most likely RSC category for the image.

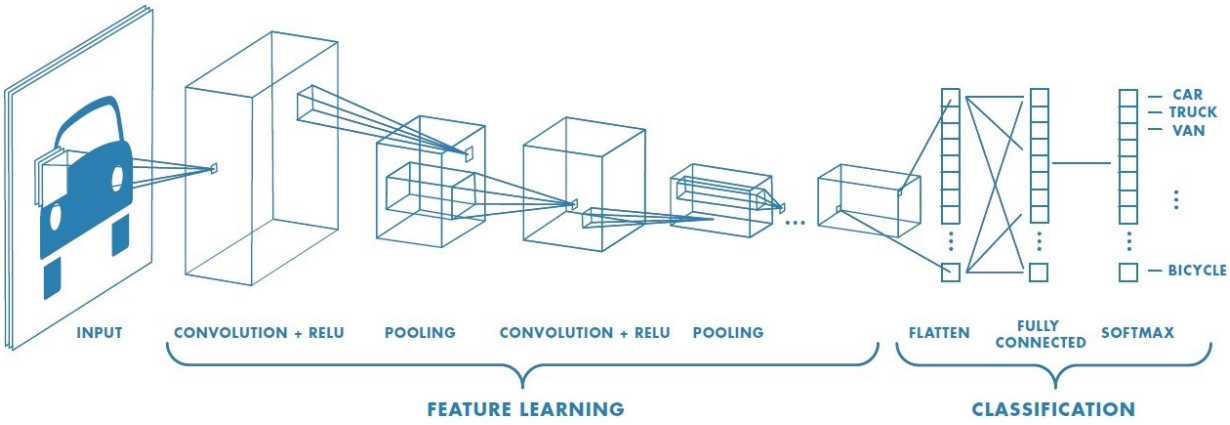


Figure 4.1: The generic architecture of a DL model for image classification. Image source: mathworks.com

The first part of DL models for image classification (feature learning) includes a series of convolutional and pooling layers that gradually convert the information contained on the input image into a compressed vector representation with smaller height and width but a larger number of channels than the original image. Convolutional layers are also known as filters that summarize an input matrix into one with smaller dimensions. In other words, these layers learn how to detect features such as shapes, contrast patterns, and color variations, and pass along a simplified representation of this features to the following layers.

The second part (classification) starts with a layer (flatten) that converts the three-dimensional vector from the first section into a one dimensional vector that is then feed into a few fully connected layers that reduce even more the size of the vector representation, and finally a softmax layer that outputs probability values for each category in the classification task. Fully connected layers are usually implemented in the last part of the models to summarize the visual features detected by previous layers. Most DL models have many convolutional layers but only a few fully connected layers because the number of parameters in the latter ones is much higher, which in turn requires more computational resources to train the models.

The design of a DL model also includes the consideration of multiple hyperparameters, and their role is crucial for the successful implementation of the model. The designer can set hyperparameters to manage multiple aspects of the model and fine-tuning them

commonly requires significant effort and domain knowledge of the application. By setting those parameters the designer can define the number of neurons per layer, the type of non-linear functions to use (activation functions), the number of times the dataset is passed through the model during training (epochs), how fast the model learns (learning rate), the Dropout regularization rate, among other characteristics of the model definition and training. In practice, researchers focus on fine-tuning the most relevant hyperparameters based on previous literature and their experimental findings.

All of the models we describe in the following sections show similar architectures to the generic architecture explained previously, with many convolutional layers arranged in sequence in the feature learning part of the model and a few fully connected layers in the end to output a vector indicative of the specific image class. Therefore, we focus on describing the particular characteristics that make each model different from the rest. Usually those characteristics include a combination of multiple ideas that span architectural settings, regularization techniques, and data augmentation.

For many applications these models are not trained from scratch, instead, researchers use a technique called finetuning for achieving maximum efficiency. Finetuning is done by training the classification models over large datasets that contain images of common objects such as animals, vehicles, food, etc. Then the last part of those models is trained again but now on datasets that contain application specific images, such as medical or satellite images. Using this technique the models can be adapted to new applications without the need for extensive training.

4.1 Inception

The Inception-v3 [61] model builds on the success of previous Inception models, which in turn originated from the GoogLeNet [60] model. It is a DL model with 42 layers that introduces the idea of Factorizing Convolutions. The goal of Factorizing Convolutions is to replace one convolution layer with two or more convolutional layers that have smaller window size. The total number of parameters in the resulting convolutional layers is smaller than that of the original convolutional with larger window size.

Figure 4.2 shows two convolutional layers of window 3 by 3 as a result of applying Convolutional Factorization to an original layer with a window size of 5 by 5. By applying his technique the authors reduce the number of parameters by approximately 28%, which in turn makes the model much more efficient due to the lower number of required computations in both the training and test stages.

The Inception-v3 model also makes use of auxiliary classifiers, a technique that mitigates vanishing gradient, an issue that arises during training of DL models due to the reduction of the gradient values in the Backpropagation algorithm. It works by allowing the model to determine the loss function not only from the last layer but also from one or more of the intermediate layers.

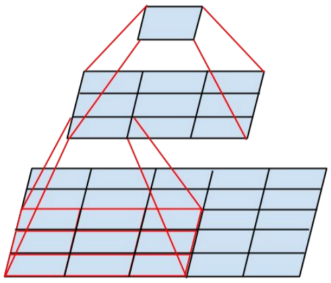


Figure 4.2: Effect of Factorizing convolutions. Image source: [61]

4.2 Inception-ResNet

The Inception-ResNet [59] model is the result of the combination of ideas from the Inception-v3 [61] and ResNet [22] DL models. This model includes the characteristics we describe in Section 4.1 and also uses residual connections between layers that are not adjacent to each other. The use of residual connections reduces the required time during training and also improves slightly the classification accuracy of the model.

The authors of the Inception-ResNet model also investigate the effects of multiple variations to combine the two previous models. Most of the variations correspond to alternate links between intermediate layers in the form of residual connections.

4.3 Xception

The Xception architecture [13] builds on top of the previous Inception type modules that use simultaneous convolutional layers with different filter sizes and then concatenates the resulting volumes into one output. According to the authors, training this model required them more than 60 GPUs; therefore, it is a model that most users might not train from

scratch. Instead, it is highly likely that further uses of the model in production correspond to instances of finetuning the model to a particular image dataset.

The Xception model implements the Pointwise Convolution technique followed by Depthwise Separable Convolutions. This combination accelerates model training while keeping the number of parameters considerably low compared to other DL models with comparable number of layers.

4.4 DenseNet

The DenseNet [25] model introduces the idea of connecting all previous layers to the current layer, this is done within groups of layers named blocks, which in turn are followed by additional blocks. While the idea of connecting layers that are not adjacent was previously implemented in ResNets [22], DenseNet models take a step forward and connect all the layers within a module as illustrated in Figure 4.3.

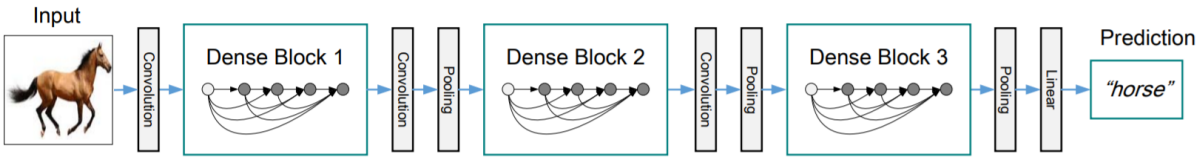


Figure 4.3: Illustration of the DenseNet architecture. Image source: [25]

4.5 MobileNet

The MobileNet [53] model also implements Depthwise Separable Convolutions, the same idea behind Xception modules. This approach reduces the computational cost and therefore is appropriate for applications running on small devices such as mobile phones and Internet of things (IoT) sensors. In MobileNet, the Depthwise Separable Convolutions capture the data structure across the height and width dimensions as well as the channels, but using smaller window sizes of maximum 3 by 3. The authors test the model in several image classification benchmarks obtaining state of the art results in accuracy while also having a reduced number of parameters.

