Development of Pavement Distress Detection Model Utilizing SOTA Deep Learning Algorithm "YOLOv5"

Ashkan Behzadian

Submitted to the
Institute of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Master of Science
in
Civil Engineering

Eastern Mediterranean University
July 2022
Gazimağusa, North Cyprus

Approval of the Institute of Graduate Studies and Research

Prof. Dr. Ali Hakan Ulusoy
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science in Civil Engineering.

Prof. Dr. Umut Türker
Chair, Department of Civil Engineering

We certify that we have read this thesis and that in our opinion it is fully adequate in scope and quality as a thesis for the degree of Master of Science in Civil Engineering.

Assoc. Prof. Dr. Mehmet Metin Kunt
Supervisor

Examining Committee

1. Prof. Dr. Mustafa Ergil

2. Assoc. Prof. Dr. Mehmet Metin Kunt

3. Asst. Prof. Dr. Hüseyin Sevay

ABSTRACT

One of the most critical issues in pavement asset management is evaluating the performance of the roads and highways. This crucial task is currently handled by regular manual inspection in many countries, which is inaccurate and sometimes dangerous. However, this inspection is processed automatically utilizing specifically designed vehicles in some developed countries; many municipalities and road agencies worldwide are still using manual methods due to the high expenses of purchasing and maintaining specific vehicles. Due to the recent advancements in computer vision, researchers and scholars use deep learning technology to enhance road inspection. Some high-tech infrastructure cities benefit from this technology to handle various infrastructural issues. Roads and pavements are no exception in this area. Currently, some intelligent cities are using deep learning technology to evaluate road performance. Pavement distress detection is one of the critical issues in this field. Many researchers worldwide use deep learning technics, expressly object detection algorithms, to automate road performance evaluation.

This study aims to provide a robust and reliable model for detecting and classifying several types of pavement distresses with high accuracy. Road agencies and municipalities could use this model to collect data on road sections conveniently and affordably. In this case, authorities could monitor the pavement condition in short intervals and make appropriate decisions on maintenance and rehabilitation strategies and methods which could result in maintaining the performance of the pavement an acceptable quality by lower costs.

This study developed a model to detect and classify pavement distresses on the surface of the road utilizing state of the art deep learning algorithm (YOLOv5) as well as most recent prominent optimization strategies such as data augmentation and hyperparameter tuning to propose an accurate, robust, and reliable model. 628 top-down view pavement images used in this study were captured in several cities in the U.S., including various distress types such as alligator cracking, longitudinal cracking, transverse cracking, block cracking, patching, sealing, and manhole. The performance of the proposed model is evaluated based on several criteria. The model's accuracy reached 0.95, 0.92, and 0.93 in precision, recall, and F1 score, respectively.

Keywords: pavement asset management, automatic road condition monitoring, pavement distress detection, artificial intelligence, deep learning.

ÖZ

Üstyapı varlık yönetimindeki en kritik konulardan biri yolların ve otoyolların performansının değerlendirilmesidir. Bu çok önemli görev, şu anda birçok ülkede hatalı ve bazen tehlikeli olabilen insan gücü ile gerçekleştirilmektedir. Bu muayene bazı gelişmiş ülkelerde özel olarak tasarlanmış araçlar kullanılarak otomatik olarak işlenmesine rağmen, dünya çapında birçok belediye ve karayolu kurumu, belirli araçların satın alınması ve bakımının yüksek maliyetleri nedeniyle hala insan gücü kullanmaya devam etmektedir. Bilgisayarlı görü alanındaki son gelişmeler nedeniyle, araştırmacılar ve akademisyenler, yol denetimini geliştirmek için derin öğrenme teknolojisini kullanıyor. Bazı yüksek teknolojili altyapı şehirleri, çeşitli altyapı sorunlarını ele almak için bu teknolojiden yararlanır. Yollar ve kaldırımlar bu alanda bir istisna değildir. Şu anda bazı akıllı şehirler, yol performansını değerlendirmek için derin öğrenme teknolojisini kullanıyor. Üstyapı tehlike tespiti bu alandaki kritik konulardan biridir. Dünya çapında birçok araştırmacı, yol performans değerlendirmesini otomatikleştirmek için derin öğrenme tekniklerini, özellikle nesne algılama algoritmalarını kullanır.

Bu çalışma, çeşitli üstyapı bozulma tiplerini yüksek doğrulukla tespit etmek ve sınıflandırmak için sağlam ve güvenilir bir model sağlamayı amaçlamaktadır. Yol ajansları ve belediyeler, yol kesimleri hakkında uygun olan ve düşük maliyetli yöntemlerle veri toplamak için bu modeli kullanabilir. Bu durumda yetkililer, üstyapının durumunu kısa aralıklarla izleyebilir ve üstyapı performansının daha düşük maliyetlerle kabul edilebilir bir kalitede sürdürülebilmesine yardımcı olacak bakım ve iyileştirme stratejileri ve yöntemleri konusunda uygun kararlar verebilir.

Bu çalışma, son teknoloji derin öğrenme algoritmasını (YOLOv5) ve ayrıca veri büyütme ve hiperparametre ayarlama gibi en son öne çıkan optimizasyon stratejilerini kullanarak yol yüzeyindeki kaplama bozukluklarını tespit etmek ve sınıflandırmak için doğru, sağlam ve güvenilir bir model geliştirdi. Bu çalışmada kullanılan, ABD'deki çeşitli şehirlerden alınmış 628 yukarıdan aşağıya görünüm kaldırım resimleri timsah sırtı çatlaması, boyuna çatlama, enine çatlama, blok çatlaması, yama, sızdırmazlık ve menhol gibi çeşitli bozukluk türlerini aktarmaktadır. Önerilen modelin performansı çeşitli kriterlere göre değerlendirildi. Modelin doğruluğu kesinlik, hatırlama ve F1 puanlarında sırasıyla 0.95, 0.92 ve 0.93'e ulaştı.

Anahtar Kelimeler: üstyapı varlık yönetimi, otomatik yol durumu izleme, üstyapı tehlike tespiti, yapay zeka, derin öğrenme.

ACKNOWLEDGMENT

This thesis would not have been possible without the help, motivation, and encouragement of my supervisor, Assoc. Prof. Dr. Mehmet Metin Kunt. I would like to thank him for his excellent guidance and professional support throughout this project.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
ANN	Artificial Neural Network
ARAN	Automatic Road Analyzer
CmBN	Cross mini-Batch Normalization
CSP	Cross-Stage-Partial-connections
DL	Deep Learning
DT	Decision Trees
IRI	International Roughness Index
LTPP	Long-Term Pavement Performance
MAP	Mean Average Precision
MIoU	Mean Intersection over Union
ML	Machine Learning
NLP	Natural Language Processing
PCE	Pavement Condition Evaluation
PCI	Pavement Condition Index
PMS	Pavement Management System
R-CNN	Region-based Convolutional Neural Network
SAT	Self Adversarial Training
SPP	Spatial Pyramid Pooling
SSD	Single Shot Detection
SVM	Support Vector Machine
WRC	Weighted-Residual-Connections
YOLO	You Only Look Once

Chapter 1

INTRODUCTION

1.1 Background

Transportation plays a significant role in today's society. Higher quality of transportation systems could result in a lower cost of accessibility for goods and services as well as less environmental contamination. Pavements are one of the most crucial parts of transportation in every society. It is essential to construct and maintain these invaluable assets in an optimized manner. Pavement Management System (PMS) represents the efficient management of the roads and highway assets (Haas, Hudson, & Zaniewski, 1994). This concept includes decision-making on constructing new roads and highways and maintaining strategies for the current roads. PMS utilizes various tools to provide the optimum decisions and funding strategies. One of the most critical issues in this topic is evaluating the performance of the pavement and predicting the serviceability for each specific road section. The accuracy of pavement evaluation helps the agencies make better decisions on pavement treatment methods, resulting in higher pavement performance and lower cost.

1.2 Problem Statement

Pavement Condition Evaluation (PCE) is one of the most crucial issues in PMS (Pavement Management System) (Haas, Hudson, & Falls, 2015). It is essential for road agencies and municipalities to acquire robust and reliable data regarding the distresses and deficiencies of the roads in various stages for the entire lifetime of the pavements. The process of collecting the road data could be done in several ways. First and the

traditional manner is collecting the data manually using virtual inspection. This method has several deficiencies, such as the low accuracy of the collected data and the danger of data collection on highways with high traffic loads.

Furthermore, it could sometimes be impossible to acquire the appropriate data quality on some roads with 24/7 traffic, such as New York City roads. The second method of collecting the data is automated data collection utilizing some specifically designed vehicles. For instance, Automatic Road Analyzer (ARAN) is a vehicle equipped with high-precision sensors to collect pavement data automatically. The cost of purchasing such a vehicle could be more than one million dollars and additional thousands of dollars monthly for maintaining the device. Although the accuracy of this method is higher than the manual inspection, it could be costly, especially for developing countries, because of the high price of purchasing these vehicles as well as the high expenses of maintenance.

Nowadays, due to the recent advancement of Artificial Intelligence (AI), it is becoming increasingly popular for road agencies and municipalities to leverage this technology to overcome complex problems such as data collection. During recent years, many types of research and studies have been conducted to automate pavement condition evaluation without utilizing expensive facilities. Due to the recent advancements in computer science, specifically computer vision, researchers are focusing on benefiting from these technologies to enhance their studies. Deep Learning technology enables researchers to develop models for automatically detecting the pavement surface's distresses (Zakeri, Nejad, & Fahimifar, 2016). However, these studies are still challenged for sufficient accuracy, robustness, and reliability.

1.3 Research Objectives

This study developed a pavement condition evaluation model based on the data collected from several sources such as Google Street Views and ARAN vehicles in various cities in the United States of America (i.e., Jefferson City, Kansas City, and Columbia in Missouri). The data collected included several distress types of the pavement, and the proposed model is based on the most recent technologies in Computer Vision and AI.

The objectives of the research are:

- Developing pavement distress detection model based on the state-of-the-art deep learning techniques
- Optimizing and generalizing the proposed model based on recent prominent strategies
- Tuning and standardizing the proposed model for detecting distresses of pavement.

Chapter 2

THEORY

2.1 Pavement Surface Distresses

The deficiencies in pavements play an essential role in evaluating the performance of the pavement. The distresses could have both external and internal causes that influence the pavement's quality. Some of these factors are heavy traffic load and environmental conditions such as hot and freezing weather, freeze and thaw process, and heavy rain or snow. These factors have a destructive impact on the pavement's surface, resulting in deterioration of the pavement throughout its lifetime.

There are several indices for measuring the performance of the pavement, such as the Pavement Condition Index (PCI) and International Roughness Index (IRI), which indicate the quality of the pavement based on specific measurements. For instance, in order to calculate the PCI of a specific road section, it is essential to record the type and severity of the existing distresses on that specific section. By counting the number of distresses and relative calculation based on (ASTM, 2018) standard, the PCI index would be calculated, which indicates the pavement's performance on a 0 to 100 scale. Thus, the accuracy of collecting the distress types and severity is crucial for measuring the pavement's performance and making proper decisions for maintenance and rehabilitation strategies. The following section demonstrates some of the main distress types of the pavement's surface and its severity level.

2.1.1 Alligator Cracking

One of the most destructive types of distresses on the pavement surface is alligator cracking, which significantly influences the performance of the pavement. These types of cracks are a series of interconnected cracks which are similar to the back of an alligator that occurs as a result of high traffic load. Figure 1 shows the orientation of this type of crack as well as different severity levels. Based on the FHWA standard (Miller & Bellinger, 2003), this type of crack is categorized into three groups based on different severity levels. The low severity level belongs to those with a few connecting cracks without any spalling or pumping. Figure 2 depicts the low severity of alligator crack. The moderate level refers to the interconnected cracks which formed a recognizable pattern. It might contain a few spalling but without any pumping. Figure 3 illustrates the moderate alligator cracking. Moreover, the last and most severe type of alligator cracking depicted in figure 4 shows a complete pattern of interconnected cracks, including spalling and pumping.

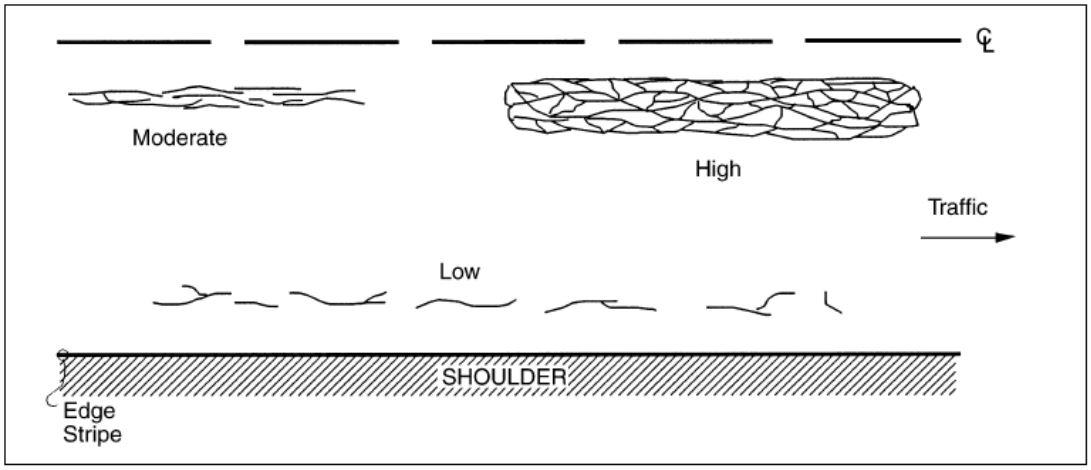


Figure 1: Schematic of alligator cracking with different severity level (Miller & Bellinger, 2003)

Figure 2: Low severity alligator crack

Figure 3: Moderate severity alligator crack

Figure 4: High severity alligator crack

2.1.2 Block Cracking

Another typical distress for the surface of the pavement is block cracking. This type of cracking consists of continuous square shape cracks; in other words, these types of cracks include connected vertical and horizontal cracks throughout the surface of the pavement. One crucial cause of this type of crack is low pavement resistance to the shrinkage occurrence due to the varying weather temperature. The FHWA standard (Miller & Bellinger, 2003) divided the severity of these types of cracks into three levels (low, moderate, and high). The smaller squares of the cracks show the higher severity of this type of distress. Figure 5 illustrates the schematic of block cracking and the possible sizes of the squares. In order to measure the severity of this type of distress, the abovementioned standard defines a numeric approach for categorizing the severity level. In this manner, the cracks with a mean width of lower than 6 mm consider as low severity. The moderate severity level consists of the mean width of the cracks between 6 to 19 mm. Moreover, the high severity level belongs to those with mean crack width higher than 19 mm. Figure 6 shows an example of this type of distress.

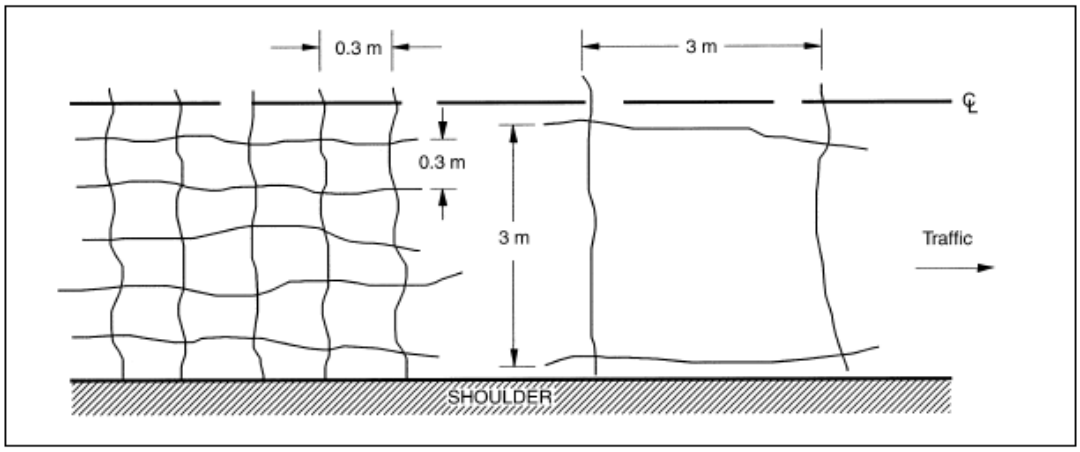


Figure 5: Schematic and orientation of block cracking (Miller & Bellinger, 2003)

Figure 6: Example of block cracking

2.1.3 Longitudinal Cracking

Longitudinal cracks are one of the most typical cracks occurring on the pavement's surface. These cracks occur parallel to the traffic flow and perpendicular to the transverse cracks. The FHWA standard (Miller & Bellinger, 2003) divides these types of distress into two general categories based on the location of the occurrence. Wheel path longitudinal cracks occur strictly under the tire of vehicles. On the other hand, non-wheel path longitudinal cracks occur between the two wheel path. Figure 7 illustrates the orientation of the longitudinal cracks as well as the difference between the wheel pas and non-wheel path cracks.

