BRAND AND SIZE DETECTION OF VEHICLE TIRES WITH DEEP LEARNING

A THESIS SUBMITTED TO
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
KOCAELİ UNIVERSITY

BY

VOLKAN ÖZDEMİR

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
ELECTRONICS AND COMMUNICATIONS ENGINEERING

KOCAELİ 2022

BRAND AND SIZE DETECTION OF VEHICLE TIRES WITH DEEP LEARNING

A THESIS SUBMITTED TO
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
KOCAELİ UNIVERSITY

BY

VOLKAN ÖZDEMİR

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
ELECTRONICS AND COMMUNICATIONS ENGINEERING

Prof. Dr. Oğuzhan URHAN		
Supervisor,	**Kocaeli University**	**.....**
Prof. Dr. Behçet Uğur TÖREYİN		
Jury member,	**Istanbul Technical University**	**.....**
Assoc. Prof. Orhan AKBULUT		
Jury member,	**Kocaeli University**	**.....**

Thesis Defense Date: 20.06.2022

ETİK BEYAN VE ARAŞTIRMA FONU DESTEĞİ

Kocaeli Üniversitesi Fen Bilimleri Enstitüsü tez yazım kurallarına uygun olarak hazırladığım bu tez/proje çalışmasında,

- Bu tezin/projenin bana ait, özgün bir çalışma olduğunu,
- Çalışmamın hazırlık, veri toplama, analiz ve bilgilerin sunumu olmak üzere tüm aşamalarında bilimsel etik ilke ve kurallara uygun davrandığımı,
- Bu çalışma kapsamında elde edilen tüm veri ve bilgiler için kaynak gösterdiğimi ve bu kaynaklara kaynakçada yer verdiğimi,
- Bu çalışmanın Kocaeli Üniversitesi'nin abone olduğu intihal yazılım programı kullanılarak Fen Bilimleri Enstitüsü'nün belirlemiş olduğu ölçütlere uygun
- olduğunu,
- Kullanılan verilerde herhangi bir tahrifat yapmadığımı,
- Tezin/Projenin herhangi bir bölümünü bu üniversite veya başka bir üniversitede başka bir tez/proje çalışması olarak sunmadığımı,

beyan ederim.

Bu tez/proje çalışması kapsamında üretilen veri ve bilgiler Brisa Bridgestone Sabancı Lastik Sanayi ve Ticaret A.Ş. tarafından Ar-Ge çalışmaları kapsamında maddi/alt yapı desteği alınarak gerçekleştirilmiştir.

Herhangi bir zamanda, çalışmamla ilgili yaptığım bu beyana aykırı bir durumun saptanması durumunda, ortaya çıkacak tüm ahlaki ve hukuki sonuçları kabul ettiğimi bildiririm.

Volkan ÖZDEMİR

YAYIMLAMA VE FİKRİ MÜLKİYET HAKLARI

Fen Bilimleri Enstitüsü tarafından onaylanan lisansüstü tezimin/projemin tamamını veya herhangi bir kısmını, basılı ve elektronik formatta arşivleme ve aşağıda belirtilen koşullarla kullanıma açma izninin Kocaeli Üniversitesi'ne verdiğimi beyan ederim. Bu izinle Üniversiteye verilen kullanım hakları dışındaki tüm fikri mülkiyet haklarım bende kalacak, tezimin/projemin tamamının ya da bir bölümünün gelecekteki çalışmalarda (makale, kitap, lisans ve patent vb.) kullanımı bana ait olacaktır. Tezin/projenin kendi özgün çalışmam olduğunu, başkalarının haklarını ihlal etmediğimi ve tezimin/projenin tek yetkili sahibi olduğumu beyan ve taahhüt ederim. Tezimde yer alan telif hakkı bulunan ve sahiplerinden yazılı izin alınarak kullanılması zorunlu metinlerin yazılı izin alarak kullandığımı ve istenildiğinde suretlerini Üniversiteye teslim etmeyi taahhüt ederim.
Yükseköğretim kurulu tarafından yayınlanan ***"Lisanüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge"*** kapsamında tezim aşağıda belirtilen koşullar haricinde YÖK Ulusal Tez Merkezi/ Kocaeli Üniversitesi Kütüphaneleri Açık Erişim Sisteminde erişime açılır.

Tezim/projem ile ilgili gizlilik kararı verilmemiştir.

Volkan ÖZDEMİR

PREFACE AND ACKNOWLEDGEMENTS

I am very grateful to BRISA BRIDGESTONE SABANCI Tire Manufacturing Inc. for all the support during the study. And it was a great honor to work with Anıl ÇALIŞKAN and Murat ÖZGEN during this study. Also, I appreciate all my colleagues and managers who contributed to this thesis.

I would like to express my heartfelt thanks to my thesis advisor Prof. Dr. Oğuzhan URHAN who always supported and directed me on this journey.