4.6 NASNet

Most DL models for image classification use groups of layers organized in sequences to form complete models. Multiple terms are used in the literature to describe these groups of layers, such as blocks, modules, or cells. The key idea behind the NASNet [68] models is the search for optimal architectures of cells in small image datasets, and then the application of those cells to form more complex models for classification of larger datasets.

More specifically, the authors look for a convolution cell in the CIFAR-10 dataset, which includes only 10 image classes, and then use the same cell for classification in the ImageNet dataset, which includes one thousand classes. The paper also introduces a regularization technique called ScheduledDropPath to mitigate overfitting.

4.7 Summary

All of the six DL models described in this chapter achieve remarkable performance in the ImageNet [17] image classification benchmark. However, the use of DL models for more specific applications in image classification must consider certain trade offs to balance the available computational resources with the particular accuracy requirements. Recent studies have compared the performance of multiple models considering aspects such as accuracy, number of parameters, and number of operations [12].

Figure 4.4 shows the comparison of multiple models considering the number of operations per model (in millions) in the x-axis and the classification accuracy in the y-axis, additionally, the size of the circles represent the total number of parameters per model. We see that models such as Inception-v4 and ResNet-152 achieve the highest accuracy at the cost of an increased number of parameters, which in turn demands more computing resources. On the other hand, models such as GoogLeNet are lightweight and have fewer parameters, making them ideal for running on small devices such as road side cameras or smart phones. Overall, we could say that Inception-v3 is a model that shows a good balance between accuracy and efficiency when compared against similar models for image classification.

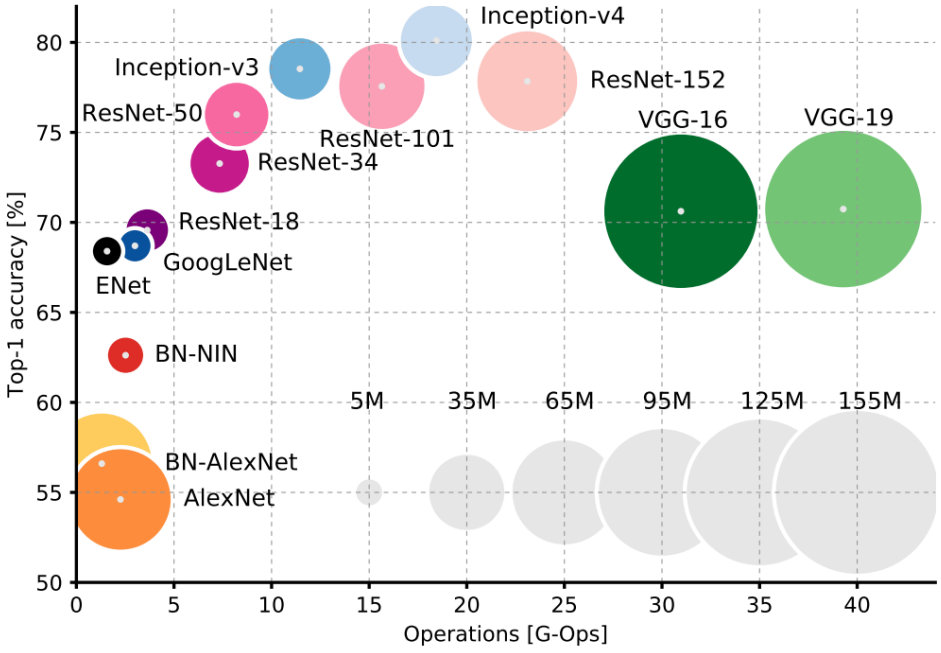


Figure 4.4: Comparison of models for image classification. Image source: [12]

Chapter 5

Deep Learning and Spatial Statistics for Winter Road Surface Monitoring

5.1 Problem Formulation

The objectives of this study are twofold. First, we aim to extend previous work regarding the use of DL methods for automated classification of RSC using images from roadside cameras with a particular focus on evaluating state-of-the-art image classification methods listed in Chapter 5. Secondly, we want to determine experimentally if the use of weather variables as a complementary data source improves the classification accuracy.

This paper is divided into three major sections. The first section describes the area of study as well as the datasets we use. In the second section, we evaluate multiple DL models for classification of RSC and in the last part, we include weather data as an additional source of data and assess the improvement on classification accuracy. To the best of our knowledge, this is the first study that looks at the combined use of roadside camera images and weather data for the automated determination of RSC. Additional contributions of this study include Python source code to replicate all the experiments presented in the paper.

5.2 Experimental Setup and Data

The area of study comprises the center and south part of the province of Ontario in Canada, which hosts approximately 38% of the population in the North American country. The

road network in the province extends along more than 275.000 km including all categories of roads, such as highways and municipal roads [1]. However, most of the population and therefore the roads are in the southern area. Weather conditions can vary drastically across the province, due mainly to the large range of latitude it covers, from approximately 42° N to 56° N, and the weather effects caused by the Great Lakes in its south-west side. For this reason, about 140 RWIS stations have been installed by the Ministry of Transportation in Ontario to help monitor the road conditions and weather in carefully selected locations across the province.

For this study, we selected a sample of 40 RWIS stations shown in Figure 5.1, which collected data approximately every 15 minutes for the winter of 2017-2018. The data includes 14.000 images taken by the roadside cameras as well as 70.000 observations from five weather variables measured by instruments installed on every station. Every image was labeled according to one of the three RSC categories listed by the Ministry of Transportation in Ontario, with approximately 45% of images corresponding to bare pavement, 40% to partial snow cover, and 15% to full snow cover. The recorded weather variables are Air Temp (C), Relative Humidity (%), Pressure (kPa), Wind Speed (km/h), and Dew Point (C).

5.3 Comparison of Models for Image Classification

We select six DL models that have scored state-of-the-art accuracy over the ImageNet image classification benchmark and evaluate how they perform for the classification of RSC over images from roadside cameras. All six models have been previously trained using the ImageNet dataset, which includes images of everyday objects such as animals, artifacts, and plants, among other classes. The goal is to fine-tune only a small portion, generally the latest portion, of those models in order to adapt them into the RSC image classification task. The underlying hypothesis is that the initial part of those DL models trained on large datasets has learned to identify basic patterns that are useful for classifying objects across multiple domains.

The DL image classification models we choose are Inception-v3 [61], Inception-Resnet-v2 [59], Xception [13], DenseNet169 [25], MobileNetv2 [53], and NASNet [68]. For all these models we kept the configuration of layers, training, hyperparameters, and regularization techniques as published in the original papers. When running the fine-tuning process, we consider previous settings from the work by Pan et. al. [45] who also explored the use of DL models for RSC classification, as well as other literature for fine-tuning image classification models [30] [49] [7].

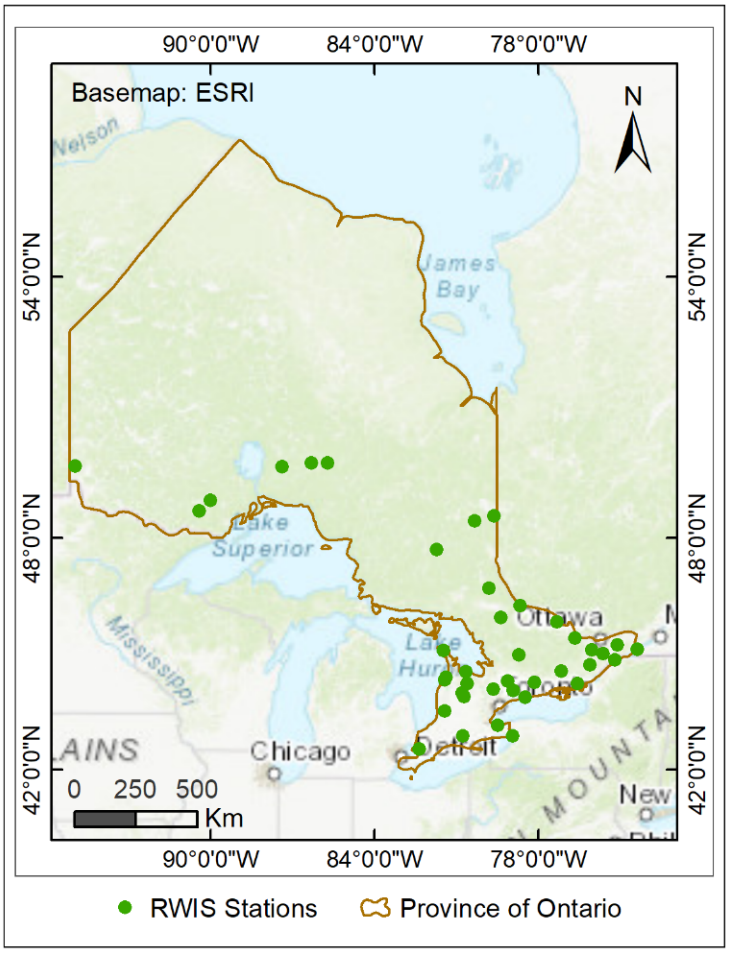


Figure 5.1: Location of the sample 40 RWIS stations in the province of Ontario.