In order to differentiate between the severity level of this distress type, the standard defines a numerical approach in a way that cracks with a mean width of below 6 mm consider as low severity, cracks between 6 and 19 mm consider moderate severity, and cracks with higher than 19 mm average width consider as high severity. Figure 8 shows the example longitudinal crack.

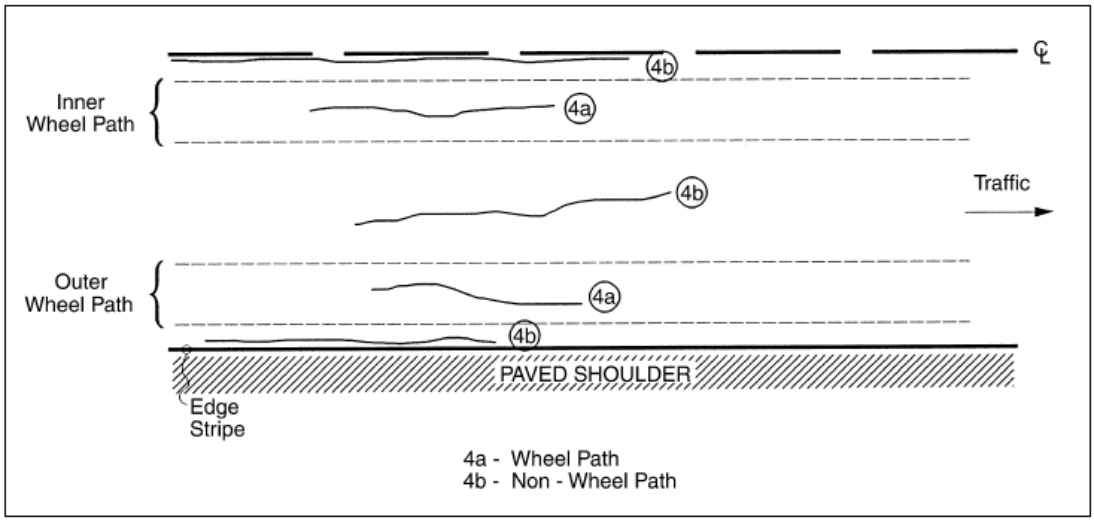


Figure 7: Schematic of longitudinal cracks (Miller & Bellinger, 2003)

Figure 8: Example of longitudinal crack

2.1.4 Transverse Cracking

Another typical type of crack is transverse crack which occurs perpendicular to the traffic flow. These types of cracks may or may not cover the pavement's whole surface. The occurrence of transverse cracking might be a result of two main factors. First is the traffic load, and the second is the climatic issues. Due to the temperature changes during the year, the freeze and thaw phenomenon could occur inside these types of cracks. Transverse cracking could have a dramatic effect on the pavement's rideability and could play a significant role in the performance of the pavement. Similar to longitudinal cracking, this type is categorized into three severity levels based on the average crack width. In this case, cracks with a width lower than 6 mm are considered as low severity, between 6 to 19 mm as moderate, and higher than 19 mm as high severity. Figure 9 shows the schematic of transverse cracks with different severities. Moreover, figures 10, 11, and 12 depicted this distress type with low severity, moderate severity, and high severity, respectively.

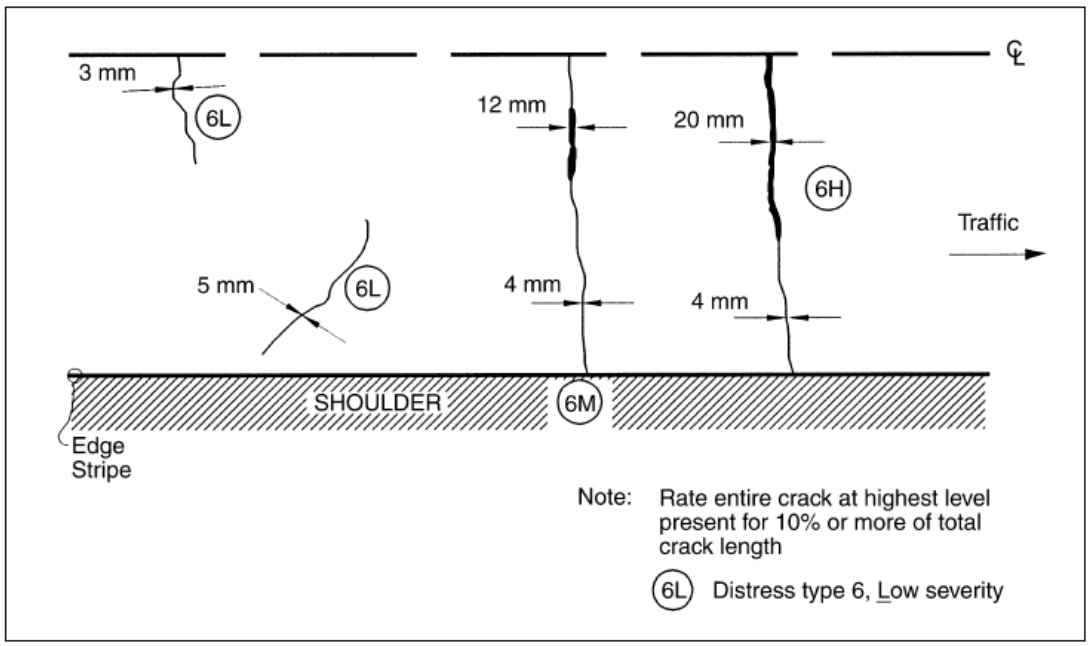


Figure 9: Schematic of transverse cracking with different severity level (Miller & Bellinger, 2003)

Figure 10: Low severity transverse crack


Figure 11: Moderate severity transverse crack

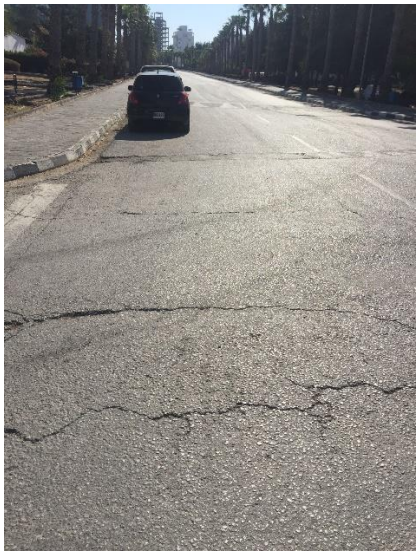


Figure 12: High severity transverse crack

2.2 Pavement Distress Detection

Pavement monitoring and evaluating the performance of the pavement is one of the crucial issues in Pavement Management Systems (PMS). After construction, the road sections need to be monitored and evaluated over their lifetime. This monitoring allows the road agencies to decide on an optimum treatment strategy and prioritize the maintenance and rehabilitation approaches. Hence, the accuracy of detecting the existing distresses of the road surface plays a significant role in providing efficient information for decision-making.

There are two main approaches for evaluating the pavement condition. The first approach is manual inspection conducted by specialists to observe and record the type and severity of pavement distresses. The other approach is automated distress detection using specifically designed vehicles. The first and traditional approach is labor-intensive and time-consuming. Furthermore, this approach might be dangerous in a particular situation, such as visual inspection of highways with high traffic flow. Also, the accuracy of this approach is not sufficient. The second and automated approach has higher accuracy than the manual inspection. However, it requires expensive tools and vehicles such as ARAN to collect the distresses of the roads and evaluate their severity.

Regarding pavement data acquisition, agencies and authorities are interested in automated and semi-automated approaches because of the improvements that have been made for decision making and better safety of procedures as well as the accuracy of the data acquired with these approaches. Nevertheless, this type of pavement

monitoring is not affordable for local agencies or municipalities in developing countries that face a budget shortage for these types of surveys.

2.2.1 Automated Pavement Distress Detection

The concept of automated pavement condition evaluation belongs to 1990, when (Hass & Hendrickson, 1990) proposed the method to evaluate the characteristics of the pavement. They developed a sensor-based model for automatically extracting the road data and properties. In fact, (Hass & Hendrickson, 1990) study introduces automatic pavement monitoring for acquiring pavement data. Figure 13 illustrates the role of automatic pavement monitoring throughout the PMS responsibilities.

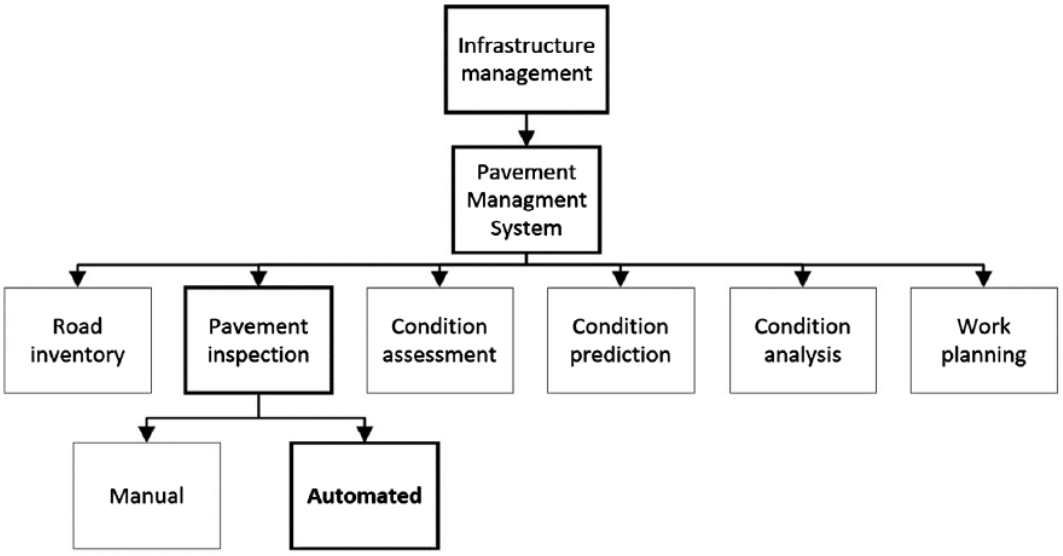


Figure 13: Position of automated pavement monitoring in PMS (Coenen & Golroo, 2017)

After eight years, another research team made significant efforts to develop this concept (Cheng & Miyojim, 1998). They studied on classification algorithm for eliminating the influence of illumination on the acquired data. This study improved the accuracy of collected data for analyzing the properties of the roads. However, these

types of research required further investigation to increase the collected data's accuracy and quality.

After several years, a research team from the Georgia Institute of Technology collaborated with the U.S. Department of Transportation to develop a laser-based technology to detect the cracks on the surface of the pavement (Tsai & Li, 2012). The device was able to detect cracks as narrow as 1 mm in width. However, the model suffered from some errors in detecting the cracks. Also, the proposed device was expensive for many road agencies and was not affordable for local and rural agencies. Figure 14 shows the proposed vehicle for pavement crack detection based on laser technology.

Figure 14: Laser-based vehicle for pavement data collection (Tsai & Li, 2012)

Although much research is conducted to evaluate and monitor the pavement's surface automatically, it still has some deficiencies in terms of accuracy, robustness, and reliability.

2.2.2 Image-Based Pavement Distress Detection

During the recent decades, many scholars and research teams focused on image-based pavement evaluation (Zakeri, Nejad, & Fahimifar, 2016). One of the essential advantages of this method is decreasing the enormous amount of budget spent on developing vehicles for this purpose. Moreover, image-based pavement distress detection considers a non-destructive evaluation of the pavement (Chambon & Moliard, 2011).

Furthermore, due to the advancement in digital image processing, the resolution and quality of collected data could be higher than in the past (Johri, Diván, Khanam, Marciszack, & Will, 2022). Thus, the image-based inspection of the roads could have higher accuracy compared to the traditional methods. In addition, many road agencies are more interested in utilizing time and cost-effective techniques for pavement distress detection.

Chapter 3

LITERATURE REVIEW

3.1 Artificial Intelligence in Civil Engineering

During recent years, Artificial Intelligence (AI) is becoming increasingly widespread among scholars and researchers in various areas of study. This technology enables researchers to solve complex problems and tackle sophisticated issues in many areas such as medicine, healthcare, business, and engineering. For instance, many researchers are developing AI-based models to detect human brain tumors in medicine (Huang, Yang, Fong, & Zhao, 2020). Furthermore, in business, researchers are developing models to predict the future behavior and trend of the stock market (Ferreira, Gandomi, & Cardoso, 2021). In geographic area of research, scholars are developing AI-based models for weather forecasting (Narvekar & Fargose, 2015). It seems that many fields of study are focusing on the use of AI in solving complicated problems and situations. In the engineering area, chemical engineers are developing models to simulate the behavior of various materials under specific circumstances (Zhang & Friedrich, 2003). Hence, civil engineering is not an exception for problem-solving utilizing AI.

Several studies show that utilizing AI in civil engineering for sustainable development has increased drastically in recent years. A recent review study analyzed 105 publications between the years 1995 to 2021 and demonstrated a significant trend in

utilizing AI in civil engineering. Figure 15 depicts the trend during the recent 26 years (Manzoor, Othman, Durdyev, Ismail, & Wahab, 2021).

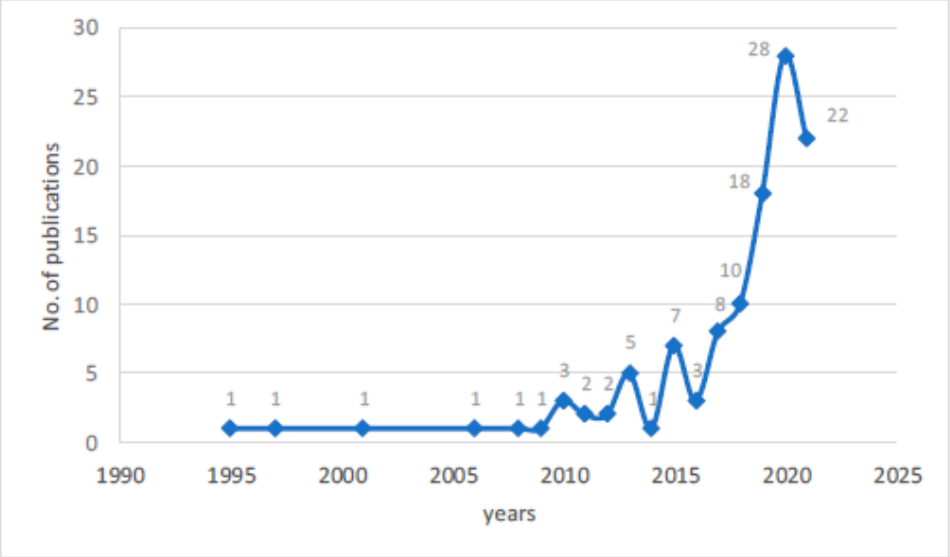


Figure 15: Annual publication trend for use of AI in civil engineering (Manzoor, Othman, Durdyev, Ismail, & Wahab, 2021)

In 2019, the Journal of Advances in Civil Engineering conducted a special edition on advanced AI technologies in civil engineering (Dede, Kankal, Vosoughi, Grzywiński, & Kripka, 2019). The state-of-the-art applications of AI introduced in this edition were related to several areas of study, such as structural engineering, construction management, hydrology, transportation, pavement engineering, geotechnical engineering, and hydrology. For instance, a research team studies the use of AI in predicting the construction time in an early phase. Another study in the transportation area was conducted on the use of AI to optimize headways and departure times in urban bus networks. In the structural area, a research team studies the damage detection methods utilizing the stress and stiffness of trusses using hybrid statistics and AI modeling.

Overall, the investigation of the use of AI in civil society is increasing at a high pace. Therefore, it could be realized that the use of AI has beneficial potential for problem-solving in this field of study.

In order to utilize AI technology and its recent advancement, it is crucial to realize the subset of this technology and the variations and different aspects of it. Figure 16 demonstrates the subsets of AI and a description of each subset.