I would like to thank my father Ali ÖZDEMİR, my mother Kezban ÖZDEMİR and my uncle Taner ÖZDEMİR who have always supported me and my dreams.

Lastly, I am eternally grateful to my wife, Berrin ÖZKAN ÖZDEMİR, who never ever hesitate to give her precious love and support. None of this would have come true without her.

May - 2022 Volkan ÖZDEMİR

CONTENTS

ETİK BEYAN VE ARAŞTIRMA FONU DESTEĞİ i
YAYIMLAMA VE FİKRİ MÜLKİYET HAKLARI ii
PREFACE AND ACKNOWLEDGEMENTS iii
LIST OF FIGURES v
LIST OF TABLES vi
INDEX OF SYMBOLS AND ABBREVIATIONS vii
ÖZET viii
ABSTRACT ix
1. INTRODUCTION 1
2. PROBLEM DEFINITION 6
2.1. Tire As a Product and Its Markings 6
2.2. Tire Purchasing and Online Shopping 8
2.3. Need for Automized Brand and Size Detection 10
2.4. Studies in the Literature 10
3. THEORETICAL BACKGROUND 13
3.1. Machine Learning 13
3.2. Artificial Neural Networks and Deep Neural Networks 14
3.3. Computer Vision Tasks 19
3.3.1. Classification 20
3.3.2. Object Detection 20
3.3.3. Segmentation 24
4. DATA SET AND PERFORMANCE METRICS 27
4.1. Data Set 27
4.1.1. Data Set Collecting 29
4.1.2. Data Set Preparation 30
4.2. Computing Environment and Tools 33
4.3. Performance Metrics 33
5. PROPOSED SYSTEM FOR DETECTION OF BRAND AND SIZE OF VEHICLE TIRES 35
5.1. Proposed Pipeline of The System 35
5.2. Preprocessing 37
5.3. Tire Segmentation Model 39
5.4. Tire Transform and Background Removal 42
5.5. Tire Brand and Size Location Detection Model 44
5.6. Size Detection Model 45
6. RESULTS AND DISCUSSION 48
7. CONCLUSION AND RECOMMENDATIONS 54
REFERENCES 55
PUBLICATIONS AND WORKS 60
BIOGRAPHY 61

LIST OF FIGURES

Figure 2. 1. Pneumatic Bicycle Tire Produced by John Boyd Dunlop in 1888..... 6
Figure 2. 2. Nomenclature of the Tire Size 8
Figure 2. 3. Tire Brand and Size..... 10
Figure 3. 1. Machine learning algorithms due to training data type..... 14
Figure 3. 2. Perceptron Model..... 15
Figure 3. 3. Multi Layer Neural Network..... 16
Figure 3. 4. Different non-linear activation funtions..... 17
Figure 3. 5. Example of Convolutional Neural Network Structure..... 19
Figure 3. 6. Different Computer Vision Tasks..... 20
Figure 3. 7. Object Detection Taxonomy 22
Figure 3. 8. Architecture of YOLOv3 23
Figure 3. 9. Architecture of YOLOv5 24
Figure 3. 10. YOLO Family Evolution Timeline..... 24
Figure 3. 11. The Architecture of SegNet Image Segmentation Model 25
Figure 3. 12. The Architecture of U-Net 26
Figure 4. 1. Types of the Unwanted Images..... 28
Figure 4. 2. Tire Images in Different Conditions 29
Figure 4. 3. Data Collection Interface 30
Figure 4. 4. Data Augmentation Techniques..... 31
Figure 4. 5. Tire and Rim Labeling for Segmentation Model and Creating Masks..... 32
Figure 4. 6. Labeling for Object Detection Methods..... 33
Figure 5. 1. Tire Size and Brand Detection Algortihm Flow Chart 37
Figure 5. 2. Blur and Darkness Detection 38
Figure 5. 3. Result Examples of Tire - Rim Segmentation Model..... 40
Figure 5. 4. Calculated Parameters from Rire-Rim Segmentation Model 41
Figure 5. 5. Tire Existence Check on Segmentation Result Mask 41
Figure 5. 6. Inverse Polar Remaps References..... 43
Figure 5. 7. Output of the Inverse Polar Transform and Background Removal Example..... 44
Figure 5. 8. Brand Detection Results Shown as Bounding Boxes 45
Figure 5. 9. Size Location Detection Example..... 45
Figure 5. 10. Size detection Model Outputs..... 46
Figure 5. 11. Flow Chart of The System with Outputs 47
Figure 6. 1. Confusion Matirces of Brand Detection Algorithms on The Same Test Set..... 49
Figure 6. 2. YOLOv3 and YOLOv5 Performance Comparison for Character Recognition..... 50
Figure 6. 3. Confusion Matrices of Size Detection Models for Character Recognition..... 51
Figure 6. 4. Test Set Distribution due to Rim Size of the Tires in The Data Set 52