In addition, we design a rather simple model inspired by seminal works in DL such as AlexNet [31] and VGG [56]. This baseline model has approximately 3.7% of the number of layers and 4.2% of the number of parameters when compared with the average characteristics of the other six models. Table 5.1 presents a summary of the characteristics of all seven models considered. Layers used for data reduction (max-pooling), model regularization (dropout), and vector concatenation are also counted as part of the total number of layers.

All seven models are trained and validated using 80% of the total number of images (14.000). The remaining 20% is held as a test set and we only use it for reporting accuracy

Table 5.1: Main characteristics of the DL models selected for comparison.

Model	Feature learning		Classification		Complete model	
	Params.	Layers	Params.	Layers	Params.	Layers
Baseline	392,608	10	603,411	7	996,019	17
Inception-v3	21,802,784	311	6,292,755	7	28,095,539	318
Inception-Resnet-v2	54,336,736	780	4,719,891	7	59,056,627	787
Xception	20,861,480	132	9,831,699	7	30,693,179	139
DenseNet169	12,642,880	595	3,915,027	7	16,557,907	602
MobileNetv2	2,257,984	155	62,739	7	2,320,723	162
NASNetMobile	4,269,716	769	51,987	7	4,321,703	776

in the end. Within the 80% of training data, 20% is taken for validation and plotting of the accuracy and loss functions during the training stage. We use the Backpropagation algorithm and Stochastic Gradient Descent (SGD) as optimizer to minimize the classification error (loss). Learning rate is kept constant at 0.001. Other parameters for the SGD optimizer are set according to recommended values in the literature, such as momentum = 0.9 and Nesterov momentum enabled. The number of times (epochs) the training set is passed through each DL model to iteratively reduce the misclassification error is kept constant at 50 and the images are feed in groups of 32 (batch size) to reduce memory usage.

Since the models have been previously trained using the ImageNet dataset, we only need to fine-tune or retrain the last portion of them in order to adapt them to the RSC classification task, except for the baseline model that needs to be trained from scratch. We consider three fine-tuning scenarios in which different percentages of the models are retrained, starting with only the fully connected layers, then fine-tuning only the last 5% and 15% of the models, including some of the last convolutional layers. During the training, a validation set is kept apart to continuously evaluate how the model performs over unseen images.

5.4 Results

Figures 5.2, 5.3, and 5.4 summarize the training and validation accuracy scored by the models when fine-tuning the fully connected layers and the last 5% and 15% percent of the models, respectively, except for the Baseline model in which all the layers are trained from scratch. In other words, in all three figures we see the exact same values for the Baseline

model. To our surprise, the baseline model achieves a competitive accuracy while most of the six state-of-the-art models do not reach an 80% accuracy when finetuning the last 5% of the model and show signs of overfitting when finetuning the last 10% and 15% of the model. The six state-of-the-art models learned successfully how to classify images in the training set but are unable to correctly classify images not seen during the training stage.

Figure 5.2 shows the final training and validation accuracy for all seven models when only the fully connected layers were fine tuned to our dataset. At first glance, we see that only three models, Baseline, Inception-Resnet, and Xception, achieve a training accuracy over 75%; however, the validation accuracy in Inception-Resnet and Xception is about 25% below their training accuracy. In other words, those two models already show early signs of overfitting. The rest of the models still require the training of more layers to learn useful mappings from the input images to the output categories.

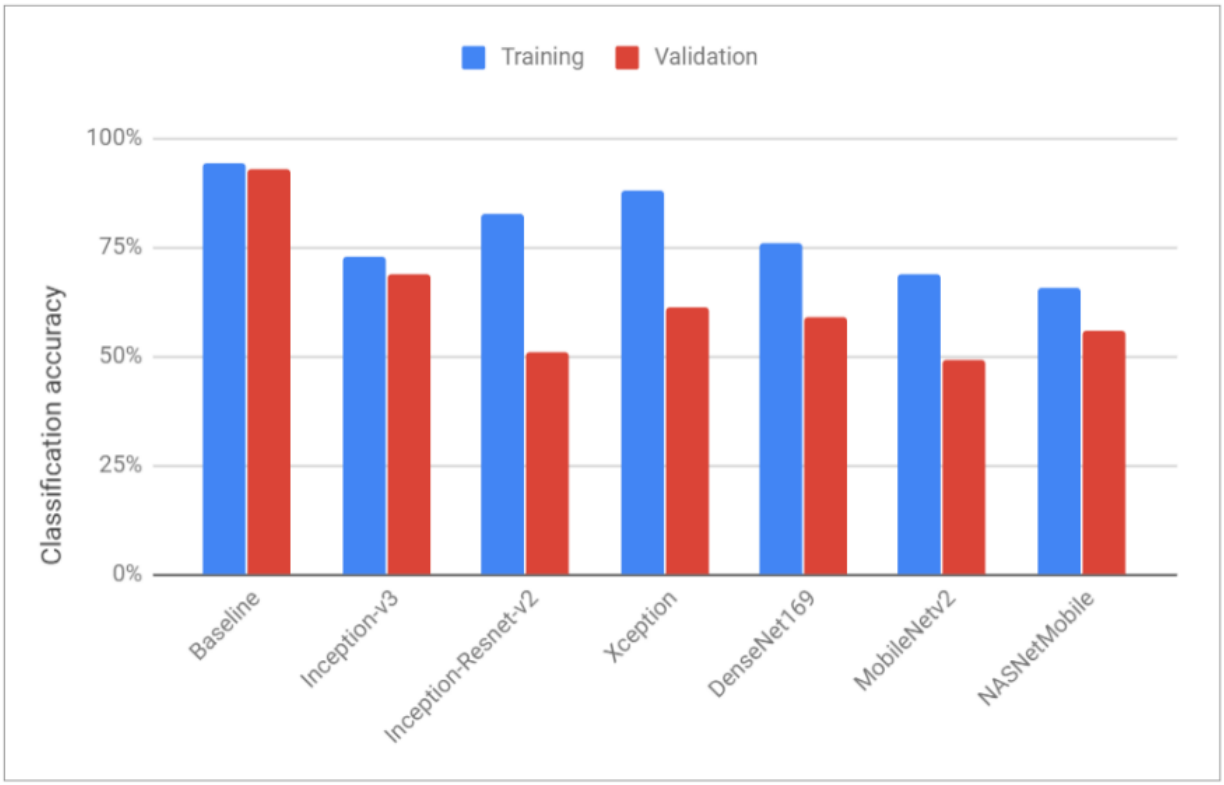


Figure 5.2: Finetuning only the fully connected layers.

The results shown in Figure 5.3 represent the training and validation accuracy when only the last 5% of layers in the models are finetuned to our dataset. All models achieve a training accuracy of more than 75%; however, the values for validation accuracy are over 75% only in the Baseline, Inception, and Inception-ResNet models. Even though we finetune a small percentage of the models (5%), some already show significant overfitting, such as Xception and MobileNet.