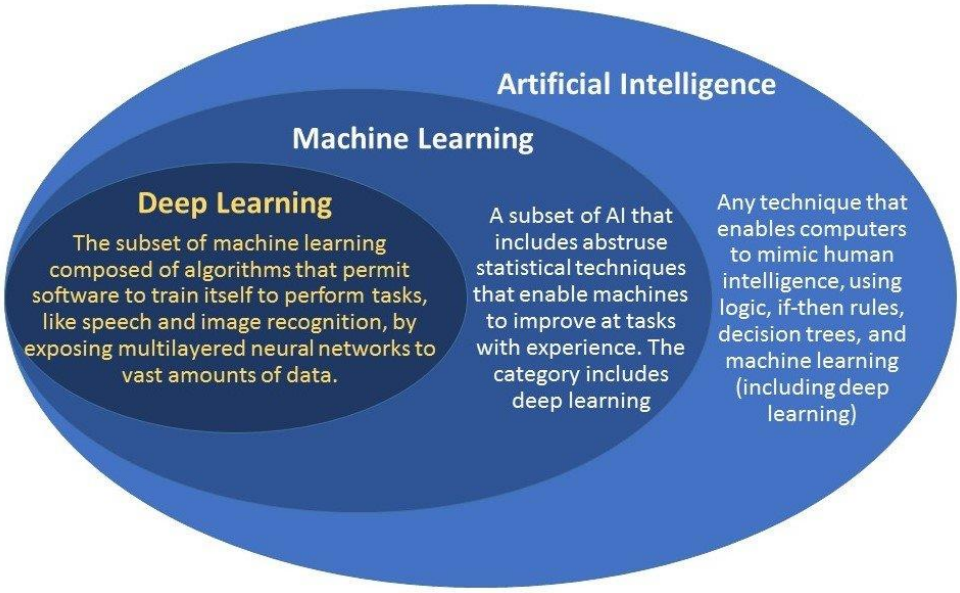


Figure 16: Virtualizing AI technology and the subsets (Dhande, 2020)

As illustrated in figure 16, AI technology has two main subsets; Machine Learning (ML) represents the statistical techniques which enhance the machine to improve the task-solving ability. Moreover, Deep Learning (DL) is the subset of ML which enables the machine to create algorithms to train itself to increase its performance on various tasks (Ji, Alfarraj, & Tolba, 2020). It is worth mentioning that the ability to conduct image processing and analyze huge amounts of data is a significant advantage of DL.

3.2 Machine Learning in Pavement Engineering

As mentioned in the previous section, Machine Learning (ML) is a subset of AI which could be divided into two main categories, supervised learning, and unsupervised learning. The difference between these categories is based on the data used for machine learning. Supervised learning refers to the methods which use a set of data consisting of the actual data and the corresponding labels. On the other hand, unsupervised learning refers to the methods which use the data without labeling. Figure 17 depicts the differences between these two groups and the applications of each one.

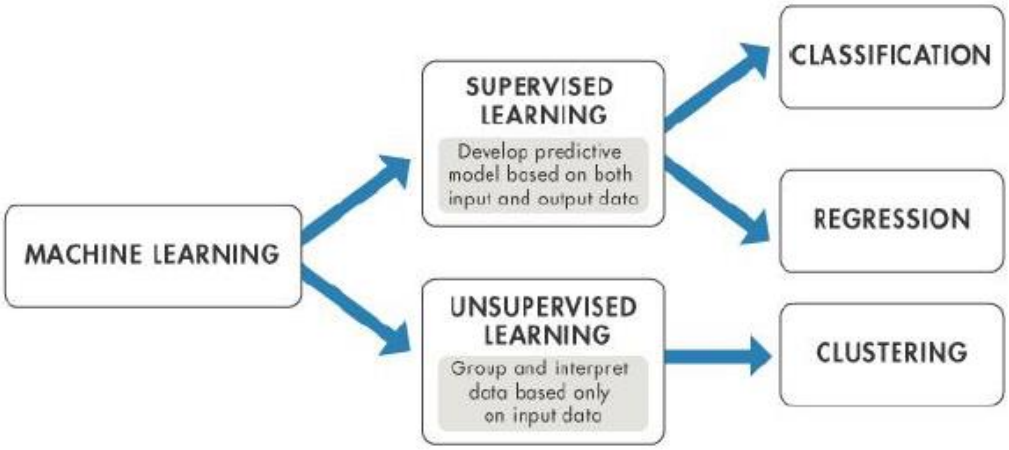


Figure 17: Supervised versus unsupervised learning (Bunker & Thabtah, 2019)

It can be seen that supervised learning is based on the input data, which feeds to the model and corresponding output data. Nevertheless, unsupervised learning is lonely based on input data without any output.

Moreover, it is shown that the application of supervised learning is divided into two main categories, which are regression and classification. Also, unsupervised learning

utilizes for solving the clustering problems. Both supervised and unsupervised learning have wide usage in transportation and pavement engineering field of study.

One of the most critical usages of ML in pavement engineering, specifically in Pavement Management Systems (PMS), is the performance prediction of the pavement for the rest of its lifetime (Abd El-Hakim & El-Badawy, 2013). This prediction assists authorities and agencies in making the proper decision for maintenance and rehabilitation strategies or making decisions for constructing or demolishing roads and highways. For example, it is possible to develop a model based on ML algorithms for predicting the International Roughness Index (IRI) based on the data such as weather conditions and traffic load for a particular pavement section (Hossain, Gopisetti, & Miah, 2017).

During recent years, many studies have been conducted on the use of various ML algorithms to predict the pavement's performance. A recent study (Marcelino, de Lurdes Antunes, Fortunato, & Gomes, 2019) developed a pavement performance prediction model based on the random forest algorithm (one of the famous algorithms in ML) to predict the condition of the pavement for 5 to 10 years. This study leveraged the data collected by Long Term Pavement Performance (LTPP) program (FHWA-HRT-15-049, 2015) as input and output data for their analysis. Their model was able to predict the future IRI value of a specific road section with high accuracy.

Another study (Issa, Samaneh, & Ghanim, 2022) was conducted on predicting the Pavement Condition Index (PCI) based on the distresses and their severity using the Artificial Neural Network (ANN) algorithm. The PCI is one of the most critical and significant indices for evaluating the performance of the pavement. Acquiring this

index is time-consuming and labor-intensive. This study leveraged the ML-based algorithm to develop a model to predict PCI value for a specific road section in Palestine based on the previously recorded data. After analyzing the model, they achieved high accuracy in terms of R^2 for correct predictions.

Overall, utilizing ML-based problem-solving strategies in pavement engineering, specifically in PMS, is becoming increasingly popular among scholars and researchers. However, further investigation and studies are still required to achieve a general, reliable, and robust model for pavement performance predictions. Despite the wide use of ML in pavement engineering, many researchers are attempting to leverage more facilities of AI Hence, many scholars focus on Deep Learning (DL) technology to tackle pavement asset management problems. DL is one of the most prominent technologies in computer science, which is becoming popular among other researchers in other fields of study. The following section demonstrates more about utilizing this technology in pavement asset management.

3.3 Deep learning in Pavement Condition Assessment

Despite the high performance of the ML techniques on numerical data for solving regression and classification problems and prediction, this technique has some deficiencies in handling image-based data. Popular ML algorithms such as Support Vector Machine (SVM), Decision Trees (DT), and XGboost struggle to handle huge image data. These algorithms could easily involve noises to deal with large amounts of data in image-based datasets. Hence, DL technology was invented to deal with more complex and sophisticated data. This technology enables the system to extract features based on heavy computational algorithms. Creating this technology relies on two main factors: the first in advanced technologies in computer science, and the second is the

availability of large datasets for analysis. During recent years, DL technology has been utilized in various fields of study, such as object detection and localization, voice recognition, and Natural Language Processing (NLP).

Due to the recent advancement of DL technology, many scholars and researchers in the pavement engineering field benefit from utilizing this technology in various areas of study, such as road network condition evaluation, pavement performance prediction, and pavement distress detection.

Utilizing DL technology in pavement condition assessment could be divided into two main categories. First, region level distress detection. Second, pixel level distress detection. These two categories will be identified and discussed in the following sections.

3.3.1 Region Level Distress Detection

In order to evaluate the performance of the pavement, there are several indices, such as IRI and PCI. To obtain the PCI value for a specific road section, several factors should be considered. One of the most key factors is the number of several distress types that occurred on the pavement surface, and the second is evaluating the severity of those particular distresses. Regarding the recording of the type of distresses and the frequency of occurrence, region-level distress detection should be utilized. On the other hand, in order to derive the severity of the distresses, pixel-level detection should be used. This section focuses on the recent prominent studies on region-level distress diction.

A recent study on automatic pavement distress detection using DL technology (Arya et al., 2021) utilizes a large image dataset of road cracks in Japan, India, and the Czech

Republic. The dataset was acquired using smartphone cameras installed on the dashboard of a car. They recorded 26,620 road images containing several distress types of the pavement. Figure 18 illustrates the position of the camera and the captured images from the road surface.

Figure 18: Road images of (Arya et al., 2021) study

After collecting the dataset, they annotated the images based on four different distress types: alligator cracks, transverse cracks, longitudinal cracks, and potholes. In order to develop and train a DL model, they applied three different algorithms, YOLOv5 (You Only Look Once), YOLOv4, and RCNN (Region-based Convolutional Neural Network). Among these algorithms, the highest accuracy belonged to the YOLOv5 model by 0.67 in terms of F1 score. Although the model was trained with a large

number of images, the accuracy performance of the test data was not high enough. Moreover, they eliminated other types of distresses, such as block cracking and patching.

In another study related to pavement distress detection and classification (Du et al., 2020), a research team prepared a large-scale high-resolution dataset of road distress using an industrial camera installed on a vehicle. The dataset consisted of 45,788 images, including 59,366 distress instances. The distress instances were alligator, longitudinal, and transverse cracking, as well as patching and manholes. In order to develop a DL model, they decided to use the YOLOv3 algorithm for training. Also, they compared this algorithm with R-CNN and SSD (Single Shot Detection) in terms of speed. After training the model, they reported accuracy of 0.73 in terms of the F1 score. Figure 19 depicts some of the sample images for detecting the distresses. Despite training with a large number of images, the accuracy metric does not show high performance for the testing phase. Moreover, capturing these images with high resolution requires industrial cameras, which could be expensive for road agencies to purchase.

Figure 19: Crack detection examples of (Du et al., 2020) study

3.3.2 Pixel Level Distress Detection

In the previous section, Region level pavement distress detection had been discussed. This section will demonstrate pixel-level detection, and some of the most recent prominent studies will be discussed.

In addition to recording and collecting different distress types of a specific pavement section, it is essential to evaluate the severity of these distresses. In this case, utilizing semantic segmentation enables the researchers to automatically record and collect each

distress type's severity automatically using ordinary images. For this purpose, many studies have been conducted to investigate the possibility of using DL technology to develop a robust and reliable model for acquiring the severity of the distresses.

In 2020 a research team from Canada conducted research to automatically collect the severity of cracks on several roads in Ontario (Mei & Gül, 2020). In order to collect the data, they utilized a GoPro camera installed on the rear side of a vehicle to capture the road images. Choosing this approach lets them capture images in a wide view without any noises such as wiper and specks of dirt on the car's mirror. Figure 20 illustrates the schematic of the data acquisition approach of their study.

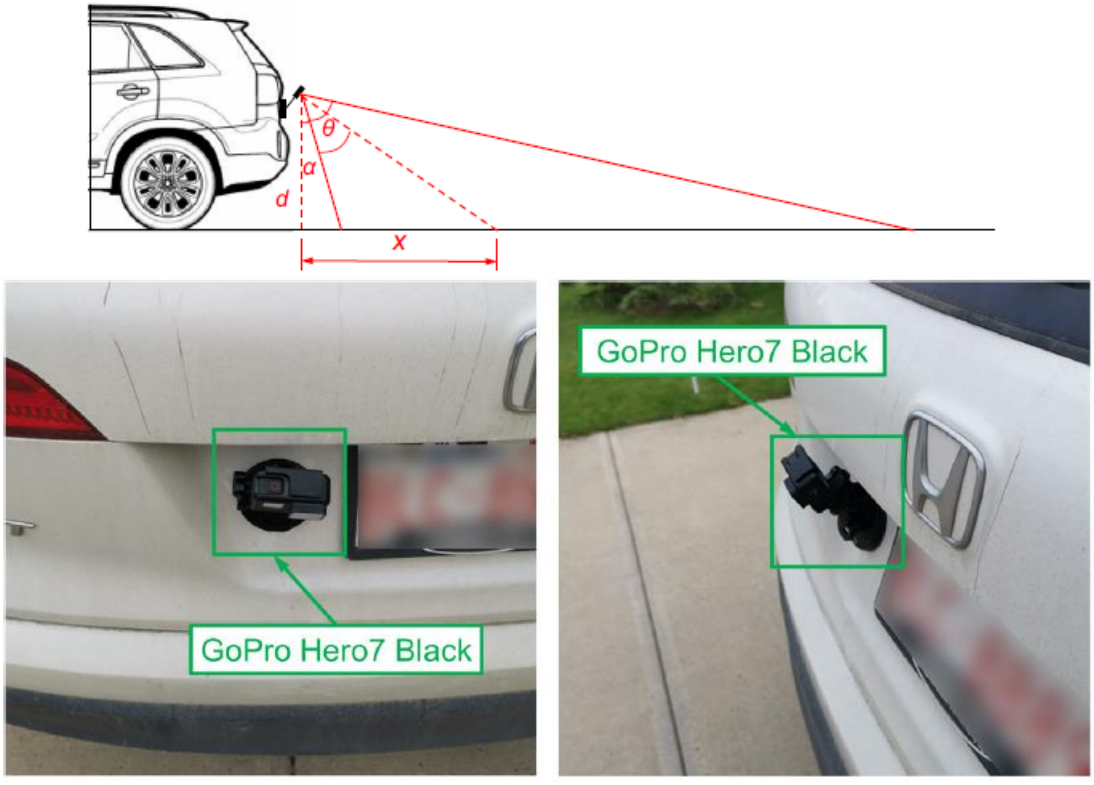


Figure 20: Data collection approach in (Mei & Gül, 2020) study

After collecting the data and annotating the cracks, they used a composite model consisting of 121 dense layers and a 5-layer convolution network called ConnCrack, and they reported the accuracy based on the F1 score, which was 0.91.

In another study, a research team from China (Ji, Xue, Wang, Luo, & Xue, 2020) conducted a study on automatically obtaining the properties of the cracks such as width, depth, and length using DL technology. In this study, they leveraged DeepLabV3+ architecture for training the network and developing the model. After training, they reported the model's accuracy based on Mean Intersection over Union (MIoU); they reached 0.83 for testing 50 images and 0.733 for testing 80 images. Moreover, they compared the accuracy with other algorithms such as FCN and DeepCrack, which had lower accuracy. Figure 21 depicts an example of analyzing the cracks using their model.

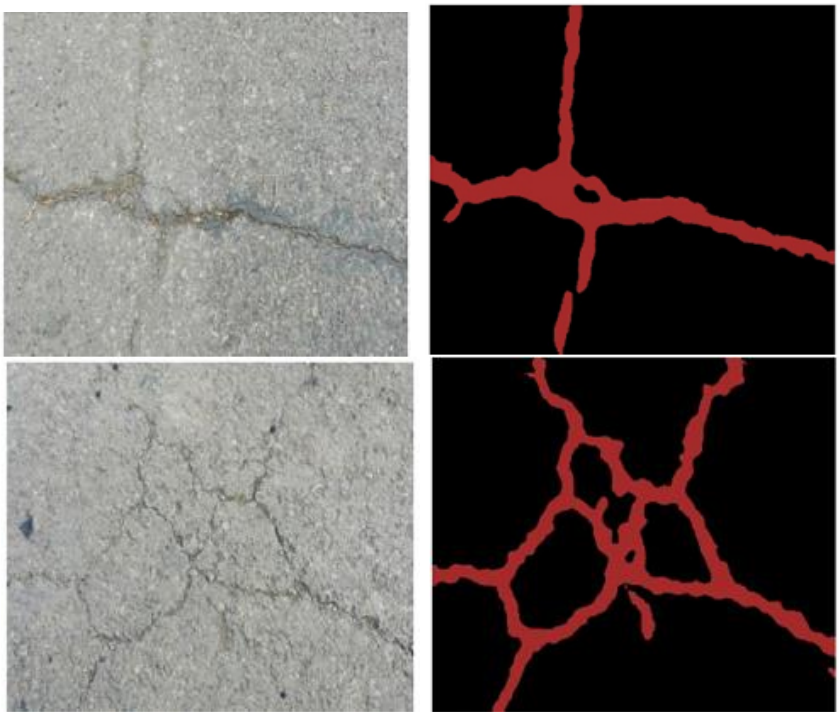


Figure 21: Example of analyzing the cracks in (Ji & Xue, 2020) study

Chapter 4

METHODOLOGY

The research gaps have been identified based on an extensive literature review demonstrated in the previous chapter. This study aims to fill these gaps and covers the research objectives mentioned in the first Chapter. For this purpose, a structured methodology has been proposed, which is illustrated in the following flowchart. Figure 22 depicts the proposed methodology structure.