LIST OF TABLES

Table 6. 1. Performance Comparison of Brand Detection Model Algorithms..... 49
Table 6. 2. Performance comparison of character recognition model algorithms..... 51
Table 6. 3. Hardware (GPU vs CPU) and Framework (OpenCV-DNN vs TensorFlow) Inference Time Comparison on the Models 53
Table 6. 4. Inference Time Breakdown for All Steps of Pipeline 53

INDEX OF SYMBOLS AND ABBREVIATIONS

Abbreviations

AI	: Artificial Intelligence
ANN	: Artificial Neural Network
B2C	: Business to Consumer
CmBN	: Cross Mini-Batch Normalization
CNN	: Convolutional Neural Network
CPU	: Central Processing Unit
CX	: Customer Experience
DCNN	: Deep Convolutional Neural Network
DL	: Deep Learning
DOT	: Department of Transportation
ETRTO	: European Tire and Rim Organization
FCN	: Fuly Convolutional Network
FN	: Folse Negative
FP	: False Positive
GPU	: Graphics Processing Unit
HSV	: Hue Saturation Value
IoU	: Intersection of Union
mAP	: Mean Average Precision
ML	: Machine Learning
MS COCO	: Microsoft Common Objects in Context
NLP	: Natural Language Processing
PAN	: Path Agfregation network
RAM	: Random Access Memory
ReLu	: Rectified Linear Unit
ROI	: Region of Interest
SAM	: Spatial Attention Module
SAT	: Self Adversarial Training
SSD	: Solid State Drive
TN	: True Negative
TP	: True Positive
UI	: User Interface
UNECE	: The United Nations Economic Commision for Europe
XML	: Extensible Markup Language
YOLO	: You Only Look Once

DERİN ÖĞRENME İLE KARA TAŞITI ARAÇLARI LASTİKLERİNDE MARKA VE EBAT TESPİTİ

ÖZET

Bu tezde, bir görüntüden uygun araç lastiklerini seçmek için gerekli bilgileri tespit etmek için bir yapay zeka sistemi geliştirilmiştir. Lastikler, kara taşıtlarının yola temas eden ekipmanı olduğu için güvenlik ve performans için bir aracın en kritik parçalarıdır. Bu nedenle, kullanıcıların araç standartlarına uygun doğru lastikleri seçmeleri gerekir, aksi takdirde araca yanlış lastik seçimi performans kaybına, ekstra yakıt tüketimine, aşırı gürültüye ve ne yazık ki birçok kazaya neden olabilir. Araç sürücülerinin çoğu, satın almadan önce eski lastiklerinden veya araç kapısının arkasında bulunan etiketten lastik satın almak için gerekli olan ebat bilgisini kontrol ederler. Ancak, genellikle bu bilgileri lastiklerden veya aracın kapısındaki etiketten bulmak zor olabilir. Kullanıcılar, lastiğin kullanım ile alakalı tüm bilgileri ve uyarıları içeren bölgeyi kontrol ettiklerinde, lastik üzerinde bulunan çok sayıda yazı ve sembol arasından ihtiyaç duydukları bilgiyi kolayca bulamazlar. Bu nedenle satın almadan kullanıcı deneyimleri kesintiye uğrayabilir. Bu soruna çözüm bulmak için, bu tez kapsamında kullanıcının mobil cihazlarından alınan tek bir görüntüden lastik satın almak için yeterli marka ve ebat bilgilerini tespit edebilen denetimli derin öğrenme algoritmalarına dayalı bir sistem geliştirilmiştir. Önerilen sistem, lastik satın almak için gerekli olan marka ve ebat bilgisini algılayarak müşteriye öneriler sunan kullanıcı dostu bir akışa sahiptir. Son kullanıcıların etkin şekilde kullanabileceği bir çözüm elde edebilmek için probleme özgü bir veri seti toplanmış ve hazırlanmıştır. Geliştirilen sistemde bulunan tespit modelleri için derin sinir ağlarına dayalı bölütleme ve nesne tespit yöntemleri ve görüntü işleme algoritmaları kullanılmıştır. Nihai sistem marka tespitinde %96.5, ebat için karakter tanımada %98.7 ve toplam ebat tespitinde %96.5 başarım göstermiştir.

Anahtar Kelimeler: Bölütleme, Derin Öğrenme, Lastik, Nesne Tespiti, Yapay Zeka.

BRAND AND SIZE DETECTION OF VEHICLE TIRES WITH DEEP LEARNING

ABSTRACT

In this thesis, an artificial intelligence system for gathering essential information in order to select suitable vehicle tires from an image has been developed. Since tires are the only equipment of ground vehicles that touch the road, they are the most critical parts of a vehicle in many aspects such as safety, performance, noise, and comfort. Hence, users should select right tires that comply with vehicle standards otherwise selecting incorrect tires to the vehicle may cause performance loss, extra fuel consumption, excessive noise and unfortunately several accidents. Most of the vehicle drivers check necessary information whether from their old tires or from the label that is located at their vehicle's door before purchasing. Especially, when users check old tires' sidewall, they cannot find easily sufficient information they need due to not recognize between lots of writing and symbols on the tire. Therefore, user's experience is interrupted. To find a solution this problem, a system based on supervised deep learning algorithms has been developed to detect and determine the required information for purchasing a tire namely brand, and size. The proposed system has a user-friendly pipeline that detects the brand and size. A specific dataset was collected and prepared for the solution. Image processing algorithms and Segmentation and object detection methods which are based on deep neural networks have been used for the detection models in developed system. The final system achieved 96.5% accuracy in brand detection, 98.7% in character recognition for size and 96.5% in total size detection.