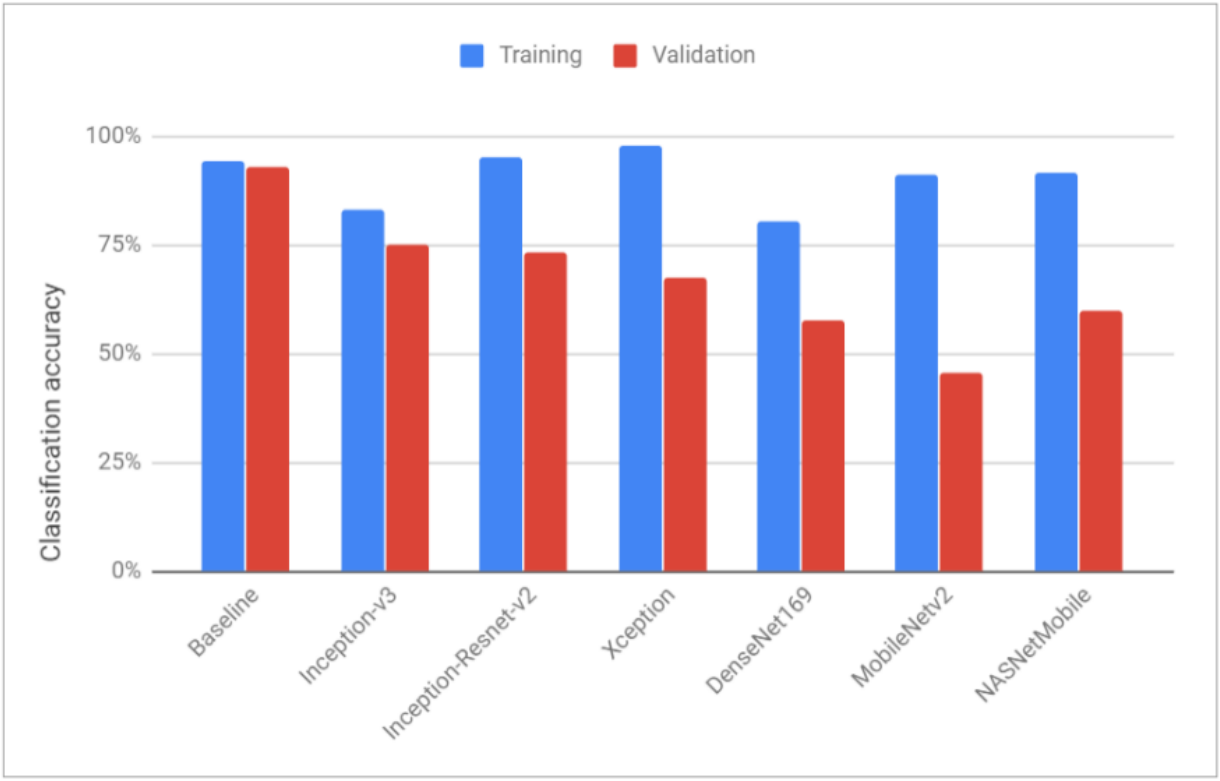


Figure 5.3: Finetuning the last 5% of layers.

Figure 5.4 shows the final training and validation accuracy for all the models when the last 15% of layers are finetuned to our dataset. Almost all models achieve a training accuracy above 90%, with DenseNet as the only exception. However, only two of the seven models are able to reach a validation accuracy equal or greater than 75%, the Baseline and Inception models. The other five models show clear evidence of overfitting, with extreme cases like MobileNet and NASNetMobile.

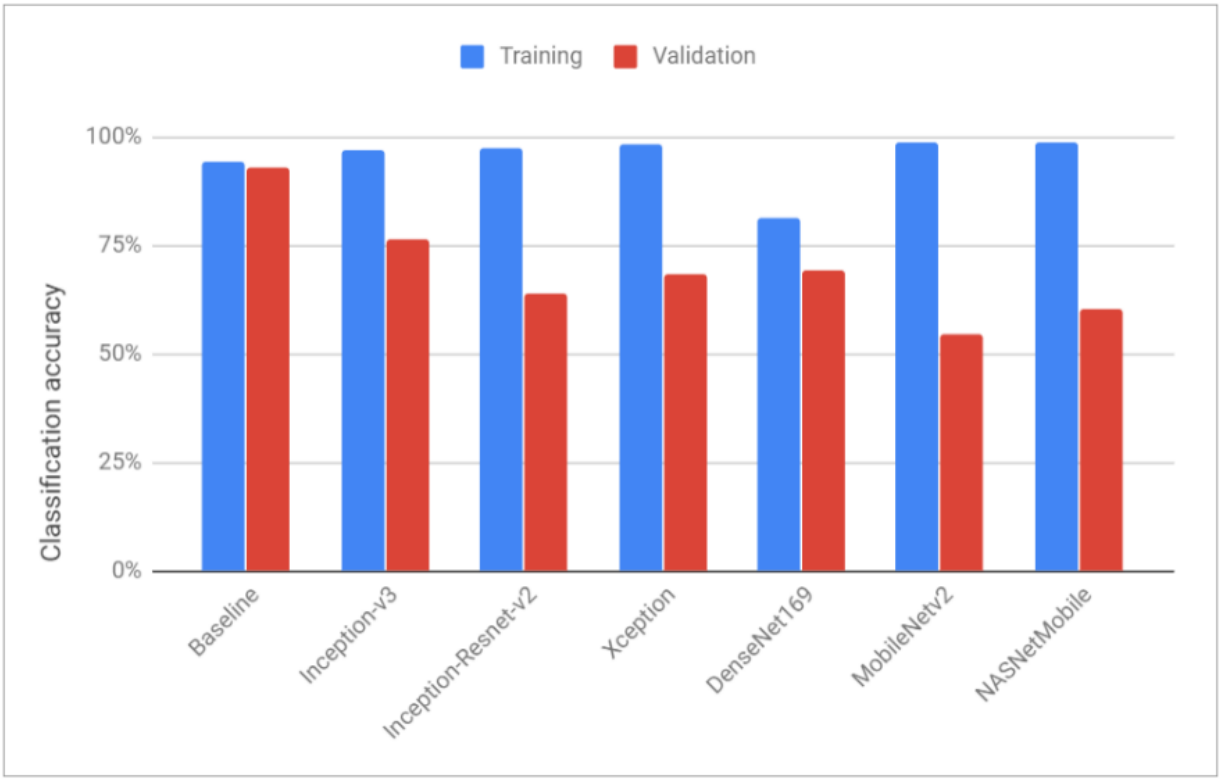


Figure 5.4: Finetuning the last 15% of layers.

Increasing the number of fine-tuned layers for the RSC classification task produces higher accuracy in the training set; however, it also affects considerably the accuracy on the validation set. In other words, the more layers we fine-tune the stronger the overfitting effect, with extreme cases such as the MobileNetv2 model, which scores 98.59% training accuracy but only 54.44% validation accuracy when the last 15% of the model is retrained. On the other hand, the baseline model scores more than 93% classification accuracy in both the training and validation sets. To better illustrate these findings, we show the variation of the loss and accuracy functions during the training of the MobileNetv2 model in Figure 5.5 and the baseline model in Figure 5.6.

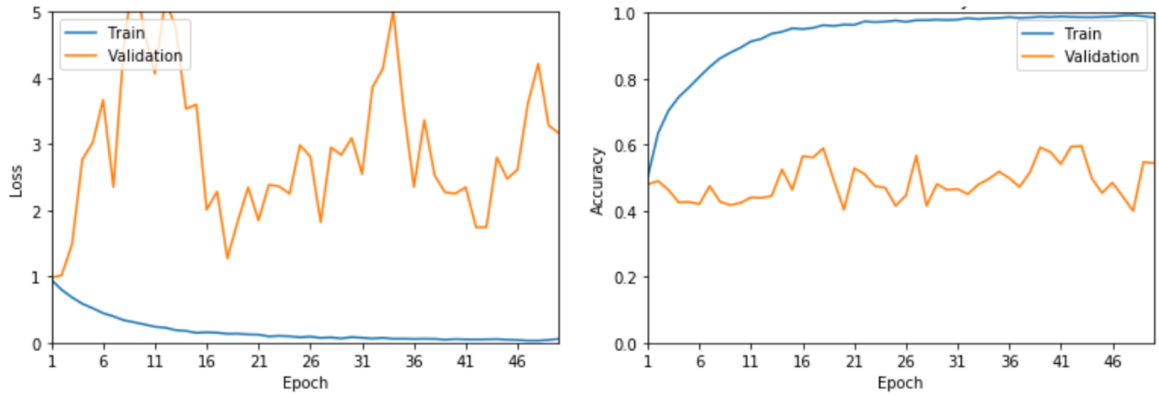


Figure 5.5: Fine-tuning the last 15% of the MobileNetv2 model. Left: Loss function. Right: Accuracy.