Despite previous studies, this study concentrates more on data-centric strategies and methods than model-centric approaches to improve the accuracy of the model. First, the data annotation proposed several points for improving the consistency of the labels. Then the dataset will be analyzed to improve the decision-making for data augmentation strategies. Finally, the model will be trained using appropriate data augmentation methods to reach acceptable precision.

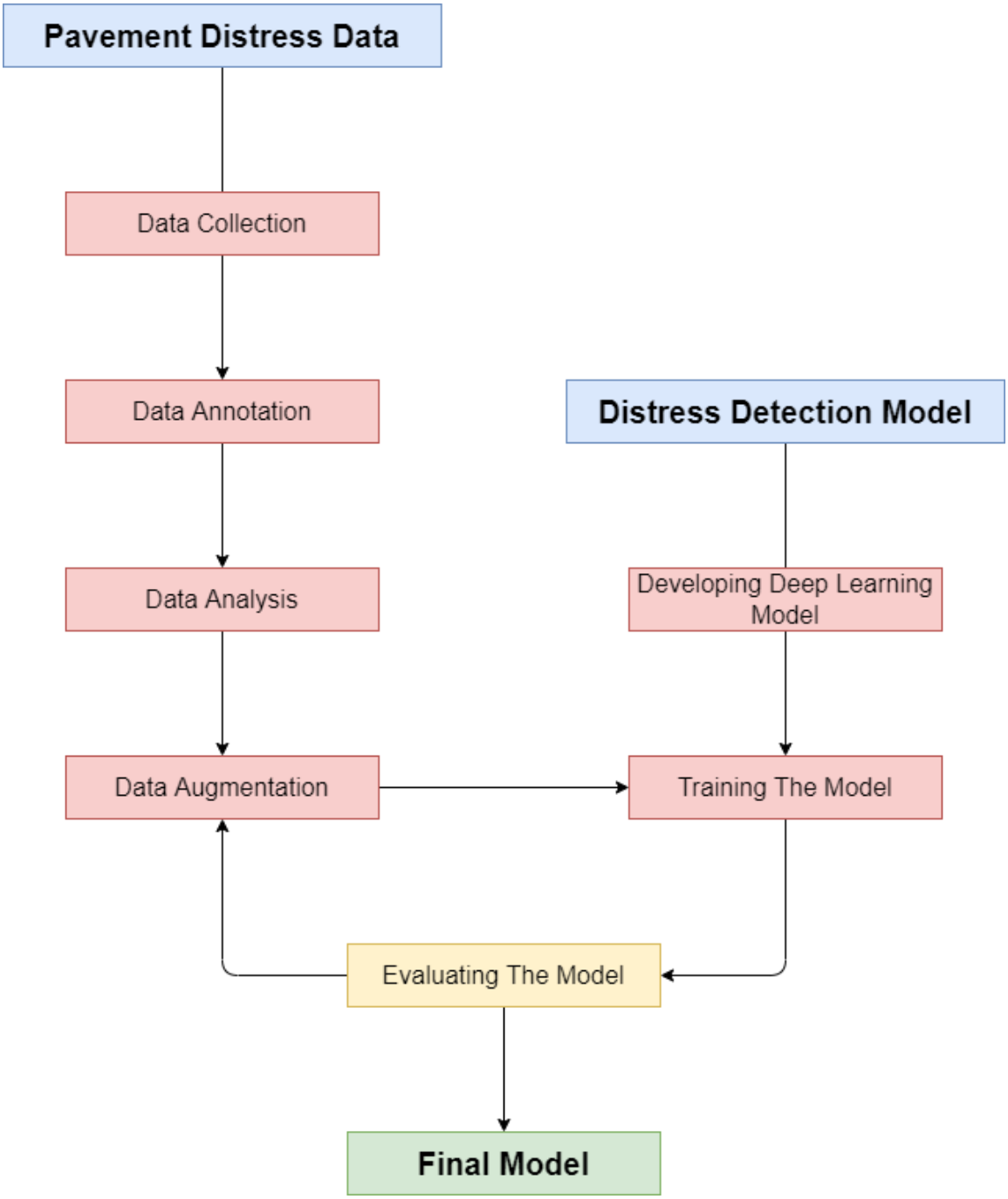


Figure 22: Methodology flowchart

4.1 Pavement Distress Data

4.1.1 Data Collection

One of the most critical issues in developing a distress detection model is acquiring sufficient data for training the model. This study utilizes a dataset collected from several cities in the United States of America. These cities are Kansas City, Jefferson

City, and Columbia in Missouri (Majidifard, Jin, Adu-Gyamfi, & Buttlar, 2020). The images were collected using two different methods. Seventy percent of the images were captured by ARAN vehicles, and the rest were collected from Google Street View API. The resolution of the collected images is 1080 by 960.

The dataset contains 628 top-down view road images that cover a single lane of the road section with different types of distresses and severity levels as well as some noise such as tree shadows and oil strains on the roads. This dataset has seven distress types: alligator cracking, transverse cracking, longitudinal cracking, block cracking, patching, sealing, and manhole. Figure 23 depicts some examples of these images.

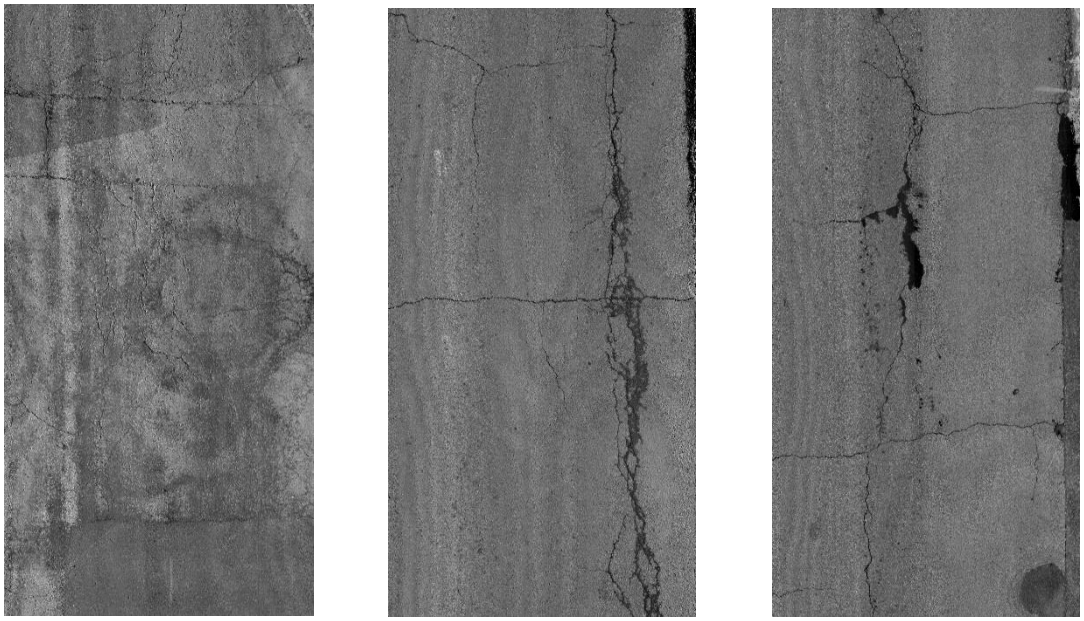


Figure 23: Sample images of the proposed dataset

4.1.2 Data Annotation

In order to use the images for training the model, the first step is to annotate the images. For this purpose, several bounding boxes should be drawn to identify a specific distress type and its corresponding location in each image. This study utilized the CVAT

annotation tool for annotating the images. This tool allows users to draw bounding boxes in various shapes and styles. Then it is possible to extract the annotations in the various format required for different algorithms. The available annotation formats are JSON, XML, and COCO, which are compatible with different algorithms. This study extracted the COCO format of the annotated images. Figure 24 shows the environment of this application.

Figure 24: CVAT annotation tool environment

It can be seen that the environment is user-friendly, containing various tools for drawing the bounding boxes. On the left-hand side, it allows users to select different shapes of the annotation, such as rectangles and polygons, and on the right side, it allows users to select the correct label for each bounding box.

One of the most critical issues in annotating the images is that the bounding boxes should be drawn precisely. It means that the bounding boxes should be drawn close to the distresses. In this way, it pretends the model to get confused the target object with the background. This approach empowers the model to distinguish between the objects

and the background. Figure 25 depicts some examples of the road images containing the bounding boxes.

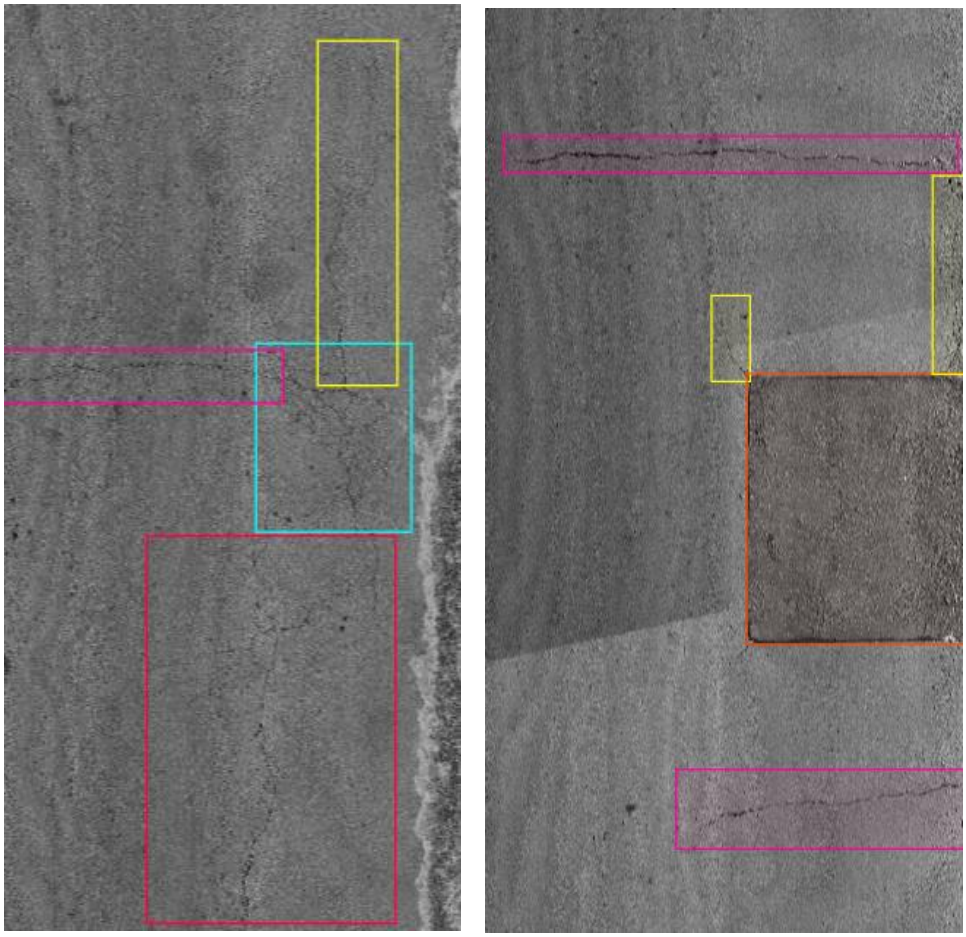


Figure 25: Distress images with bounding boxes of the proposed dataset

It can be seen that a specific bounding box distinguishes each distress on the surface of the road. Also, each color of the bounding boxes represents a particular type of distress. For instance, the pink color shows the transverse cracking, and the yellow color shows the longitudinal cracking.

4.1.3 Data Analysis

Before making decisions on choosing appropriate strategies for data augmentation, it is essential to analyze the dataset and realize the consistency of the images and

annotations and make a brief understanding of the data. This section aims to analyze the entire dataset and its labels.

First of all, it is crucial to analyze the number of distresses for the entire dataset and realize the distribution of each distress type. For this purpose, a bar chart had been prepared. Figure 26 depicts this bar chart.

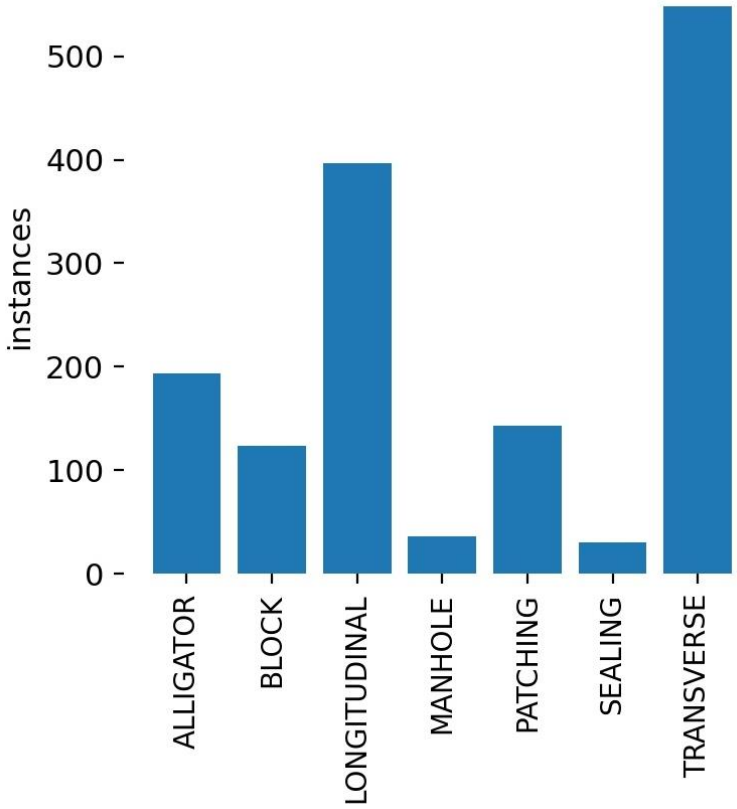


Figure 26: Distribution of the distresses in proposed dataset

The above chart shows that the highest number of instances belongs to transverse cracking by more than 500 instances. Then the longitudinal cracking by the number of instances of around 400. on the other hand, the lowest number of instances belongs to sealing with lower than 40 instances. After that manhole has around 30 instances, and the other distresses have a number of instances between 100 and 200.

It can be realized that longitudinal and transverse cracking with the highest number of instances in the entire dataset could easily be recognized by the model. Nevertheless, the model might not be able to learn the features of sealing and manhole, which has the lowest number of instances. This type of information about the dataset benefits the decision-making on selecting appropriate strategies for data augmentation.

Another fundamental analysis for the dataset is heat map evaluation. In this case, it is possible to realize the location of occurrence of the distresses throughout the entire dataset. In order to create the heat map figure for each distress in the dataset, this study leveraged the Roboflow application. This application allows users to conduct a heatmap analysis based on the images and the location of the corresponding annotations for each label. Figure 27 illustrates these heat maps. It can be seen that most of the longitudinal cracks are located on the edges of the images, which demonstrates that in order to choose a cropping strategy for data augmentation, if it crops the center of the images, many of the longitudinal cracks will be removed.

On the other hand, transverse cracking occurs more in the center of the image than at the edges of the image. Also, patching has the same distribution. Regarding other types of distresses, it could be seen that they are distributed almost over the entire image. In fact, this analysis confirms that in order to select a cropping strategy for data augmentation, the cropping should be based on the occurrence of the damages. Otherwise, the cropping strategy might not be efficient for this dataset.