Keywords: Segmentation, Deep Learning, Tire, Object Detection, Artificial Intelligence.

1. INTRODUCTION

In recent years, artificial intelligence (AI) has made it possible to use many applications that have been replace human effort and human hand and eye control in many sectors such as retail, manufacturing, health, marketing, telecommunication, finance, education etc. The enablers of this succession are high speed Internet, improvements in machine learning, open-source knowledge sharing and high performance and accordingly affordable hardware's such as GPUs, CPUs, RAMs. Thanks to those improvements, developers and engineers have been started to solve important problems in the above-mentioned fields and find tremendous solutions which yield using less effort and source compared to existing systems. These solutions, in which techniques such as machine learning and deep learning in the field of artificial intelligence, are used, provide cost advantage and efficiency to users and manufacturers in almost every sector. While this is the case, the increasing accessibly and usability of these technologies has enabled more research and development regardless of the sector.

For a long time, classification, recognition, and prediction have been one of the biggest challenges in computer science world. Deciding the correct class in an image, signal, pattern, pixel or even a word is vital for many problems of today. Sometimes, detecting an outlier or predicting the exact value of an output can be the main aim of the technical study. In accordance with the purpose of the study, many difficult problems have had a new solution with an artificial neural network structure to be selected, whether they learn from the past data set or not. After the availability of the deep learning models with enhanced hardware and open-source frameworks, libraries, data sets and techniques, plenty of use cases in many fields can get solved with ease. These advances in AI have created opportunities that can be a solution to many problems in the retail sector. Applications that will assist customers in product selection and replace non-value-added information detection on the products with artificial intelligence offer better customer experience.

Undoubtedly, the tire is one of the parts that requires the most attention due to its safety and suitability compared to other retail products. Therefore, selection the suitable tire for one's vehicle is not a simple task. Mostly, customers prefer to consult a retail store or an

expert, but this process interrupts the purchasing experience especially in online shopping. The main information which is needed to complete a tire purchasing experience is tire size which is consist of alphanumerical characters indicates the dimensions of the tire and tire's brand. However, when the majority of tire customers search the required information for tire purchasing at the sidewall of the corresponded tire, it is difficult to find location of their size marking since a tire contains numerous texts, symbols, and signs on its black surface. Since tire is an equipment that wearing out through the usage, due to the erosion, stretches and dust on the writings makes the size finding problem more difficult.

The United Nations Economic Commission for Europe (UNECE) is the authority that sets the vehicle regulations to ensure the standard all over the world. UNECE defined the tire definitions and markings by vehicle and transportation agreement (Gas, 2008). According to this prescription, there are more than 20 markings on the tire sidewall that a pneumatic tire manufacturer should state on the tire such as trade name or brand, tire size designation, speed index, load capacity, the date of manufacture, inflation pressure, bead setting pressure, warnings, ply rating number, the category of use, structure etc. However, the locations of these markings are not specified. In other words, the location, font, size, and type of engraving style varies due the tire designer's selection. As mentioned before, during the purchasing process of a new tire, users need to notify tire size information, which can occur in 9 characters or more, usually located on a random place of the sidewall. It is arduous to distinguish the necessary markings which are tire size and brand between multiple markings. Along with this problem, the rising online shopping trend, especially after COVID-19, is altering customer habits. Even the tire as a product, which is preferably bought by touching and feeling with traditional methods, is sold on many online shopping sites with different possibilities. However, because of not having tire size and brand clearly in the tire buying process, the purchasing experience is interrupted before the purchase takes place. This condition rises the need of extracting the required tire information in an easy way to create an end-to-end excellent tire purchasing experience.

In last decade, people use mobile devices which are an indispensable part of our lives, to meet many different needs with the help of increasing speed of the Internet such as social

networking, communication, online shopping, health control etc. Hence, to solve this a problem, it would be effortless and a smart way to choose a function that can be fulfilled by the mobile devices that customers use in many moments of their lives. For this reason, it is aimed to develop a system that will detect the information necessary for the purchase of tires through just a tire image taken by the customers with their mobile devices.