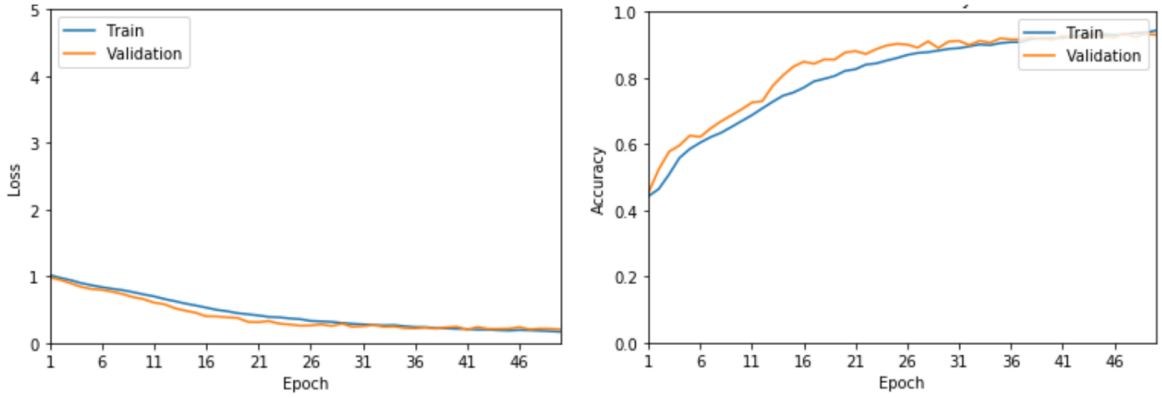


Figure 5.6: Training of the baseline model. Left: Loss function. Right: Accuracy.

On average, the difference between the final training and validation accuracy for the six state-of-the-art models is 18.2% when finetuning only the fully connected layers, 26.6% when finetuning the last 5% of layers, and 29.7% when finetuning the last 15% of layers. In contrast, the difference between the final training and validation accuracy for the baseline model is just 1.3%.

To summarize, the desired performance for any Machine Learning model is to have high accuracy in both the training and validation sets; however, in our experiments, this goal was only achieved by the baseline model. While multiple explanations can clarify these results, the most reasonable one is that the RSC image classification task might not require those complex DL models and a competitive accuracy can be scored using a simpler DL model like the baseline.

Overfitting is a common issue in Machine Learning, and it happens in cases where the model is too complex for a task and as a result, it memorizes the characteristics of the training set. In other words, the model does not generalize well for data samples not observed during training, like the images in the validation or the test sets. Multiple regularization techniques are available to avoid or mitigate overfitting in DL models, such as reducing the number of layers, reducing the number of neurons, or using Dropout layers that randomly disconnect certain connections between neurons, among others.

To avoid overfitting in our study, we use three Dropout layers with a rate of 0.5 between the fully connected layers of all models. The rate defines the frequency at which the connections between neurons in those layers are randomly dropped, the higher the rate the stronger the regularization effect; however, using a rate that is too high could hinder the ability of the model to learn any meaningful patterns during training.

The number of Dropout layers as well as the rate we use are appropriate according to the practices commonly found in the literature for these type of DL models. It is worth to note that each of the DL models we used from the literature already includes their own strategies to mitigate overfitting; therefore, we do not apply additional measures. Instead, we introduce the baseline model with the goal of evaluating the performance of a much simpler architecture with a significantly smaller number of layers and parameters.

5.5 Ablation Study

We also investigate how much we can simplify the baseline model without affecting the resulting classification accuracy. Experiments that remove some parts of a Machine Learning model and look at the effects it produces on the model performance are known as ablation

studies and are recommended especially for DL models [67] [39]. We explore two different ways to reduce the complexity of the baseline model while maintaining its classification accuracy. The first is focused on reducing the number of channels in the intermediate convolutional layers (feature learning part) and the second looks at reducing the number of neurons in the fully connected layers (classification part). Figure 5.7 shows the initial architecture of our baseline model.

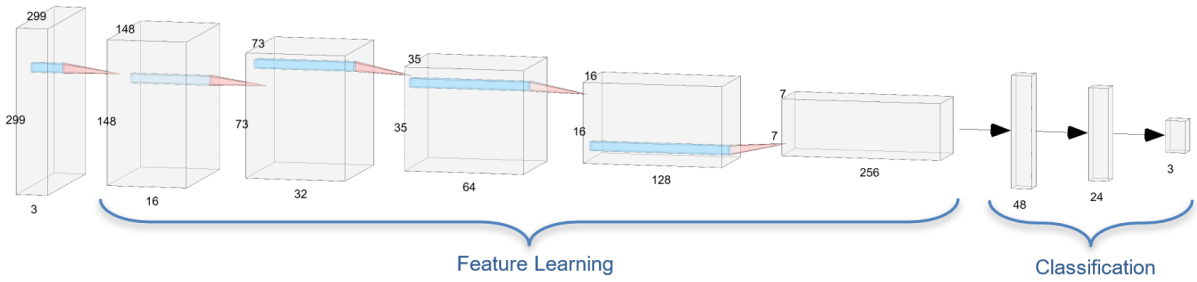


Figure 5.7: The architecture of the baseline model.

In this model, the number of channels doubles from one convolutional layer to the following, the same as in the architecture of VGG [56]. We call the rate at which the number of channels grows as Incremental Channels Factor (ICF) and explore the effect of using an ICF smaller than two. Figure 5.8 shows the effects of reducing the ICF in the baseline model where the area of the circles is representative of the total number of parameters in the model. Lowering the ICF reduces the validation accuracy; however, the decrement is not significant when the value is kept equal or above 1.7; therefore, we move forward using this number. Using a smaller ICF also reduces the number of parameters by approximately 54%, from 996,019 parameters in the original baseline model down to 460,247, which means a significant gain in memory efficiency.

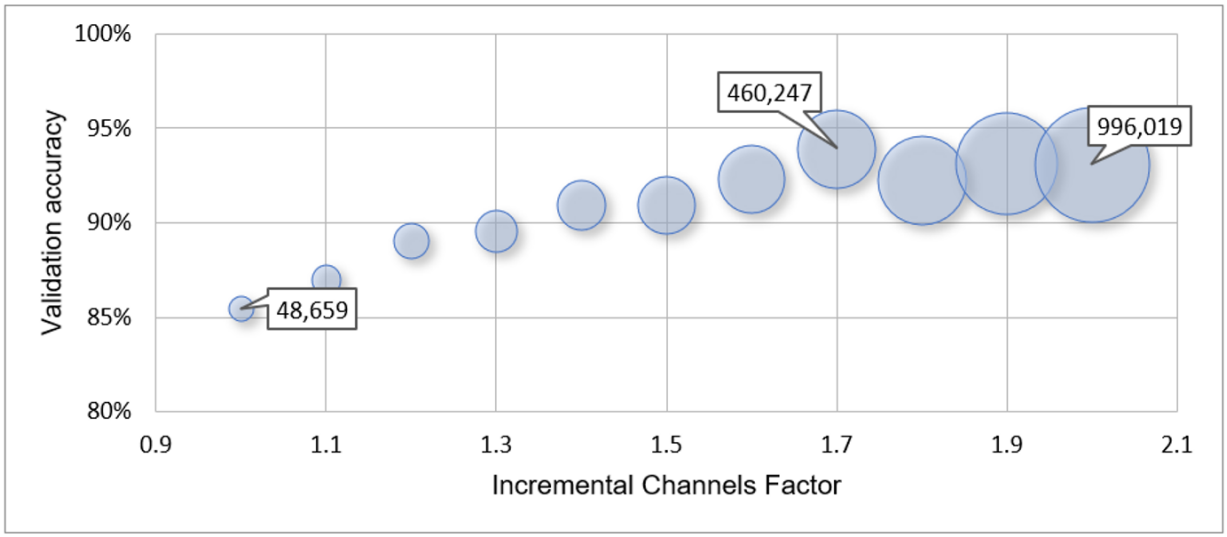


Figure 5.8: Change in validation accuracy as a result of reducing the number of channels between convolutional layers.