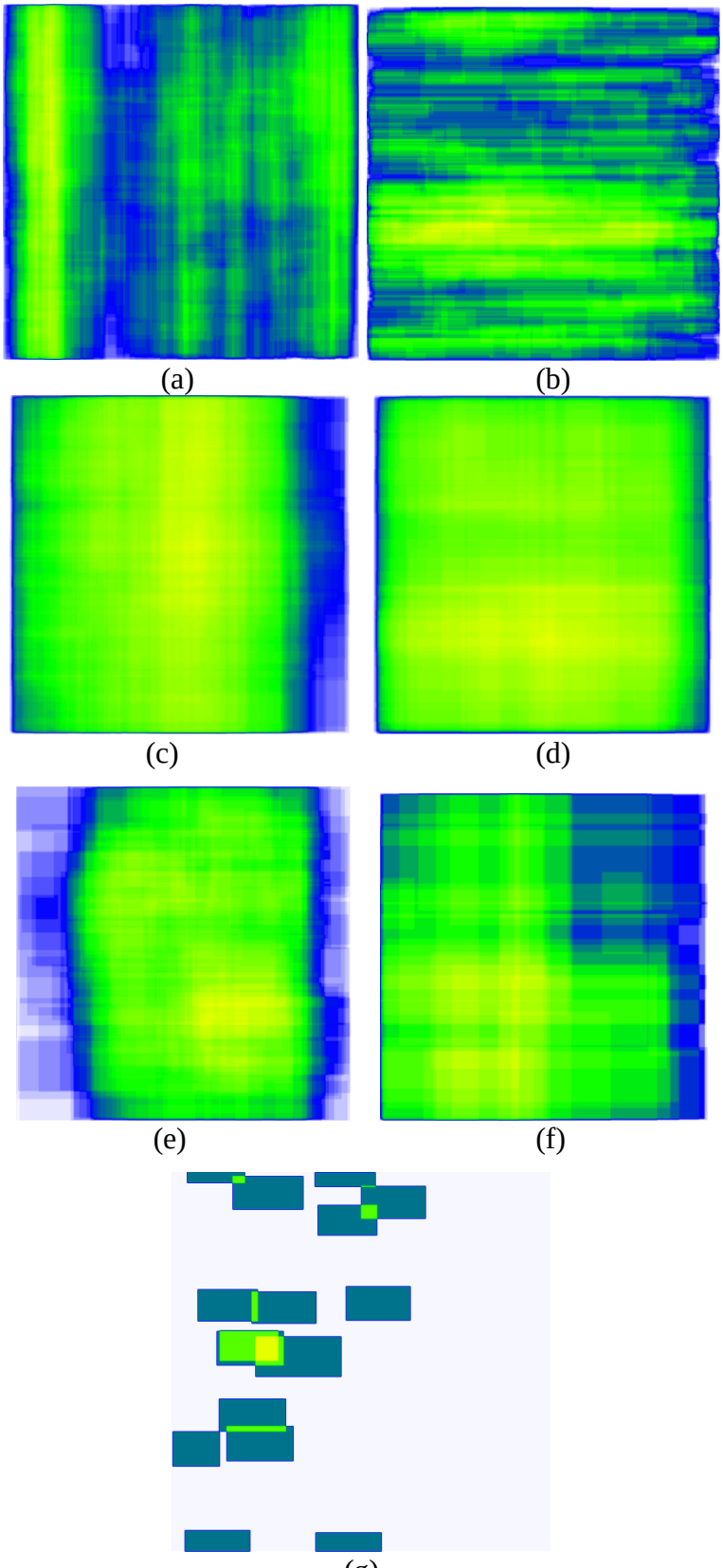


Figure 27: Heat map analysis of the annotations of proposed dataset. (a) longitudinal (b) transverse (c) alligator (d) block (e) patching (f) sealing (g) manhole.

Another vital analysis of the dataset is understanding the sizes of the bounding boxes. This type of analysis allows the researcher to choose the scaling strategy for augmenting the dataset. If the dataset suffers from a wide variety of sizes of objects, the scaling strategy should be considered the primary strategy. Figure 28 shows the sizes of the bounding boxes, which represent the volume of the distresses on the entire dataset.

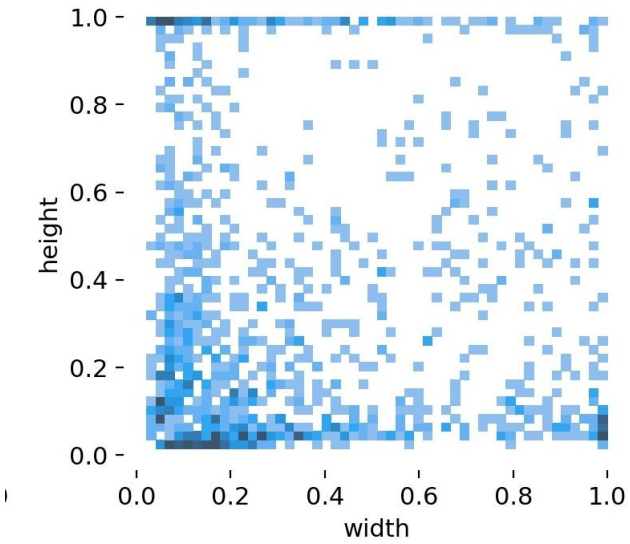


Figure 28: Distress sizes on the entire proposed dataset

In figure 28, the blue points show each distress with corresponding height and width. The sizes of each bounding boxes scaled between 0 to 1. The lighter blue shows a low density of points. Visa versa, the darker blue shows the high density of the points in the figure.

The density of the point on the lower left side of the plot demonstrates that most distress contains low, high, and low width. However, some high-density areas on the

bottom right and the upper left of the plot shows some distresses with high width and low height (transverse cracks) and high height and low width (longitudinal cracks).

4.1.4 Data Augmentation

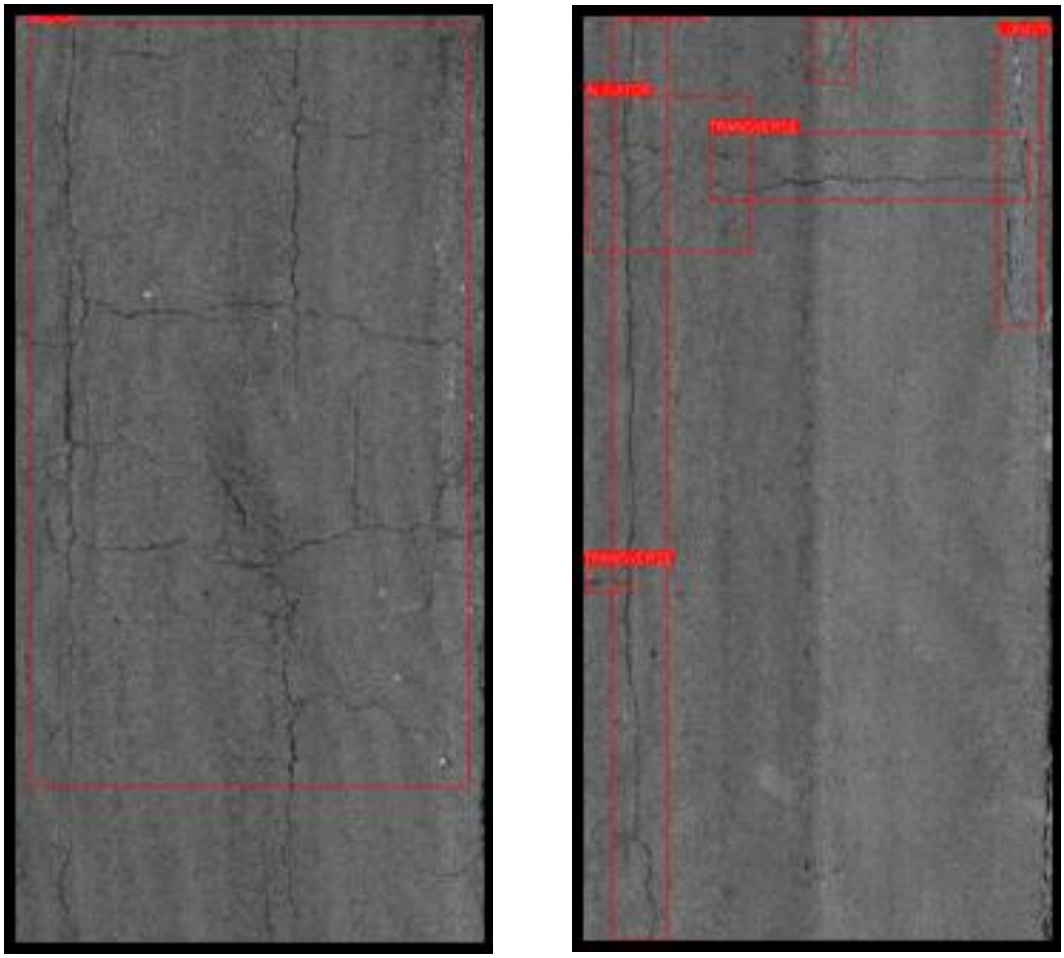
One of the most critical challenges in developing a deep learning model is overfitting. Overfitting refers to the situation in which the model is memorizing the data instead of learning the features. The model could suffer from overfitting when there is not a sufficient number of images, and the model cannot extract the necessary features. In order to overcome this problem, data augmentation should be applied. By increasing the number of images in the dataset, it is possible to decrease the effect of overfitting. However, it is essential to choose an appropriate data augmentation method which is compatible with the images in the dataset. There is no rule of thumb for selecting data augmentation methods for a specific dataset. The performance of applying each data augmentation method might differ for each particular dataset.

In order to apply data augmentation methods for pavement images, it is essential to evaluate the dataset beforehand. Some sort of augmentation method could harm the accuracy result. For instance, choosing the rotating augmentation method could confuse the model in distinguishing longitudinal cracking and transverse cracking. The shear method could have the same effect. In contrast, flipping left to right or up to down could be appropriate.

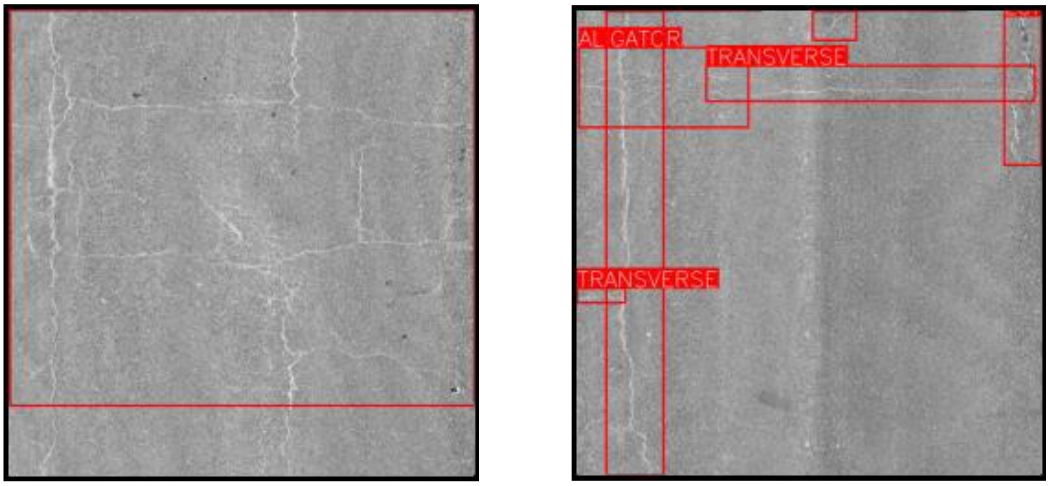
In this study, two types of augmentation methods were applied:

- Bounding Box Safe Crop
- Image Inversion.

Based on the analysis conducted in the previous section, it could be realized that these two types of augmentations are the appropriate choice for the proposed dataset. Bounding box safe crop, make new images by cropping the existing ones in a way that it does not lose any bounding boxes in each image. Furthermore, the image inversion method was applied to the dataset for creating new images. Figure 29 depicts the effect of applying these two augmentation methods on the dataset's images.



(a)



(b)

Figure 29: Example of augmented images of the proposed dataset: (a) actual image (b) augmented image

4.2 Distress Detection Model

Recent developments in deep learning, precisely computer vision, resulted in various deep convolutional neural network architectures that perform accurately and fast on different data types and are utilized in a wide range of scientific and industrial areas.

This study benefited deep learning technology for developing a model to detect and classify different distresses on the surface of the roads. One of the state-of-the-art object detection and classification algorithms is the YOLO (You Only Look Once) algorithm. This algorithm is popular among researchers in the data science domain due to its accuracy and speed. This study utilizes the fifth version of this algorithm for developing a distress detection model.

4.2.1 History of YOLO Algorithm

In 2015, the YOLO (You Only Look Once) algorithm evolved object detection tasks in computer vision, outperforming the other models in accuracy and detection speed (Redmon, Divvala, Girshick, & Farhadi, 2016). This algorithm divides each image into a specified number of grid cells, and each cell is responsible for detecting objects

within itself. This method allows the algorithm to detect all the objects and their locations in a single glance. This mechanism helps the model improve inference time drastically and become a real-time object detection model.

This algorithm has three main steps for detecting the objects. First, it resizes the input images to 448 by 448 and prepares them to feed the network. Then, the images run through a deep convolutional neural network to extract the features of each class of objects. After that, each specific object will be detected and classified using non-max suppression mechanisms. Figure 30 shows the initial architecture of the YOLO algorithm.

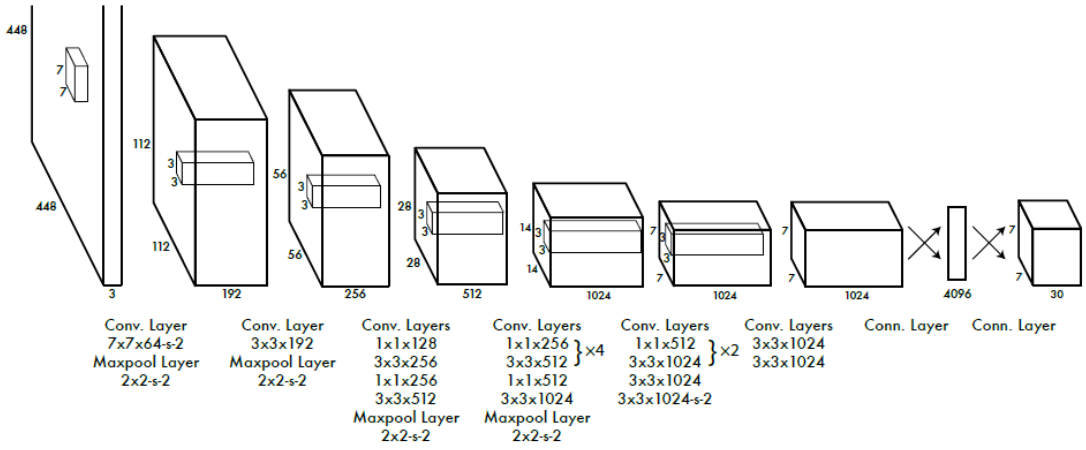


Figure 30: Initial architecture of YOLO algorithm (Redmon, Divvala, Girshick, & Farhadi, 2016)

Figure 30 demonstrates that the initial YOLO algorithm contains 24 convolutional layers followed by two fully connected layers. Each convolutional layer is responsible for extracting a specific feature of the image. The maxpooling layer enables the network to focus on a particular part of the image.

After a year, (Redmon & Farhadi, 2017) introduced the second version of the YOLO (YOLOv2) algorithm. In addition to optimizing the model's architecture, they applied several functions such as batch normalization and scaling to increase the accuracy and inference speed of the algorithm. Figure 31 compares the performance of the YOLOv2 algorithm with other prominent object detection algorithms.

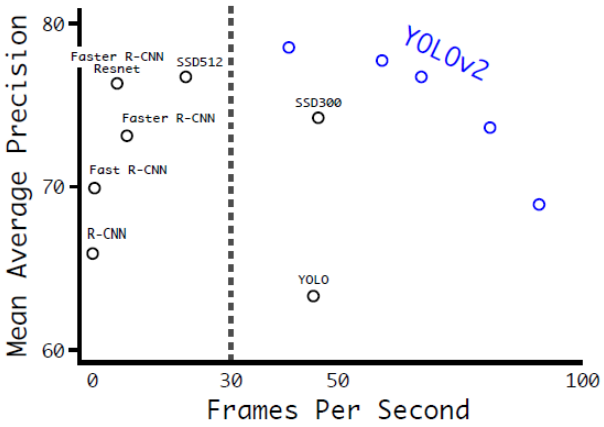


Figure 31: Performance comparison of YOLOv2 with other algorithms (Redmon & Farhadi, 2017)

It can be seen that the performance of the YOLOv2 algorithm outstands among other algorithms such as Faster R-CNN, SSD512, and Resnet in terms of both speed and accuracy measured by mean average precision. It should be mentioned that the reported accuracy and speed of the algorithms had measured on the Pascal VOC 2007 dataset, which contains thousands of images, including real-life images.