With the purpose of meeting such a need computer vision technology has been chosen. Computer vision analyses images or videos and can provide the desired information with advanced machine learning and image processing algorithms. Computer vision serves to researchers in many tasks such as image classification, object detection, segmentation, and image generation. These tasks become a different technical research area depending on the type of information needed to be sought from an instance. If the information which object is in the image or which class the object belongs is search, this technique called classification. When the class and location or coordinates of specific object class is important together, this technique is called object detection. What is more, if one is seeking to find the pixels' location of a specific object with its class, it is called segmentation. Besides, if the main target is to generate new images consist of specific object classes, this technique called image generation. Occasionally, some of these techniques are used successively to reach target of the one's research. Hence, in order to extract the size and information from an image in a user-friendly way computer vision techniques have been selected in this thesis.

Motivation of this thesis is to develop a system that will create an end-to-end seamless tire purchasing experience with detecting the required information from an image of entire tire. It is a critical step for the online tire purchasing process to determine the information required for the purchase of tires in vehicle tires that have many signs and inscriptions on them, as required by the regulations, through an image that the users can take with their mobile devices, including the entire tire. However, due to the rubber structure containing carbon black in the compound of the tire (Donnet, 1998), detection of the black texts and symbols on a black background becomes challenging. Furthermore, the main difficulties that are aimed to be overcome with this motivation can be given as follows:

- The markings are reversed or angled due to the arc distribution caused by the position and outformation of the tire.
- The camera angle of the images that the users will take with their mobile devices are varying.
- The ratio of the area of the searched objects on the entire image to the entire image area is less than 5%.
- Every brand uses its own design approach, so each tire pattern is unique due to font of markings and decorations.

The main objective of the thesis to detect the necessary size and brand information for purchasing a tire. The side objectives can be given as to warn the user when taking an unsuitable image, transmit user-friendly messages for the customer to correct the usage mistake, and to process all this service in the optimum inference time, in order to achieve a better customer experience.

Methodology of the thesis to solve the problem of determining the brand and size of the is briefly that create an inclusive dataset, prepare, and clean the dataset for selected methods, search and implement the appropriate computer vision algorithm pipeline to perform analysis and detection and lastly compare the performance of the methods to conclude the studies in aspects of accuracy, inference time and usability. The created dataset will be divided into training set, validation set and test set, and deep neural network based supervised learning models will be trained with the training set. With the validation set, hyperparameter optimizations of these models will be carried out and performance comparisons and interpretations will be performed on the test set. In order to determine the best method, the appropriate deep learning and image processing algorithms to be used will be searched by making analyzes on the dataset. State of the art methods whose performance will meet the expectations will be investigated within the pipeline, which will consist of preliminary image processing steps, classification, segmentation, and object detection methods. By using various methods deemed appropriate as a result of the research, their performances will be compared and reported with the test set, which was previously created by random selection and covering the entire sample universe. In order for the thesis to reach its unique goal, a method consisting

of models that will successfully detect tire brand and size information will be developed over an image taken with mobile devices.

The main result expected from this thesis is to develop and deploy an algorithm system that can detect the brand and size of the tire in the image with high performance and with an optimum inference time from an image that the customers will take with their mobile devices, including the whole tire. In addition, there is another expected result of this thesis to contribute to the literature by eliminating the deficiencies of similar studies in the literature and to identify and report the development suggestions for the future work.

The thesis is divided into five chapters and conclusion. After introduction, chapter 2, description of tire and its markings, problem definition and literature studies are revealed. Chapter 3 explains basic technical background of the used methods. Next, Chapter 4 reveals the main reasons of need for creating new dataset and its structure and the computing environment as well. Next, Chapter 5 contains detailed explanation of the proposed system, its pipeline, and implementation. Then, Chapter 6 gives the results, performance, and comments about performance. Toward the end, conclusion and future work suggestions are given in the conclusion section.

2. PROBLEM DEFINITION

In this chapter, definition of tire and its markings, the need for automized detection and similar studies in the literature are described in following order.

2.1. Tire As a Product and Its Markings

Tire is the only equipment of a ground vehicle that contacts the surface. As outside visual, tire is round and black, but it consists of several components and different rubber compounds inside. Pneumatic tire means it is inflated by the air and the space between rim and tire filled by air pressure. The first tire which is pneumatic was patented in 1847 by Scottish inventor and engineer Robert William Thomson. After, John Boyd Dunlop developed the first pneumatic tire for bicycle usage and patented it in 1888. The first pneumatic bicycle tire which is given in Figure 2.1 is still exhibited in the National Museum of Scotland. After 1950's tubeless tires were used in the automobile industry.



Figure 2. 1. Pneumatic Bicycle Tire Produced by John Boyd Dunlop İn 1888 (National Museums Scothland, 2018)

The main role of the tires in an automobile is to provide the essential friction and grip between the tread of the tire and the surface of the ground. This composite structure is highly engineered to provide safety, handling, traction, and performance expectations of the customers. An average vehicle tire spins at least 500 revolutions per kilometers. Considering that the average life of a tire is 70000 km (Weissman et al., 2003), all the components of tire face more than 40 million loading-unloading cycles (Gent et al., 2006).

When all these are considered, the importance of the tire in transportation is once again demonstrated, so using the appropriate tire as well.