The next strategy we use to lessen the complexity of the baseline model focuses on the gradual reduction of the number of neurons in the classification part of the model. The original baseline model uses a combination of 48 and 24 neurons in the first and second fully connected layers, and 3 neurons in the last classification layer, which adds up to a total of 75 neurons. We note that the number of neurons in the last layer must be kept equal to the number of categories in the classification task, which is 3 for determining RSC (bare pavement, partial cover, full cover).

On each consecutive experiment, we eliminate 6 neurons in the first fully connected layer and 3 neurons in the second, until we end up with a combination of 6, 3, and 3 neurons for a total of 12 units in the classification part of the model. Figure 5.9 shows how the accuracy is affected by gradually reducing the number of neurons from a total of 75 down to 12, along with the correspondent reduction in the total number of parameters in the model represented by the area of the circles. The validation accuracy drops drastically when the total number of neurons goes below 30; therefore, we move on using a total of 39 units in order to keep the accuracy above 90%, which left us with a total of 301,727 parameters.

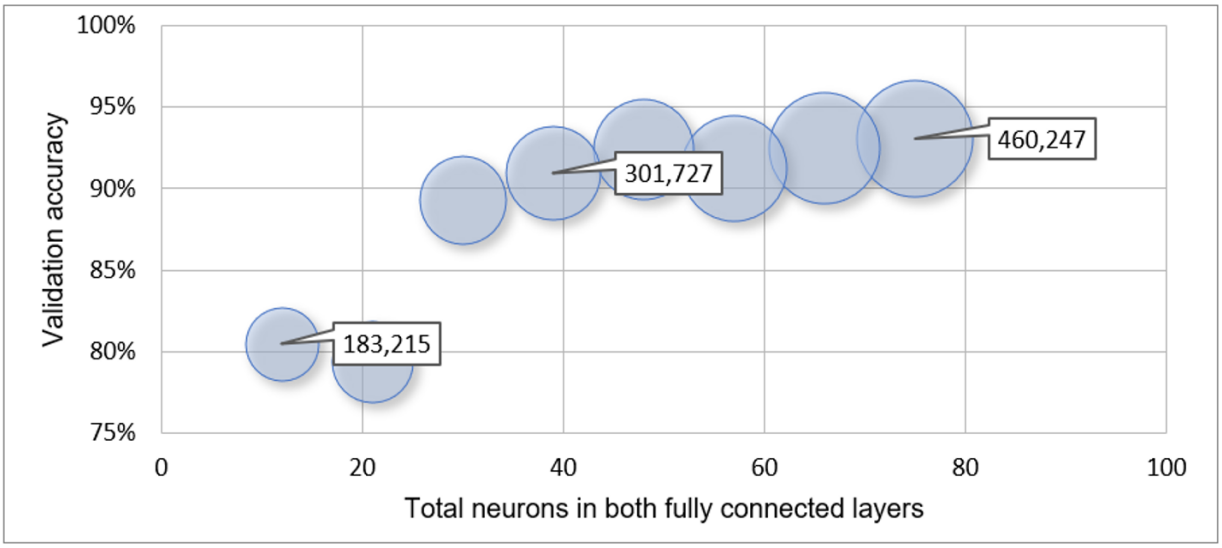


Figure 5.9: Change in validation accuracy as a result of reducing the number of neurons in the classification part of the model.

The results of the ablation study allow us to lower the complexity of the baseline model without a significant reduction in validation accuracy. This *simplified* version of the baseline model includes about 30% the number of parameters compared to the original one and only 1.3% when compared to the average number of parameters in the six state-of-the-art models we evaluate in the previous section.

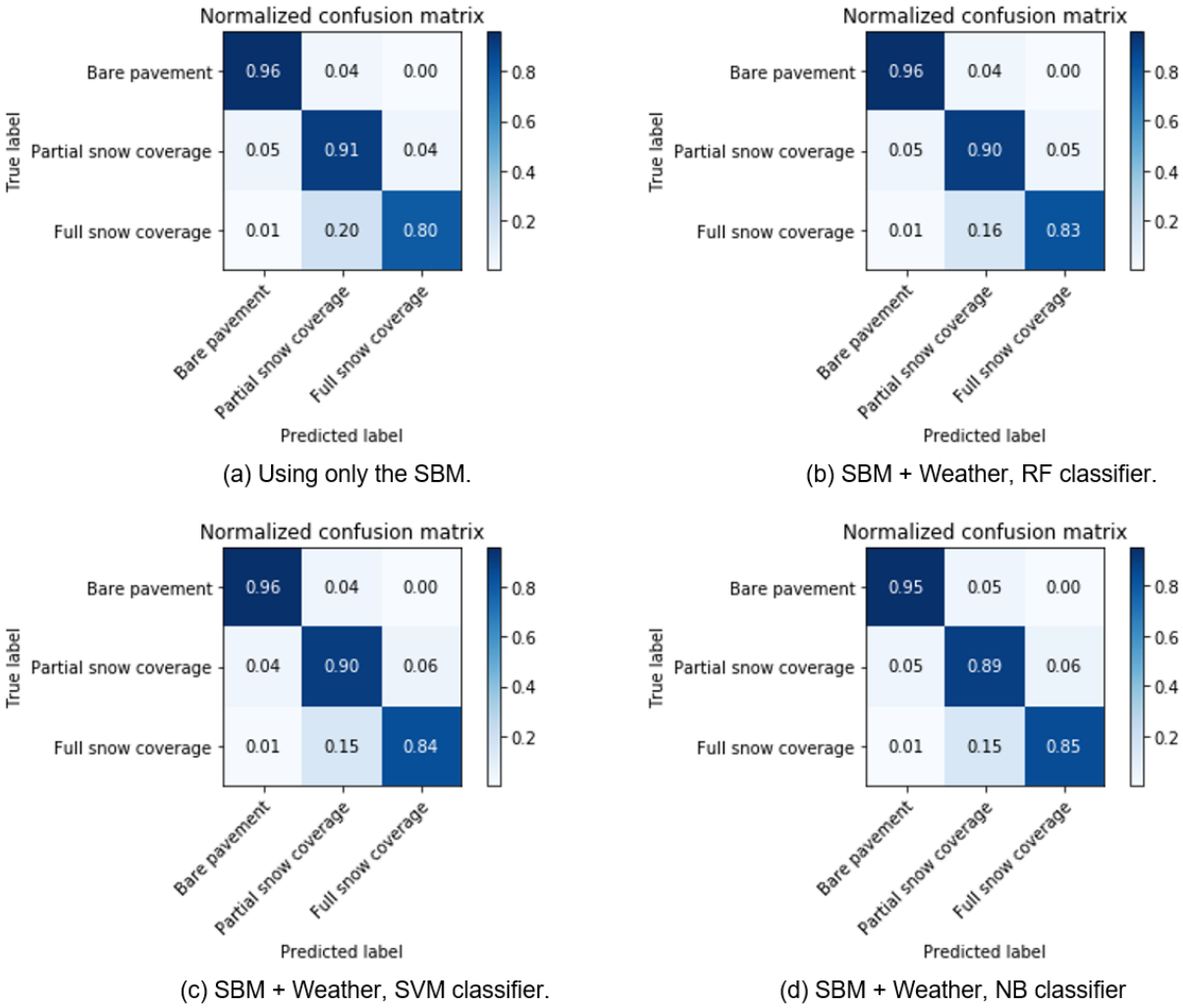
Even though the accuracy went from 93% to 91% during the ablation study, our purpose here is to explore the effects of multiple design criteria rather than scoring the maximum possible accuracy. Using DL models with a fewer number of parameters is highly desirable for reducing training time during the design stage and for deployment in a production environment, especially for applications where input images come from hundreds of cameras every 10 to 15 minutes like in the case of determining RSC for roads in the province of Ontario.