After two years, (Redmon & Farhadi, 2018) published the third version of the algorithm (YOLOv3). In order to optimize their proposed algorithm, they made significant modifications to the architecture of the algorithm. They added the Darknet-53 backbone to improve the accuracy of the model. Figure 32 illustrates the performance of the YOLOv3 algorithm compared to the other models. It can be seen

that the YOLOv3 model outperforms the RetinaNet-50 and RetinaNet-101 in terms of both accuracy and inference time. It should be mentioned that the performance of the model tested on the COCO dataset.

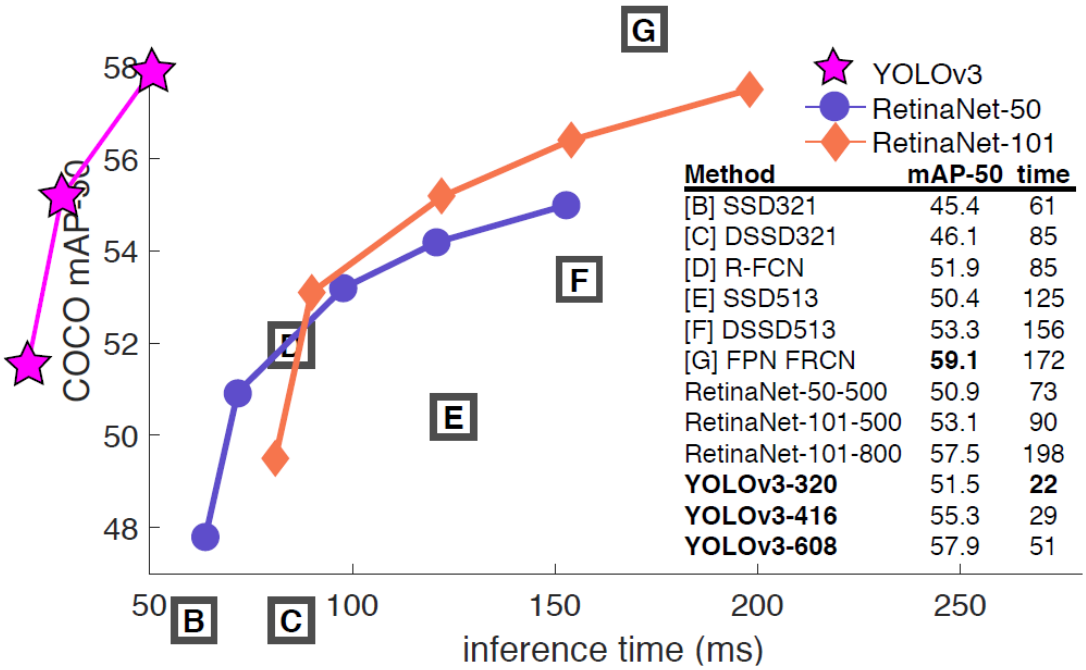


Figure 32: Performance of YOLOv3 algorithm (Redmon & Farhadi, 2018)

After two years, another research team introduced the YOLOv4 algorithm (Bochkovskiy, Wang, & Liao, 2020). They added several functions to the architecture to improve the model's accuracy. These functions are Weighted-Residual-Connections (WRC), Cross-Stage-Partial-connections (CSP), Cross mini-Batch Normalization (CmBN), Self-adversarial-training (SAT), and Mish-activation. Moreover, they leveraged the Mosaic augmentation strategy to reduce the overfitting effect of the proposed model. Figure 33 shows the performance of the YOLOv4 algorithm compared to the other pioneer object detection algorithms. Although the accuracy of YOLOv4 is lower than EfficientDet, the inference time reached 120 frames per second, which could be considered a real-time object detector algorithm. Moreover, the

YOLOv4 outperforms YOLOv3 in both accuracy and inference time. The accuracy was recorded based on average precision, and the Microsoft COCO dataset was used (Lin et al., 2014).

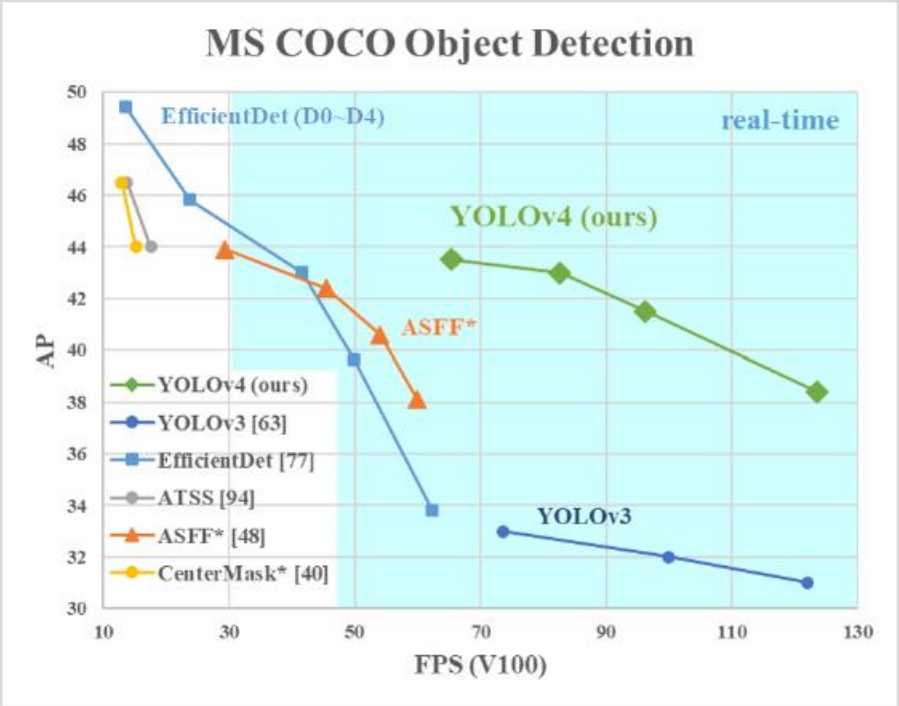


Figure 33: Performance of YOLOv4 algorithm (Bochkovskiy, Wang, & Liao, 2020)

4.2.2 YOLOv5

Up to now, YOLOv5 is the latest version of the YOLO algorithm introduced by Glen Jocher in 2020 (Jocher et al., 2020). YOLOv5 is one of the most accurate and fast algorithms among all object detection algorithms. This new version added several significant functions such as Upsampling and Focus to increase the robustness of the model. Additionally, the architecture of the algorithm was optimized by several modifications in the neck and head part of the model. This algorithm benefited from adaptive learning rates as well as adaptive anchor boxes to produce an efficient training process for various types of custom datasets.

Recently, the Ultralytics team implemented the YOLOv5 algorithm on the PyTorch framework using Python programming language, and the source code is publicly available on their GitHub page. This algorithm was developed in various versions to utilize different types of datasets. However, the architecture of all the versions is the same. Figure 34 depicts the different versions of the YOLOv5 algorithm and demonstrates the differences.

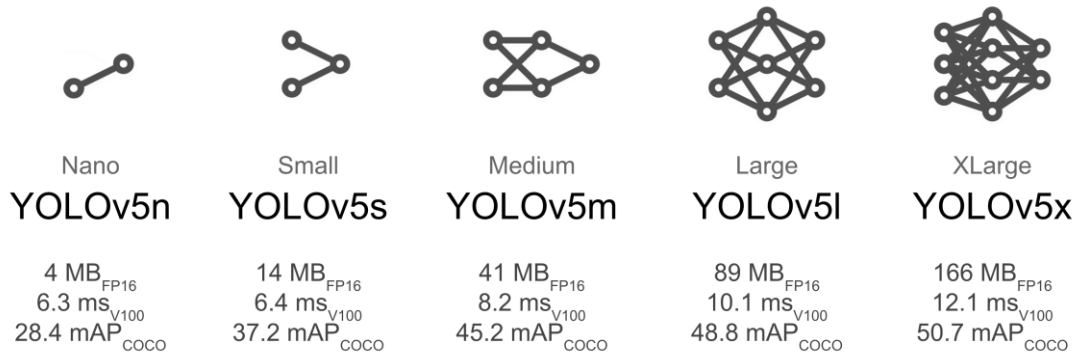


Figure 34: Different versions of the YOLOv5 algorithm (Jocher et al., 2020)

As discussed earlier, the architecture of all five versions, from Nano to XLarge, are the same. The difference is related to the size of the network for each version. In fact, from Nano to XLarge, the network's height and width will gradually increase by a prefixed factor. Figure 34 illustrates the performance metrics of each version on the COCO dataset. It can be seen that by increasing the size of the network, the corresponding accuracy increases. The performance of the Nano version on the COCO dataset was reported as 28.4 in terms of mean average precision (mAP). By enlarging the network to XLarge, it was observed that the accuracy increased to 50.7 mAP.

In contrast, by increasing the size of the network, the inference time will increase as well. It is shown that the inference speed of the Nano version on the abovementioned dataset is 6.3 millisecond, and the inference time of the largest network size is 12.1

milliseconds. This study utilized the XLarge version due to reach higher accuracy. In the following section, the architecture of this algorithm will be discussed.

4.2.3 Architecture of YOLOv5

The architecture of the YOLO algorithm has been improved gradually from 2015, the initial architecture to 2020, the latest architecture up to now, by making significant efforts of many research teams. This version of the YOLO algorithm benefited from recent prominent advancements in deep learning technology found in the recent years. The structure contains several enhancement functions for smooth training and preventing overfitting or convergence issues. Furthermore, many complex deep convolutional networks have been implemented to improve the feature extraction process. This section attempts to perform a brief demonstration of the YOLOv5 architecture, including various functions utilized in this algorithm and several convolutional layers and their functionality and effects.

The architecture of YOLOv5 consists of three main parts:

- Backbone
- Neck
- Head.

Each of these sections has several functions and layers inside. The Backbone extracts the features related to the overall shape and orientation of the objects. After that, the images pass through the Neck part to extract more sophisticated and narrow shapes of each class. Finally, the images pass through the Head part as the final part for detecting the objects and the location of their occurrence throughout the image. Figure 35 depicts the architecture of YOLOv5, including identifying the three main parts as well as the layers inside each section.

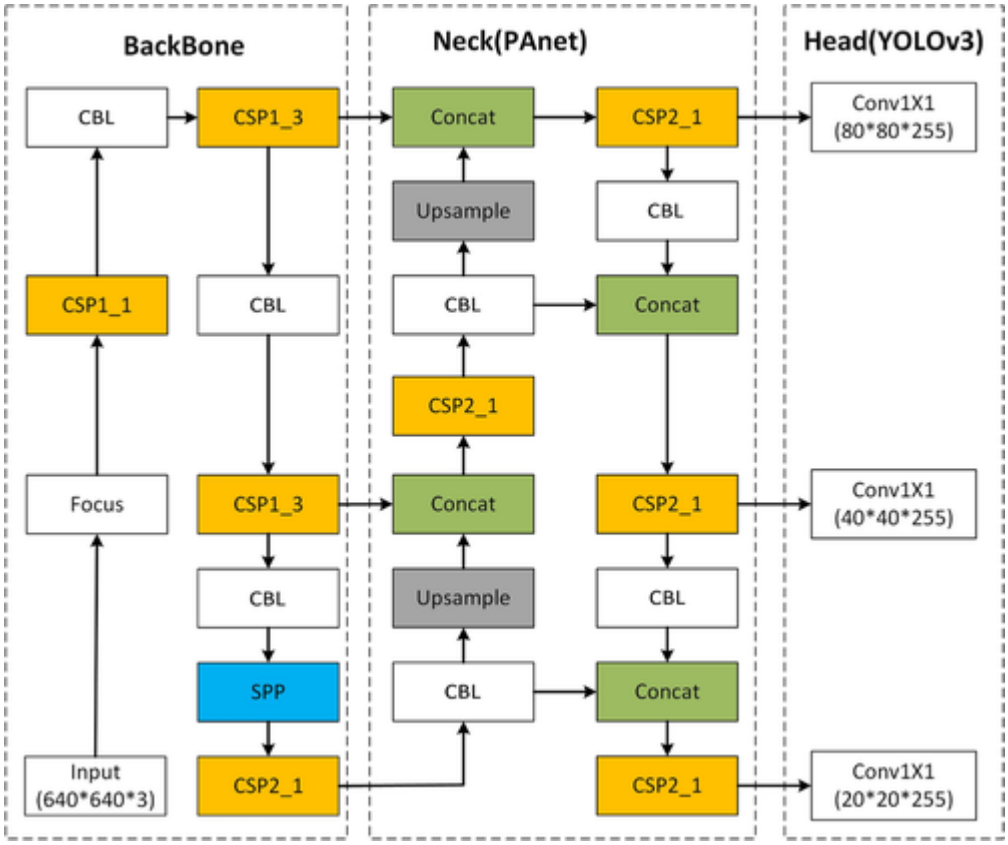


Figure 35: Architecture of YOLOv5 (Jia et al., 2021)

As illustrated in figure 35, the network consists of the Backbone, Neck, and Head. The first stage of the Backbone is the input size. It is shown that the input size of the images is 640 by 640 by 3, which means that the network is only compatible with this image size. Thus, in order to feed images to the network, it is essential to resize them to the abovementioned scale. The second stage is Focus. This stage prepares the images by reconstructing them in a low-resolution format. Another essential layer in the Backbone is CSPNet (Wang et al., 2020). In figure 35, CSP1_x represents the block containing the CBL block and x residual connection unit. Moreover, CSP2_x block contains the x CBL blocks. The CBL block represents a convolutional layer followed by a batch normalization and an activation function. Another function in this section is Spatial Pyramid Pooling (SPP) (He, Zhang, Ren, & Sun, 2015) which is applied after the third CBL block.

The architecture of the Neck part is derived from the idea of PANet (Wang, Liew, Zou, Zhou, & Feng, 2019) to improve the feature extraction of the images. It can be seen that this part includes some concatenation and upsampling functions as well. The Head part of the architecture is similar to the previous versions (YOLOv3 and YOLOv4). It applies anchor boxes on features and generates final output vectors with class probabilities, objectness scores, and bounding boxes, and it performs the final detection part. It is shown that the final detection part includes one by one convolutional layer in three different image sizes.

It is worth mentioning that the activation function used in this architecture is Sigmoid Linear Unit, and the optimization function is Stochastic Gradient Decent (SGD). Furthermore, the choice of loss function for this architecture is Binary Cross-Entropy with Logits Loss.

4.3 Evaluation Metrics

In order to evaluate the model, there are several metrics that are widely accepted among scholars and researchers of this domain. For evaluating the performance of the object detection and classification model, the following metrics have been provided. Equations 1 to 3 demonstrate the computational methods of these metrics.

$$\text{Precision} = \frac{True\ Positive}{True\ Positive + False\ Positive} \quad (1)$$

$$\text{Recall} = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (2)$$

$$\text{F1 Score} = 2\times \frac{Precision \times Recall}{Precision + Recall} \quad (3)$$

Precision indicates the portion of true positive prediction over the total positive prediction. On the other hand, recall indicates the portion of true positives over all actual positives. Finally, the F1 Score accumulates precision and recall equally by dividing multiplication over their summation.

Chapter 5

ANALYSIS AND RESULTS

5.1 Training Criteria

in order to train the model, there are several essential criteria which will be discussed in this section. The proposed dataset contains 628 images, including several road distress types such as alligator, transverse, longitudinal, and block cracking as well as sealing, patching, and manhole. Regarding preparing the dataset for training, the dataset is divided into two parts:

- Training set contains 80 percent of the dataset (505 images)
- Validation set contains 20 percent of the dataset (123 images).

All images in three sets are fully annotated leveraging the CVAT annotation tool platform and resized to 640 by 640 to ensure compatibility with the proposed model.

This study utilizes the YOLOv5 algorithm as a model for training the dataset. The largest version of the model was selected to obtain high accuracy. The algorithm was developed using the PyTorch framework (Paszke et al., 2019) scripted in Python programming language (Python 3.10). The GPU used for computational progress was GeForce RTX 3050. In order to tune the hyperparameters of the algorithm, several experiments were done, and the results were compared. The following hyperparameters represent the optimized values for each hyperparameter.

- initial learning rate (0.01)
- optimizer weight decay (0.0005)
- bounding box loss gain (0.01)
- class loss gain (0.1)
- object loss gain (0.1)
- image HSV-Hue augmentation (0.015) (fraction)
- image HSV-Saturation augmentation (0.7) (fraction)
- image HSV-Value augmentation (0.4) (fraction)
- image rotation (0.0)
- image scale (0.2)
- image shear (0.0)
- image flip up-down (0.5) (probability)
- image flip left-right (0.5) (probability)
- image mosaic (1) (probability)
- bounding box safe crop (0.5) (probability)
- image inversion (0.5) (probability)
- image mix-up (0)
- optimizer (Stochastic Gradient Descent)
- activation function (Sigmoid Linear Unit).