When choosing a tire, it is critical that the product to be selected is in the size recommended by the vehicle manufacturers. Vehicle manufacturers recommend the appropriate tire size to customers by making vital calculations and designs such as load per tire, grip, rolling resistance and handling dynamics. It is always recommended to choose the tire brand and size that comes on the vehicle as the first production (original equipment). For this reason, it is essential to know the necessary information in order to buy tires in the brand and size of this tire. Choosing the wrong tire size can be avoided without using it in conditions where the tire is not suitable for the vehicle rim. However, even this situation leads to loss of time and cost. Unsuitable tires that can be fitted to the rim may cause various undesirable situations by showing low performance values such as safety, excessive noise, excessive fuel consumption, longer braking distance, worse wet grip, and comfort. Therefore, obtaining the required tire size and tire brand information is the most critical step in purchasing the right tire. After this step, both purchasing from dealers and online purchasing processes will be completed perfectly.

The most important factor that users should pay attention to when choosing their tires is that the tire is suitable for the conditions of use and seasons. The precarious point in the following process is that for both vehicle and road safety, users must choose tires with dimensions determined by vehicle manufacturers. Users who make this selection must understand the information on the tire and choose the right tire size. The tire size includes information such as section width, aspect ratio, construction type and rim size. This information is determined and described by the European Tire and Rim Technical Organization (ETRTO), an authority in the tire industry (The European Tyre and Rim Technical Organization, 2018). The variety of tires which can be produced around the standards set by ETRTO exceeds thousands of combinations of section width, aspect ratio and rim size. In figure 2.2, nomenclature of the tire size is given. It can be seen that the illustrated tire size is 205/55R16. The first three character represents the nominal section width in millimeters for inflated tire. Section width is always expressed as a multiple of 5 and varies from 130 to 320 for passenger cars. After the separator "/" the following two character defines the tire section height as a percentage of the section width so the "55"

is called as the aspect ratio. Aspect ratio is always defined as a multiple of 5 and can take values from 30 to 95 for a passenger car. “R” typifies radial construction. Moreover, “B” or “-“ are used for bias construction. Lastly, “16” is the rim diameter size in inches. For rim diameter passenger car, it is always expressed as an integer and its value can vary from 11 to 23 inches. Truck and bus radial (TBR) group can take half values for tires such as “21.5”. Moreover, there are other standards for the tire markings such as Department of Transportation (DOT) marking for usage in United States of America. These regulations require that the tire size and brand must be written on the tire sidewall. In this case, it is guaranteed that the searched objects can be found on the tire, except in cases where it is wiped, wearied, aged, or covered with mud. This will also enable an expert to label size and brand objects within a data set. Thus, since the ground truth exists in each situation, it would be wise to seek help from artificial intelligence models trained with supervised learning over an image that will enclose the entire tire as a solution to such a problem.



Figure 2. 2. Nomenclature of the Tire Size

2.2. Tire Purchasing and Online Shopping

Rapidly developing technologies in recent years lead users to online shopping with their mobile devices. Indeed, tire has also become a product that can be purchased online. At

this point, the issue that confuses users is not only the performance or price of the tire, but also choosing the right tire size. Besides the difficulty of tire selection, the tendency of users to compare tire prices and performances encourages them to online channels. In 2018, 6.5% of the global tire market was estimated by Smithers' business-to-consumer (B2C) e-commerce channel as the fastest growing channel (Smitters, 2018). Another Market research supporting this prediction shows that according to simpletire.com, the largest tire e-commerce company, more than 10% of tire customers prefer to order their items online; but the number of tire customers who will order from an online channel is thought to rise when increasing number of customers search online for information before they buy (McPhee, 2016). Over the past decade, online shopping has experienced slow and steady growth. From 2008 to 2019, e-commerce grew by about 1% per year to reach 16% of total retail sales, an output expected by many authorities. When the COVID-19 effect emerged in 2020, changes in shopping habits resulted in outcomes that exceeded expectations in e-commerce channels. Bridgestone Global Digital Solutions assumes that over 90% of consumers start their tire shopping journey online. According to studies, repetitive steps and "heavy" digital interfaces and the lack of tire identification information by users are causing disruptions to the online experience, so digital solutions have begun to be designed and served to assist online tire buyers (Henry, 2021).

Considering this increasing mobility to online channels, it has been clearly demonstrated that systems consisting of user-friendly interfaces and messages should be established in order for customers to do their shopping more easily. Users who purchase tires with online methods sometimes do not know the size information clearly and this interrupts their e-commerce experience. To solve this problem, traditional methods such call center call line or chatbots constantly encountered with today's natural language processing (NLP) technologies are used (Cui et al., 2017; Gupta et al., 2015; Asadi et al., 2018). Still, all these methods do not provide a lean solution for the problem. For this purpose, the study, which is the subject of the thesis, has emerged. The perception of tire brands and sizes through tire photos taken by users with their mobile devices further reduced this risk. It is focused on enabling users to have the simplest and most accurate e-commerce experience on their own.