5.6 Combined use of Weather Variables

In this section, we evaluate the benefits of using weather data along with images from roadside cameras as input for monitoring RSC. RWIS stations can have different sensors depending on the vendor; however, most stations record at least the following five variables:

Air Temperature (C), Dew Point (C), Relative Humidity (%), Pressure (kPa), and Wind Speed (km/h). For each image in the dataset (14,000) we obtain these five variables, for a total of 70,000 weather observations; however, since Dew Point is usually inferred from Air Temperature and Relative Humidity, we discard that variable for the subsequent analysis.

We concatenate the output of the simplified baseline model (SBM), which is a vector of probabilities for each one of the three RSC categories listed by the Ministry of Transportation in Ontario, with the values of the weather variables recorded in the same instant and location as the input image. The resulting vector includes seven features in total. About 1.7% of observations for Relative Humidity and 0.26% for Wind Speed appear as Null values; therefore, we fill those gaps with their corresponding average per feature. Z-score standardization per feature was also used. Three Machine Learning models for classification are considered: Random Forest (RF), Support Vector Machine (SVM), and Naive Bayes (NB). For all three we use grid search and cross-validation to find an appropriate set of parameters and plot the normalized confusion matrices over the test set as shown in Figure 5.10.



(a) Using only the SBM.

(b) SBM + Weather, RF classifier.

(c) SBM + Weather, SVM classifier.

(d) SBM + Weather, NB classifier

Figure 5.10: Normalized confusion matrices calculated over the test set.

The normalized confusion matrices show that including weather data improves the ability of the models to distinguish images with full snow coverage, with increments of 3%, 4%, and 5% in test accuracy when using the RF, SVM, and NB models accordingly. However, these improvements are in part due to a slight reduction in the classification accuracy for the other two categories. Based on the experimental results, we suggest using the Naive Bayes classifier to achieve the highest accuracy per RSC category, although the F1-scores for the three models are very similar, with 91.78% for RF, 91.80% for SVM, and 91.24% for NB.

The optimal combination of parameters found during the grid search procedure for the three methods are as follow:

- Random Forest
  - Number of trees = 50
  - Max depth = 6
  - Minimum samples per leaf = 4
- SVM
  - Kernel is radial basis function
  - Penalty of the error term = 100
  - gamma = 0.1
- Naive Bayes
  - Variable smoothing = 0.01

5.7 Summary

The results of this study confirm the effectiveness of DL models for determining RSC from roadside camera images. We evaluate the performance of seven DL models under multiple training and fine-tuning scenarios and select the baseline model as the one that produces the best experimental accuracy. We then explore the simplification of the baseline model through an ablation study and find that we can achieve 91% accuracy using a model that is 98.7% smaller than recently published state-of-the-art DL models. However, we are aware that these results are only indicative of the performance of these models for our specific application. Furthermore, we find that using weather data slightly improves the classification accuracy for images in the fully covered RSC category.

Chapter 6

Conclusions and Future Directions

While there is a direct applicability of the methods presented in this paper to improve winter road maintenance operations, some challenges are worth to note, such as the required changes for the current technology infrastructure in order to accommodate the new software components running these automated tasks, as well as the introduction of visualization dashboards in the monitoring centres for the aggregation of results.

We highlight the value of the experimental results presented in this study toward the automated detection of RSC for winter road monitoring, not only for improving road safety but also for resource optimization. The contributions of this thesis are aligned with one of the strategic objectives of Canadas Road Safety Strategy 2025: Leveraging technology and innovation and two of the key road safety suggested interventions: use of technology and data-driven research [11]. Moreover, the use of Machine Learning techniques for automating data processing can complement other applications on winter road safety such as accident prediction models [66].

Future research could focus on replicating this study in other provinces across Canada and countries in the Northern hemisphere, as well as exploring the integration of these automated methods with the software architecture of the current monitoring systems.

References

[1] Ontario Road Network: Road Net Element — Ontario.ca.

[2] Location Optimization of Road Weather Information System (RWIS) Network Considering the Needs of Winter Road Maintenance and the Traveling Public. *Computer-Aided Civil and Infrastructure Engineering*, 32(1):57–71, 2017.

[3] Early winter storms risk burying city's cleanup budget — Ottawa Citizen, 2019.

[4] Here's what it costs to keep Toronto safe in the winter - The Weather Network, 2019.

[5] GA Achilleos. The inverse distance weighted interpolation method and error propagation mechanism–creating a dem from an analogue topographical map. *Journal of Spatial Science*, 56(2):283–304, 2011.

[6] Ana M. T. Amorim, Alexandre B. Gonçalves, Luís Miguel Nunes, and António Jorge Sousa. Optimizing the location of weather monitoring stations using estimation uncertainty. *International Journal of Climatology*, 32(6):941–952, may 2012.

[7] Hossein Azizpour, Ali Sharif Razavian, Josephine Sullivan, Atsuto Maki, and Stefan Carlsson. Factors of Transferability for a Generic ConvNet Representation. jun 2014.

[8] Matts-Åke Belin, Per Tillgren, and Evert Vedung. Vision Zero a road safety policy innovation. *International Journal of Injury Control and Safety Promotion*, 19(2):171–179, jun 2012.

[9] B Sue Bell, Richard E Hoskins, Linda Williams Pickle, and Daniel Wartenberg. Current practices in spatial analysis of cancer data: mapping health statistics to inform policymakers and the public. *International journal of health geographics*, 5(1):49, 2006.

[10] Finlay Buchanan and S E Gwartz. Road Weather Information Systems at the Ministry of Transportation, Ontario. Technical report.

[11] Canadian Council of Motor Transport Administrators CCMTA. Canada's Road Safety Strategy 2025 Towards Zero: The Safest Roads in the World. Technical Report January, Canadian Council of Motor Transport Administrators, 2016.

[12] Alfredo Canziani, Adam Paszke, and Eugenio Culurciello. An analysis of deep neural network models for practical applications. *CoRR*, abs/1605.07678, 2016.

[13] François Chollet. Xception: Deep Learning with Depthwise Separable Convolutions. oct 2016.

[14] City of Toronto. Vision Zero: Toronto's Road Safety Plan. Technical report, Transportation Services City of Toronto, Toronto, Ontario, Canada, 2017.

[15] M Coulibaly and S Becker. Spatial interpolation of annual precipitation in south africa-comparison and evaluation of methods. *Water International*, 32(3):494–502, 2007.

[16] Michael John De Smith, Michael F Goodchild, and Paul Longley. *Geospatial analysis: a comprehensive guide to principles, techniques and software tools*. Troubador publishing ltd, 2007.

[17] Jia Deng, Jia Deng, Wei Dong, Richard Socher, Li-jia Li, Kai Li, and Li Fei-fei. Imagenet: A large-scale hierarchical image database. *IN CVPR*, 2009.

[18] Philip M Dixon. Ripley's K Function. *Wiley StatsRef: Statistics Reference Online*, 2014.

[19] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016.

[20] Ministry of Transportation Government of Ontario. Winter road conditions.

[21] Murtaza Haider and Eric J Miller. Effects of transportation infrastructure and location on residential real estate values: application of spatial autoregressive techniques. *Transportation Research Record*, 1722(1):1–8, 2000.

[22] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[23] Brian Hirt. Installing Snowplow Cameras and Integrating Images into MnDOT's Traveler Information System. Technical report, 2017.

[24] Jaroslav Hofierka, Juraj Parajka, Helena Mitasova, and Lubos Mitas. Multivariate interpolation of precipitation using regularized spline with tension. *Transactions in GIS*, 6(2):135–150, 2002.

[25] Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q. Weinberger. Densely Connected Convolutional Networks. aug 2016.