In order to train the model, pretrain weights were applied as the initial weight for the training. These weights were trained with the COCO dataset for 1000 epochs.

Regarding the training, the number of batch size defied as 16, and the model trained for 200 epochs. The computational time for the entire training process was

approximately 3 hours. Figure 36 depicts a sample of the training batch for the training process.

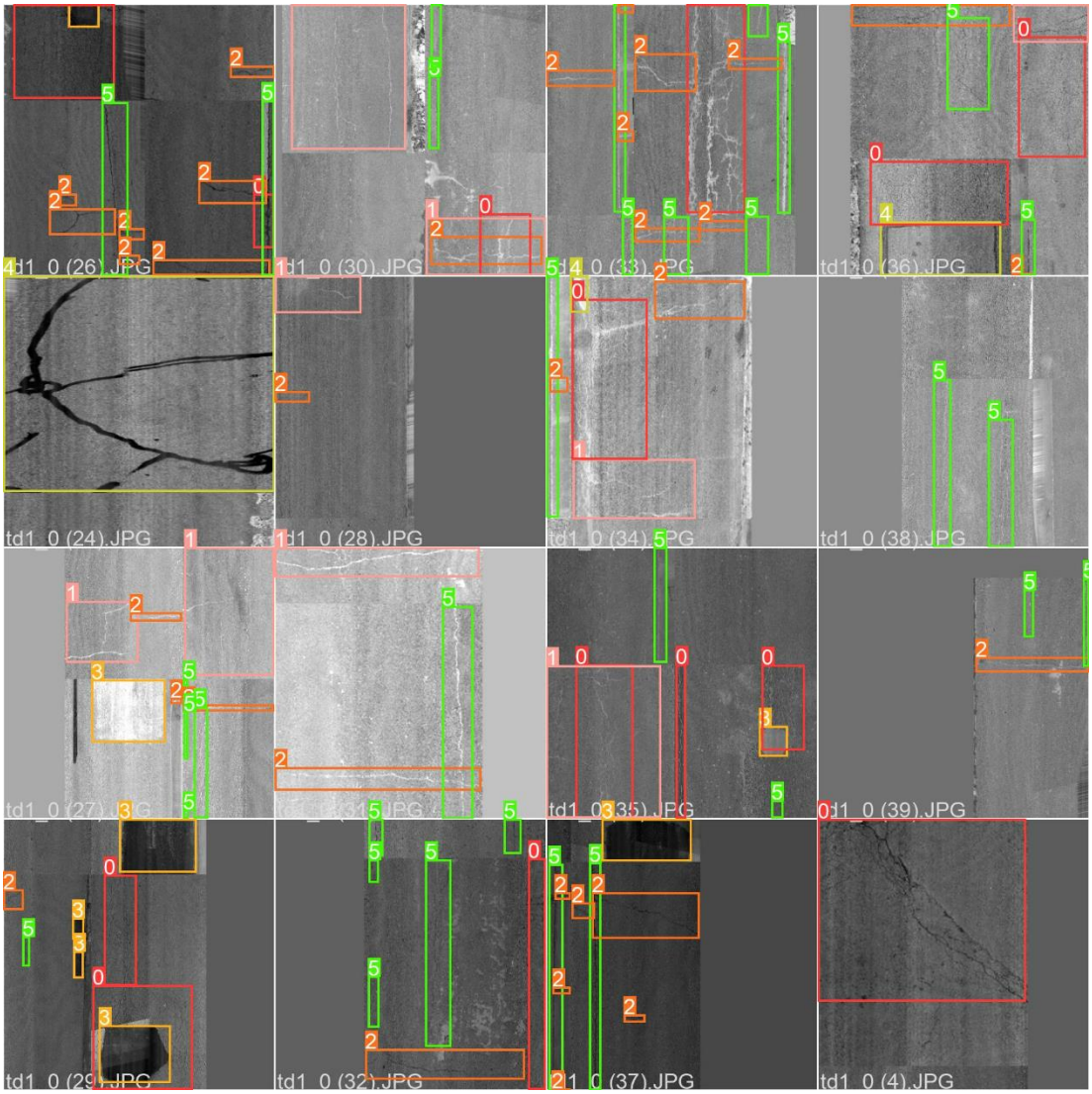


Figure 36: Sample batch of the training phase

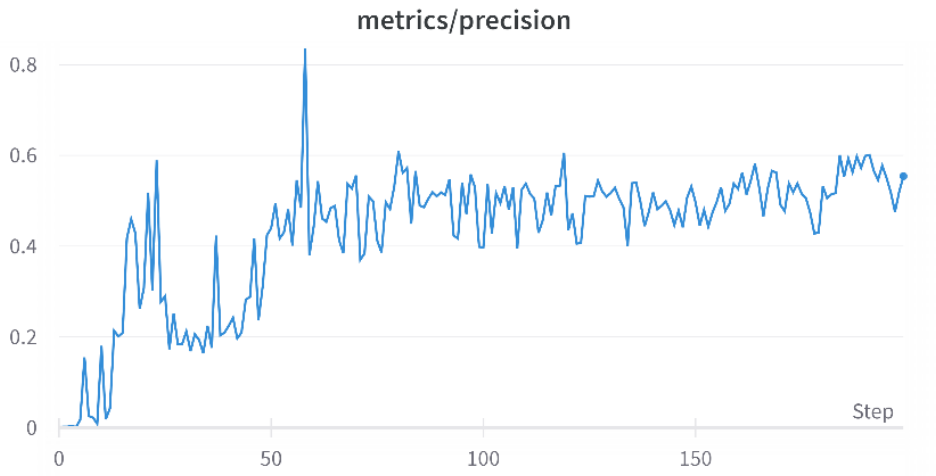
It can be seen that each training batch consists of 16 images, and each image contains several types of pavement damage identified with a specific color of the bounding box.

5.2 Validation Results

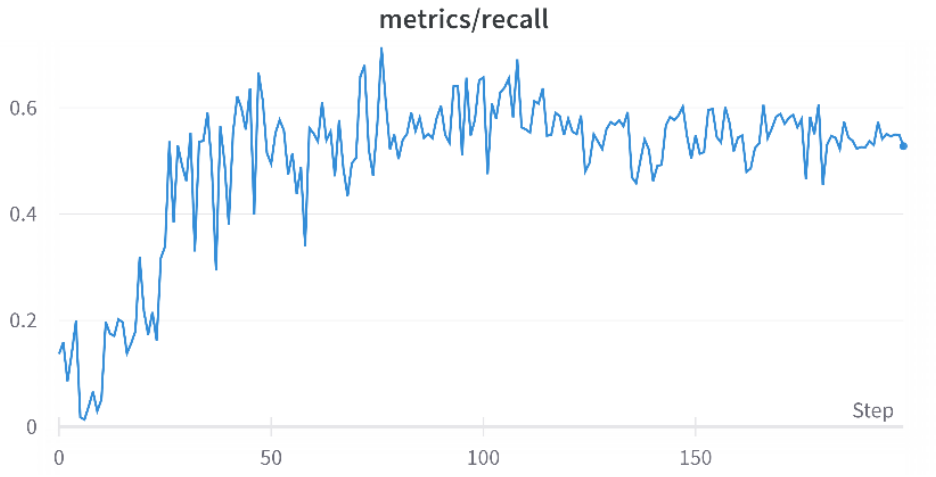
During the training process, the performance of each epoch is evaluated. The evaluation process for 200 training epochs is measured through several metrics such

as precision, recall, mean average precision, and loss trend for objects, classes, and bounding boxes. This section reports and elaborates on the validation results.

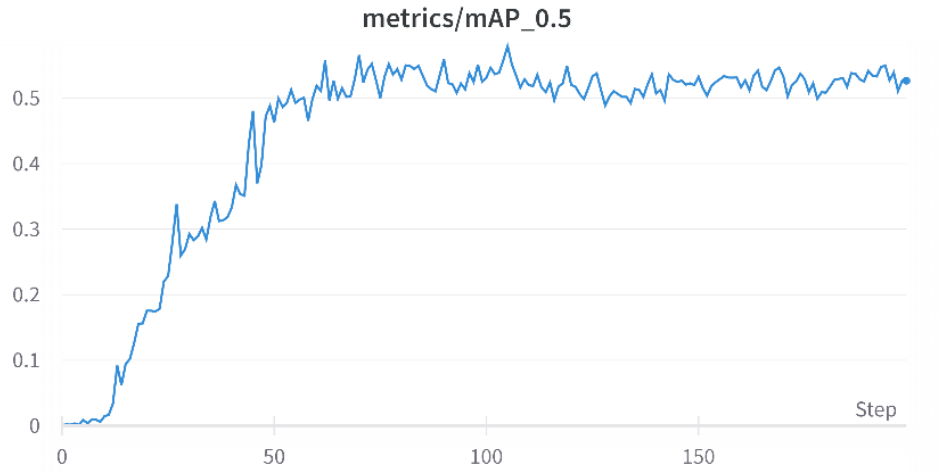
One of the most essential criteria is evaluating the precision, recall, and mean average precision during the training phase. Figure 37 illustrates the trends of these three metrics during the training phase.



(a)



(b)

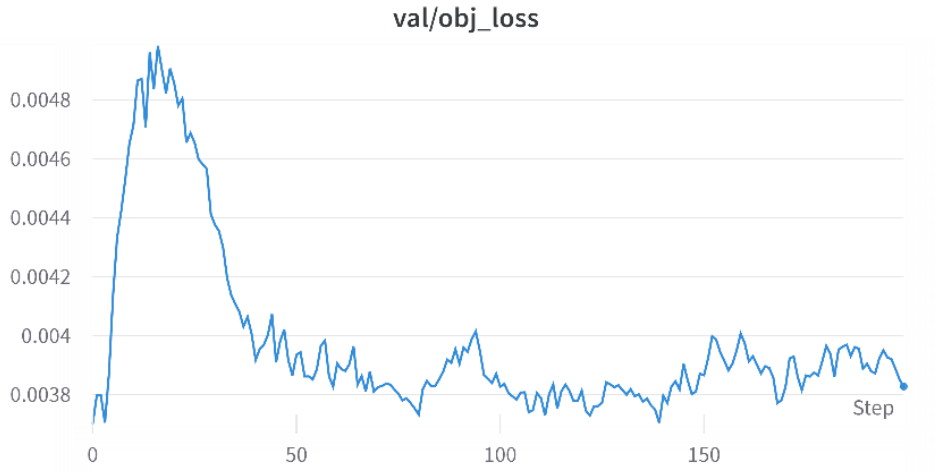


(c)

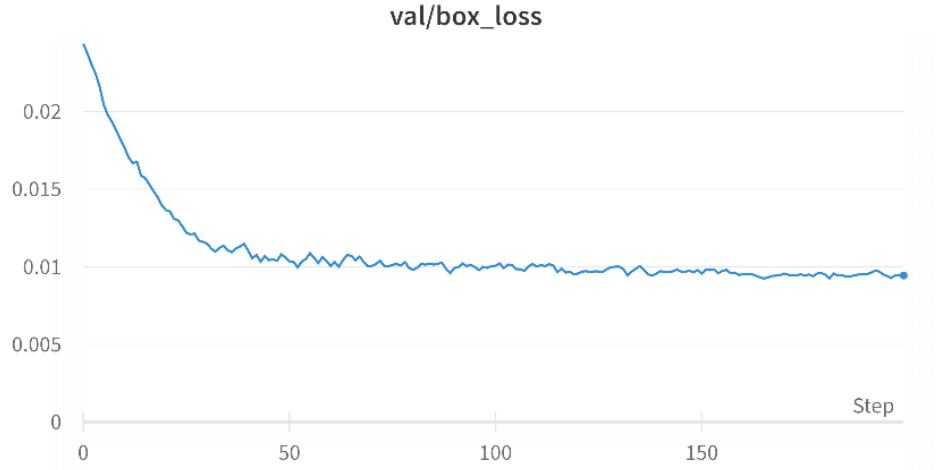
Figure 37: Validation metrics plot. (a) precision, (b) recall, (c) mean average precision

It is shown that precision, recall, and mean average precision have up-trending until approximately 100 epochs, and it fluctuated between 100 and 200 epochs which shows that the model is not able to learn the features after 100 epochs. Figure 37 demonstrates that the model performs acceptably in the training process.

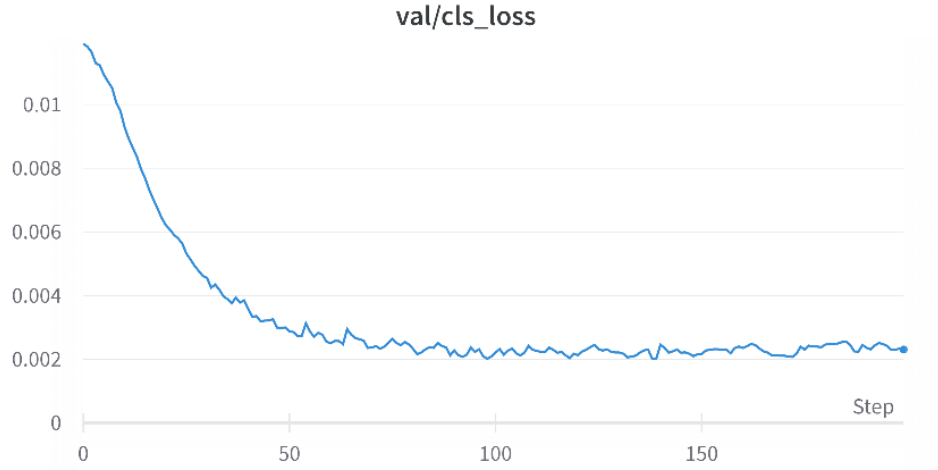
Another essential criterion during the training process is the value of losses. YOLO algorithm contains three types of losses. "Object loss" represents the loss value related to detecting the correct object (distress), "class loss" represents the loss value related to identifying the correct label for each class (distress type), and "box loss" indicates the loss value for identifying correct location of the objects (distress locations). Figure 38 illustrates the loss values for the validation set during the training process.



(a)



(b)



(c)

Figure 38: Validation losses plot. (a) object loss, (b) box loss, (c) class loss

The figure above shows that object loss value increased during the first 20 epochs, which is the warmup stage. After 20 epochs, the object loss value decreased drastically until epoch 50, and after that, the value fluctuated for the rest of the training. Moreover, the loss values for classes and bounding boxes have a decreasing trend until 50 epochs, and they remained steadily till the rest of the training process. The validation loss values show an acceptable performance of the model during the training phase. Figure 39 illustrates the confusion matrix for the accuracy of the predicted images for different distress types.

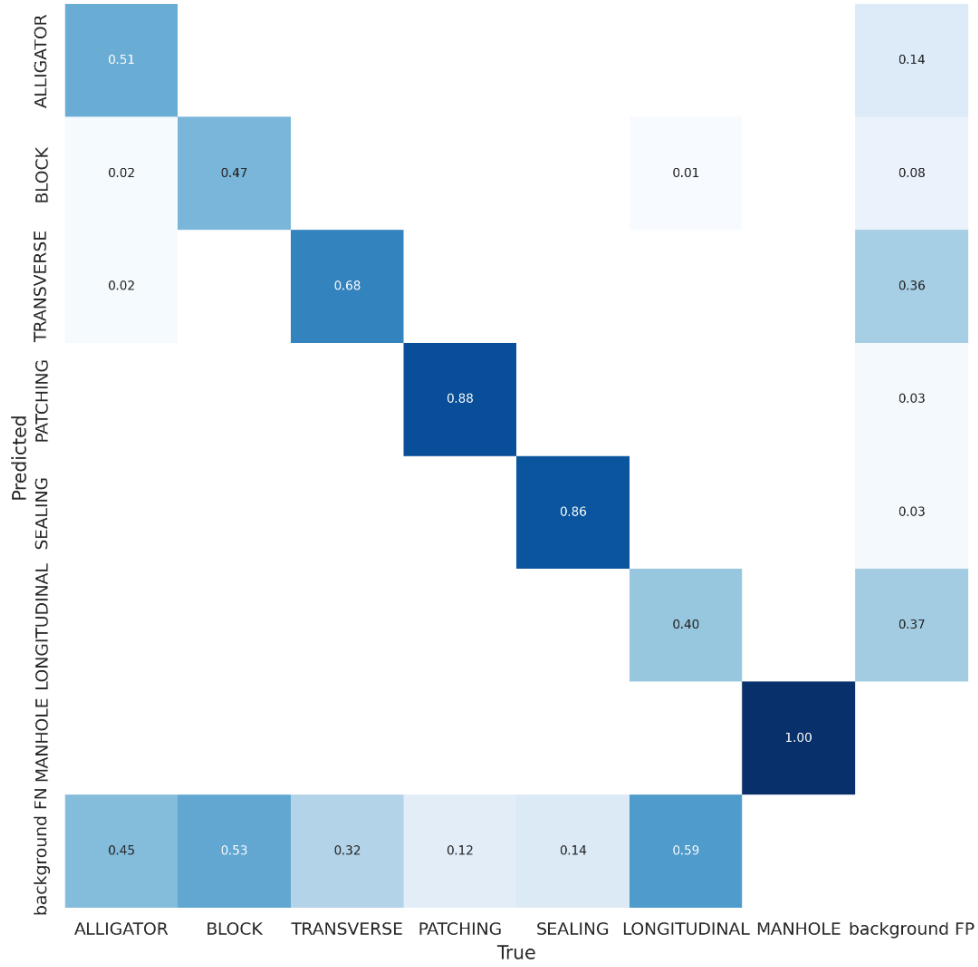


Figure 39: Confusion matrix of predicted images

It can be seen that among all distress types, the highest accuracy belongs to manhole, patching, and sealing with 1, 0.88, and 0.86, respectively. Furthermore, the lowest accuracy belongs to longitudinal and block cracking with 0.4 and 0.48. Figure 40 depicts a sample batch of the predicted images. Although it has a few prediction mistakes, considerable damages have been successfully identified and localized.