2.3. Need for Automized Brand and Size Detection

Thanks to the advancing level of technology, developing artificial intelligence algorithms and the high speed of Internet, most of the work one used to do manually has been automated such as bill payments, banking, shopping, mailing etc. In particular, shortening and rationalizing the selection processes that do not create added value has become an inevitable necessity. Considering increasing e-commerce and clear need of a smart system for seamless tire purchasing experience, it has emerged as a real need to automate a value-added selection process by developing an artificial intelligence-based system. As seen in Figure 2.3, this technology can be used to effortlessly detect size and brand information that can come in any orientation due to the structure and nature of the tire.

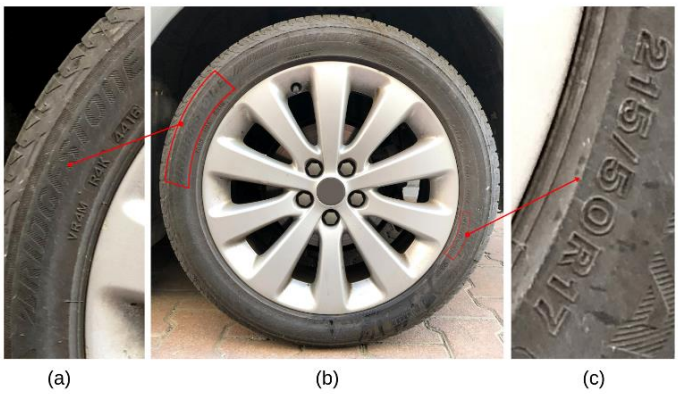


Figure 2. 3. Tire Brand and Size. (a) Tire Brand, (b) Tire, (c) Tire Size

2.4. Studies in the Literature

When the literature is examined for tire size detection, applications that use basic image processing techniques and deep learning-based techniques can be found. The history of the studies is mostly based on the last 4 years. This situation also shows how up to date the research area is. Still, in the literature, there is not a direct solution that detects tire brand and size on an image taken with mobile devices in a way that includes the entire tire.

In 2013, researchers at Hefei University of Technology developed an algorithm to decide the dynamic tire pressure by using image processing techniques (Chen et al., 2013). Even

though there is not a size detection in the study, authors use Hough circle detection (Hough, 1962) to obtain rim diameter and its center.

In 2013, a study on tire sidewall marking recognition was conducted by academics from Shenyang University of Technology. To extract the information from the sidewall of a tire, a moment-based method was presented. This method reports an accuracy of 94.7% (Xia et al., 2013). Study based on the idea of invariant of rotation, translation, and scaling Hu moments (Huang et al., 2010). However, experiments in the study do not cover the real vehicle tires which may be in thousands of variations due to usage, design, or lightening.

In 2019, Kazmi et al. developed deep learning based tire detection and text recognition algorithm with a stationary camera system (Kazmi et al., 2019). They proposed a system that detects the tire, unwrap the tire with inverse polar transform (Zhang et al., 2008) and uses a traditional object detection method which is Histogram of Oriented Gradients (HOG) (Freeman et al., 1995) combination with convolutional neural networks (CNN). The method detects the tire and reads DOT of moving vehicle using a stationary camera. They reported an accuracy of 95% for each classifier and also an inference time of 22 seconds on a Core i7 3.6GHz CPU and 2 seconds on a GTX-1080 GPU.

In 2020, Kazmi et al. improved their study in 2019 by adding the description of the cameras and their locations as dual camera/light housing towers (Kazmi et al., 2020). Authors do not improve the efficiency of the system in this study, they describe the product and the algorithm in more details and some add-ons.

In 2021, Gao F. et al. conducted a study to detect the tire text code by using deep convolutional neural networks (DCNN). In this study, firstly, the text parts are found with the DCNN model. By fitting the tire circle curve, the image is rotated to the counterclockwise orientation by determined angle. Then, a strategy of creating text boundary points based on each character's boundary points' generation method is proposed. Authors created a dataset and reported 94.5% detection accuracy (Gao et al., 2021). This system contains a stationary camera that perpendicular to the tire.

The above-mentioned research clearly shows the need of the detection the texts on the tire and have yielded successful results for reading the writings on the tire as well. The basic algorithms used inspire to solve similar problem. However, all these studies do not fully reflect the large sample population due to the limited datasets used. In addition, since it requires the installation of devices that see the tire from the opposite side, it does not fully cover the angle differences in the shots to be made with mobile devices.

3. THEORETICAL BACKGROUND

In this chapter, the brief technical background of used methods is given. First, supervised learning concept is explained. Second, artificial neural network (ANN) and deep neural networks are described. Third, deep learning-based object detection and segmentation algorithms and image processing algorithms used in thesis are provided with technical their backgrounds.