[26] Michael F Hutchinson, Dan W McKenney, Kevin Lawrence, John H Pedlar, Ron F Hopkinson, Ewa Milewska, and Pia Papadopol. Development and testing of canada-wide interpolated spatial models of daily minimum–maximum temperature and precipitation for 1961–2003. *Journal of Applied Meteorology and Climatology*, 48(4):725–741, 2009.

[27] Peter J Jin, Andrew Walker, Meredith Cebelak, C Michael, Walton P J Jin, M Cebelak, C M Walton, and P J Jin. Determining Strategic Locations for Environmental Sensor Stations with Weather-Related Crash Data TX 78701. Current affili-ation for. *Transportation Research Record: Journal of the Transportation Research*, 2440:34–42, 2014.

[28] Ilkka Juga, Pertti Nurmi, and Marjo Hippi. Statistical modelling of wintertime road surface friction. *Meteorological Applications*, 20(3):318–329, sep 2013.

[29] Diane Kelsall and Donald A. Redelmeier. Winter road safety is no accident. *Canadian Medical Association Journal*, 188(4):241–241, mar 2016.

[30] Simon Kornblith, Jonathon Shlens, and Quoc V. Le. Do Better ImageNet Models Transfer Better? 2018.

[31] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[32] Tae J. Kwon and Liping Fu. Evaluation of alternative criteria for determining the optimal location of RWIS stations. *Journal of Modern Transportation*, 21(1):17–27, mar 2013.

[33] Tae Jung Kwon, Liping Fu, and Chaozhe Jiang. Road weather information system stations where and how many to install: a cost benefit analysis approach. *Canadian Journal of Civil Engineering*, 42(1):57–66, jan 2015.

[34] Michael A. Linton and Liping Fu. Winter Road Surface Condition Monitoring: Field Evaluation of a Smartphone-Based System. *Transportation Research Record: Journal of the Transportation Research Board*, 2482(1):46–56, jan 2015.

[35] Michael A. Linton and Liping Fu. Connected Vehicle Solution for Winter Road Surface Condition Monitoring. *Transportation Research Record: Journal of the Transportation Research Board*, 2551:62–72, 2016.

[36] Franz Loewenherz, Victor Bahl, and Yinhai Wang. Video Analytics Towards Vision Zero. *ITE Journal*, 87, may 2017.

[37] George Y Lu and David W Wong. An adaptive inverse-distance weighting spatial interpolation technique. *Computers & geosciences*, 34(9):1044–1055, 2008.

[38] Bonnie Lysyk. Special Report, April 2015: Winter Highway Maintenance. Technical report, Office of the Auditor General of Ontario, 2015.

[39] Ningning Ma, Xiangyu Zhang, Hai-Tao Zheng, and Jian Sun. Shufflenet v2: Practical guidelines for efficient cnn architecture design. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 116–131, 2018.

[40] Katrin Meusburger, A Steel, P Panagos, L Montanarella, and Christine Alewell. Spatial and temporal variability of rainfall erosivity factor for switzerland. *Hydrology and Earth System Sciences*, 16(1):167–177, 2012.

[41] Helena Mitášová and Lubos Mitáš. Interpolation by regularized spline with tension: I. theory and implementation. *Mathematical geology*, 25(6):641–655, 1993.

[42] TG Mueller, SRK Dhanikonda, NB Pusuluri, AD Karathanasis, KK Mathias, B Mijatovic, and BG Sears. Optimizing inverse distance weighted interpolation with cross-validation. *Soil science*, 170(7):504–515, 2005.

[43] Marcus Nolte, Nikita Kister, and Markus Maurer. Assessment of Deep Convolutional Neural Networks for Road Surface Classification. 2018.

[44] Raqib Omer and Liping Fu. An automatic image recognition system for winter road surface condition classification. In *13th international IEEE conference on intelligent transportation systems*, pages 1375–1379. IEEE, 2010.

[45] Guangyuan Pan, Liping Fu, Ruifan Yu, and Matthew Muresan. Evaluation of Alternative Pre-trained Convolutional Neural Networks for Winter Road Surface Condition Monitoring. Technical report.

[46] Guangyuan Pan, Liping Fu, Ruifan Yu, and Matthew D Muresan. Winter Road Surface Condition Recognition Using a Pre-trained Deep Convolutional Neural Network. In *Transportation Research Board 97th Annual Meeting*, Washington DC, United States, 2018.

[47] Clark Pennelly, Gerhard W Reuter, and Stevanus Tjandra. Effects of Weather on Traffic Collisions in Edmonton, Canada.

[48] Dieter Pennerstorfer. Spatial price competition in retail gasoline markets: evidence from austria. *The Annals of Regional Science*, 43(1):133–158, 2009.

[49] Nikiforos Pittaras, Foteini Markatopoulou, Vasileios Mezaris, and Ioannis Patras. Comparison of fine-tuning and extension strategies for deep convolutional neural networks. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 10132 LNCS, pages 102–114, 2017.

[50] Paddy M Prenter et al. *Splines and variational methods*. Courier Corporation, 2008.

[51] Damian Rzeznikiewiz, Hala Tamim, and Alison K. MacPherson. Risk of death in crashes on Ontario's highways. *BMC Public Health*, 12(1):1125, dec 2012.

[52] Mitsuru Saito and Seishi Yamagata. Effect of Video Camera-Based Remote Roadway Condition Monitoring on Snow Removal-Related Maintenance Operations Final Report. Technical report, Engineering Department at Brigham Young University, 2014.

[53] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. MobileNetV2: Inverted Residuals and Linear Bottlenecks. jan 2018.

[54] Luis Fernando Santa Guzman, Roverth Steven Pinzón Rodríguez, Juan Sebastián Pulido Ávila, Luis Fernando Gómez Rodríguez, et al. Diseño de una red de muestreo espacial que optimiza la estimación de parámetros de la función de covarianza para la medición pluviométrica en el departamento de cundinamarca. *UD y la geomática*, (7):75–85, 2013.

[55] Xianming Shi and Liping Fu. *Sustainable Winter Road Operations*. Wiley Online Library, 2018.

[56] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[57] Amit Kumar Singh, Jia Li, Mike Murphy, C Michael, Walton A K Singh, and A K Singh. Optimizing Environmental Sensor Station Locations for Road Weather Management Overview and a Generalized Modeling Framework. *Transportation Research Record: Journal of the Transportation Research*, 2551:18–28, 2016.

[58] Paul W. Steely. Vision Zero Cities. *International Journal of Traffic Safety Innovation*, 3, 2018.

[59] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alex Alemi. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. feb 2016.

[60] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Computer Vision and Pattern Recognition (CVPR)*, 2015.

[61] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, and Zbigniew Wojna. Rethinking the Inception Architecture for Computer Vision. dec 2015.

[62] Eric Tang, Kara Peach, Joshua DeFisher, and Kim Eccles. Transportation Safety Planning and the Zero Deaths Vision: A Guide for Metropolitan Planning Organizations and Local Communities. Technical report, 2018.

[63] Russell Tessier and University of Massachusetts at Amherst. Evaluation of Portable Road Weather Information Systems, jun 2016.

[64] Finlay Transportation Association of Canada. and S E Gwartz. *2005 Annual Conference and Exhibition of the Transportation Association of Canada : Transportation - Investing in Our Future.* Transportation Association of canada, 2005.

[65] Taimur Usman, Liping Fu, and Luis F. Miranda-Moreno. Quantifying safety benefit of winter road maintenance: Accident frequency modeling. *Accident Analysis & Prevention*, 42(6):1878–1887, nov 2010.

[66] Taimur Usman, Liping Fu, and Luis F Miranda-Moreno. Accident prediction models for winter road safety: Does temporal aggregation of data matter? *Transportation research record*, 2237(1):144–151, 2011.

[67] Bolei Zhou, Yiyou Sun, David Bau, and Antonio Torralba. Revisiting the importance of individual units in cnns via ablation. *arXiv preprint arXiv:1806.02891*, 2018.

[68] Barret Zoph, Vijay Vasudevan, Jonathon Shlens, and Quoc V. Le. Learning Transferable Architectures for Scalable Image Recognition. jul 2017.