5.3 Detection Results

In order to optimize the accuracy of the detection phase, several modifications were made as follows:

- Confidence Threshold (0.4)
- NMS-IoU Threshold (0.99)
- Test Time Augmentation (applied).

Optimizing the abovementioned detection parameters was based on trial and error. Non-Max Suppression IoU threshold represents the confidence to deny or approve a predicted bounding box. Moreover, the images are augmented using methods mentioned in the previous section, such as bounding box safe crop and image inversion. After optimizing the detection parameters, the precision, recall, and F1 scores reached 0.95, 0.92, and 0.93, respectively.

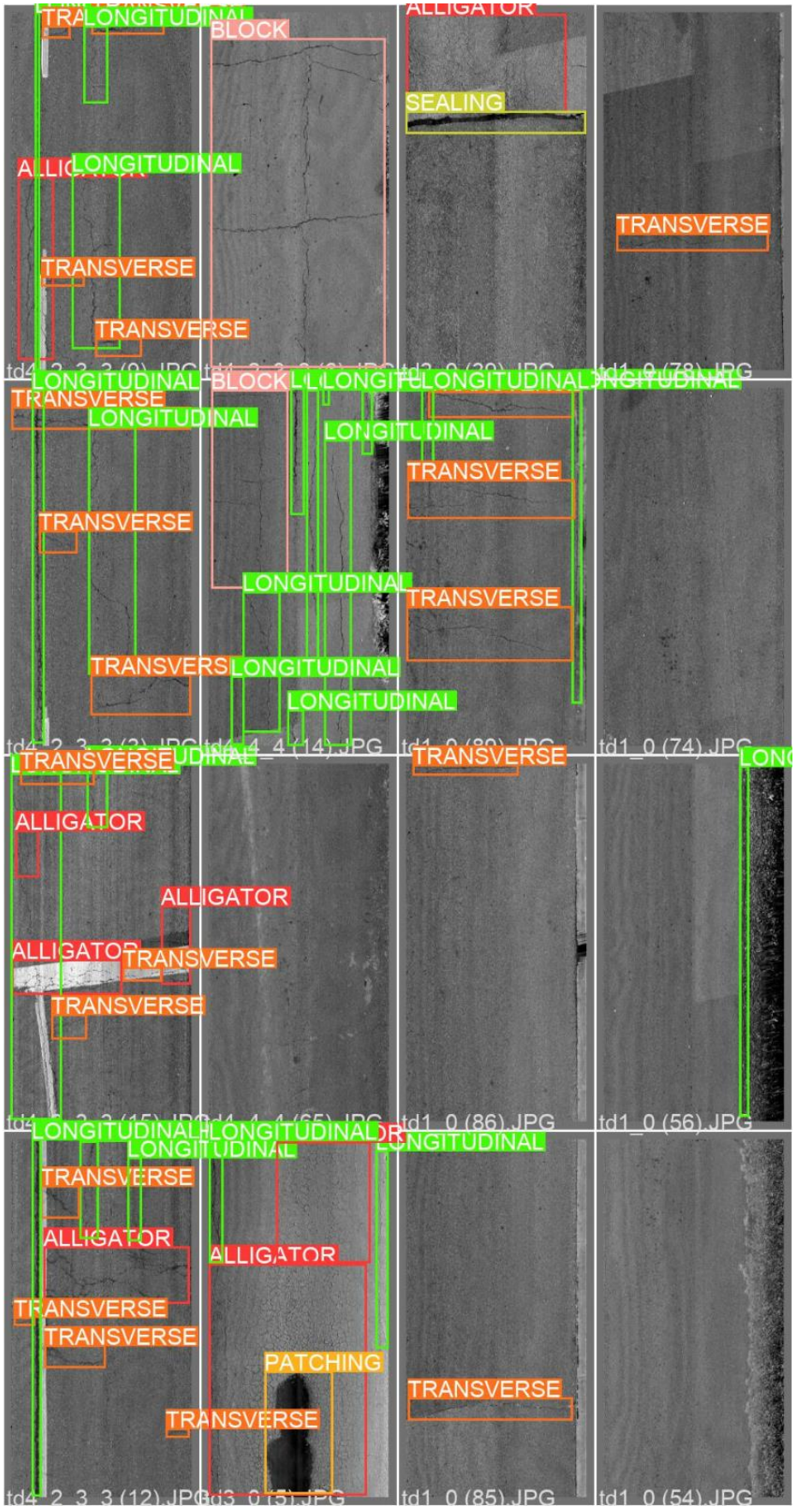


Figure 40: A sample batch of the predicted images

Chapter 6

CONCLUSION AND RECOMMENDATIONS

This study contains an extensive literature review on pavement distress types, pavement condition evaluation, and the importance of artificial intelligence, specifically deep learning, in pavement management systems. In this regard, essential research gaps have been identified and discussed. The main objectives of this study aimed to fill the gaps elaborated in the literature review. This study developed a pavement distress detection model utilizing the most recent prominent techniques, strategies, and algorithms in deep learning technology to achieve an accurate, robust, and reliable model for detecting, classifying, and localizing several pavement distress types. Moreover, optimizing data-centric and model-centric strategies and approaches have been analyzed to improve the proposed model's generalization.

The dataset used in this study contained 682 top-down view images captured in several cities in the U.S. (i.e., Jefferson City, Kansas City, and Columbia in Missouri), including various distress types such as alligator, longitudinal, transverse, and block cracking as well as patching, sealing, and manhole. In order to train the model, the YOLOv5 algorithm was developed on the PyTorch framework using Python programming language. Moreover, the proposed model is optimized using data augmentation techniques such as bounding box safe crop and image inversion. After evaluating the performance of the proposed model, the accuracy reached 0.95, 0.92, and 0.93 in terms of precision, recall, and F1 score, respectively.

In the future, the author aims to develop a model based on advanced deep learning algorithms such as the U-Net to measure the severity of pavement distress as well. In this regard, a fully automated pavement condition evaluation would be possible. Also, measuring and calculating the pavement performance indices, such as the pavement condition index, would be achieved efficiently.

REFERENCES

Abd El-Hakim, R., & El-Badawy, S. (2013). International roughness index prediction for rigid pavements: An artificial neural network application. *Advanced Materials Research, 723*, 854-860. doi:10.4028/www.scientific.net/amr.723.854

Arya, D., Maeda, H., Ghosh, S. K., Toshniwal, D., Mraz, A., Kashiyama, T., & Sekimoto, Y. (2021). Deep learning-based road damage detection and classification for multiple countries. *Automation in Construction, 132*, 103935. doi:10.1016/j.autcon.2021.103935

ASTM, Standard practice for roads and parking lots pavement condition index surveys. (2018). West Conshohocken, PA, United States: ASTM International.

Bochkovskiy, A., Wang, C., & Liao, H. (2020, April 23). Yolov4: Optimal Speed and accuracy of object detection. Retrieved June 6, 2022, from https://arxiv.org/abs/2004.10934

Bunker, R. P., & Thabtah, F. (2019). A machine learning framework for sport result prediction. *Applied Computing and Informatics, 15*(1), 27-33. doi:10.1016/j.aci.2017.09.005

Chambon, S., & Moliard, J. (2011). Automatic road pavement assessment with image processing: Review and comparison. *International Journal of Geophysics, 2011*, 1-20. doi:10.1155/2011/989354

Cheng, H., & Miyojim, M. (1998). Automatic pavement distress detection system. *Information Sciences, 108*(1-4), 219-240. doi:10.1016/s0020-0255(97)10062

Coenen, T. B., & Golroo, A. (2017). A review on automated pavement distress detection methods. *Cogent Engineering, 4*(1), 1374822. doi:10.1080/23311916.2017.1374822

Dede, T., Kankal, M., Vosoughi, A. R., Grzywiński, M., & Kripka, M. (2019). Artificial Intelligence Applications in civil engineering. *Advances in Civil Engineering, 2019*, 1-3. doi:10.1155/2019/8384523

Dhande, M. (2020, July 03). What is the difference between AI, Machine Learning and deep learning? Retrieved June 23, 2022, from https://www.geospatialworld.net/blogs/difference-between-ai%ef%bb%bf-machine-learning-and-deep-learning/

Du, Y., Pan, N., Xu, Z., Deng, F., Shen, Y., & Kang, H. (2020). Pavement distress detection and classification based on Yolo Network. *International Journal of Pavement Engineering, 22*(13), 1659-1672. doi:10.1080/10298436.2020.1714047

Ferreira, F. G., Gandomi, A. H., & Cardoso, R. T. (2021). Artificial Intelligence applied to stock market trading: A Review. *IEEE Access, 9*, 30898-30917. doi:10.1109/access.2021.3058133

Haas, R., Hudson, W. R., & Falls, L. C. (2015). *Pavement Asset Management*. Salem, MA: Scrivener Publishing.

Haas, R., Hudson, W. R., & Zaniewski, J. P. (1994). *Modern Pavement Management*. Malabar, FL: Krieger.

Hass C., & Hendrickson C. (1990). Computer-based model of pavement surface. *Transportation Research Record: Journal of the Transportation Research Board, 1260*, 91-98.

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 37*(9), 1904-1916. doi:10.1109/tpami.2015.2389824

Hossain, M. I., Gopisetti, L. S., & Miah, M. S. (2017). Prediction of international roughness index of flexible pavements from climate and traffic data using artificial neural network modeling. *Airfield and Highway Pavements 2017*. doi:10.1061/9780784480922.023

Huang, S., Yang, J., Fong, S., & Zhao, Q. (2020). Artificial Intelligence in cancer diagnosis and prognosis: Opportunities and challenges. *Cancer Letters, 471*, 61-71. doi:10.1016/j.canlet.2019.12.007

Issa, A., Samaneh, H., & Ghanim, M. (2022). Predicting pavement condition index using artificial neural networks approach. *Ain Shams Engineering Journal, 13*(1), 101490. doi:10.1016/j.asej.2021.04.033

Ji, A., Xue, X., Wang, Y., Luo, X., & Xue, W. (2020). An integrated approach to automatic pixel-level crack detection and quantification of Asphalt Pavement. *Automation in Construction, 114*, 103176. doi:10.1016/j.autcon.2020.103176

Ji, H., Alfarraj, O., & Tolba, A. (2020). Artificial Intelligence-empowered edge of vehicles: Architecture, enabling technologies, and applications. *IEEE Access, 8*, 61020-61034. doi:10.1109/access.2020.2983609

Jia, W., Xu, S., Liang, Z., Zhao, Y., Min, H., Li, S., & Yu, Y. (2021). Real-time automatic helmet detection of motorcyclists in urban traffic using improved Yolov5 detector. IET Image Processing, 15(14), 3623-3637. doi:10.1049/ipr2.12295

Jocher, G., Stoken, A., Borovec, J., NanoCode012, Christopher STAN, Changyu, L., . . . Rai, P. (2020, October 29). Ultralytics/Yolov5: V3.1 - bug fixes and performance improvements. Retrieved June 6, 2022, from https://zenodo.org/record/4154370

Johri, P., Diván, M. J., Khanam, R., Marciszack, M., & Will, A. (2022). *Trends and advancements of image processing and its applications*. Cham: Springer International Publishing.

Lin, T., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., . . . Zitnick, C. L. (2014). Microsoft Coco: Common Objects in Context. *Computer Vision – ECCV 2014*, 740-755. doi:10.1007/978-3-319-10602-1_48

Long-term Pavement performance program. (2015). Washington, DC: U.S. Department of Transportation, Federal Highway Administration.

Majidifard, H., Jin, P., Adu-Gyamfi, Y., & Buttlar, W. G. (2020). Pavement image datasets: A new benchmark dataset to classify and densify pavement distresses. Transportation Research Record: Journal of the Transportation Research Board, 2674(2), 328-339. doi:10.1177/0361198120907283

Manzoor, B., Othman, I., Durdyev, S., Ismail, S., & Wahab, M. (2021). Influence of artificial intelligence in civil engineering toward Sustainable Development—a systematic literature review. *Applied System Innovation, 4*(3), 52. doi:10.3390/asi4030052

Marcelino, P., De Lurdes Antunes, M., Fortunato, E., & Gomes, M. C. (2019). Machine Learning Approach for Pavement Performance Prediction. *International Journal of Pavement Engineering, 22*(3), 341-354. doi:10.1080/10298436.2019.1609673

Mei, Q., & Gül, M. (2020). A cost effective solution for pavement crack inspection using cameras and deep neural networks. *Construction and Building Materials, 256*, 119397. doi:10.1016/j.conbuildmat.2020.119397

Miller, J. S., & Bellinger, W. Y. (2003, June 01). Distress identification manual for the LTPP (Fourth revised edition). Retrieved June 6, 2022, from https://www.fhwa.dot.gov/publications/research/infrastructure/pavements/ltpp/reports/03031/

Narvekar, M., & Fargose, P. (2015). Daily Weather Forecasting using artificial neural network. *International Journal of Computer Applications, 121*(22), 9-13. doi:10.5120/21830-5088

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., . . . Chintala, S. (2019, January 01). Pytorch: An imperative style, high-performance deep learning library. Retrieved July 21, 2022, from http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library

Redmon, J., & Farhadi, A. (2017). Yolo9000: Better, faster, stronger. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. doi:10.1109/cvpr.2017.690

Redmon, J., & Farhadi, A. (2018, April 08). Yolov3: An incremental improvement. Retrieved June 6, 2022, from https://arxiv.org/abs/1804.02767

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. doi:10.1109/cvpr.2016.91

Tsai, Y. J., & Li, F. (2012). Critical assessment of detecting asphalt pavement cracks under different lighting and low intensity contrast conditions using emerging 3D laser technology. *Journal of Transportation Engineering, 138*(5), 649-656. doi:10.1061/(asce)te.1943-5436.0000353

Ultralytics. (n.d.). Ultralytics/yolov5: Yolov5 in PyTorch ONNX CoreML TFLite. Retrieved June 6, 2022, from https://github.com/ultralytics/yolov5

Wang, C., Mark Liao, H., Wu, Y., Chen, P., Hsieh, J., & Yeh, I. (2020). CSPNet: A new backbone that can enhance learning capability of CNN. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. doi:10.1109/cvprw50498.2020.00203

Wang, K., Liew, J. H., Zou, Y., Zhou, D., & Feng, J. (2019). Panet: Few-shot image semantic segmentation with prototype alignment. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. doi:10.1109/iccv.2019.00929

Zakeri, H., Nejad, F. M., & Fahimifar, A. (2016). Image based techniques for crack detection, classification and quantification in Asphalt Pavement: A Review. *Archives of Computational Methods in Engineering*, *24*(4), 935-977. doi:10.1007/s11831-016-9194-z

Zhang, Z., & Friedrich, K. (2003). Artificial Neural Networks applied to Polymer Composites: A Review. *Composites Science and Technology*, *63*(14), 2029-2044. doi:10.1016/s0266-3538(03)00106-4