3.1. Machine Learning

Artificial intelligence (AI) as a concept has been in people's lives for a long time starting with science fiction movies to robotic improvements and computer science. AI is basically the concept given to the ability of any processing unit to take actions or decisions similar to human behaviors or decisions. As a sub-branch of AI, machine learning (ML) can be defined as a task force that establishes the link between input data and output without hardly programming to produce a particular outcome (El Naqa et al., 2015). In other words, ML is a computational journey which has an ability to learn through the input samples and produce accurate and robust outcome for the desired task. Learning could be outlined in this manner as extracting features and defining the dependencies from data (Cherkassky et al., 2007).

Machine learning is part of a broader family of AI methods based on learning data features. Learning can be semi-supervised, unsupervised, or supervised as shown in Figure 3.1 according to the nature of data. Supervised learning is aimed to approximate an unknown mapping between input and output from labeled samples. A human effort is needed to label the data. Whereas unsupervised learning only uses input samples without labels. Semi-supervised learning is a mixture of supervised and unsupervised learning which means some portion of the data is annotated. The annotated data is applied to infer the unlabeled part of the data in this case.

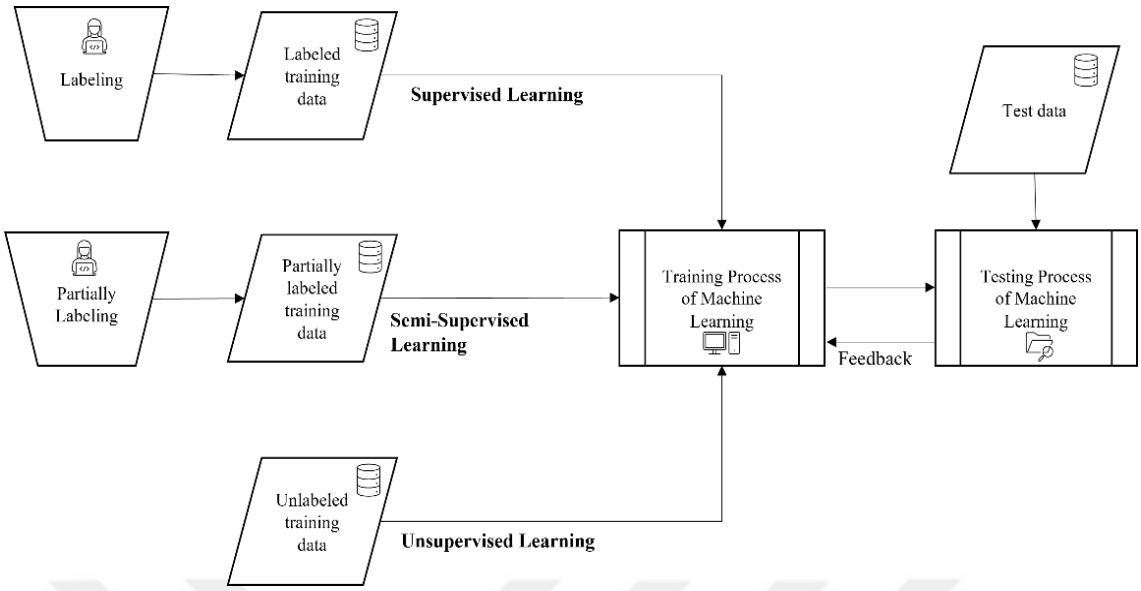


Figure 3. 1. Machine Learning Algorithms due to Training Data Type

3.2. Artificial Neural Networks and Deep Neural Networks

To understand the deep neural networks and their concepts better, it is crucial to start from the beginning of the artificial neural networks (ANN) history: the perceptron. Rossenblantt explains the way to get to know familiar with the mechanism of learning capability of organisms in 1958 and be the prior of the perceptron to use it for the first time in pattern classification task (Rosenblatt, 1958). Perceptron consists of one and only one neuron which can be trained according to data. It is also useful to know what constitutes the basis of artificial neural networks. The problem of minimizing or maximizing the target function with respect to the target value leads us to the optimization theory. Similar to that, perceptron tries to optimize the coefficients which are called weights. An illustration of perceptron can be seen in figure 3.2. It uses step function to calculate the output and decide the direction of the searching algorithm during the recalculations of weights by using the learning rate. The function which is used to calculate the output of a node is called activation function.

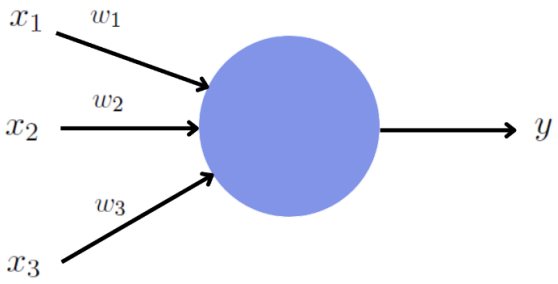


Figure 3. 2. Perceptron Model (Rosenblatt, 1958)

The delta rule is nothing but least squares method in order to re-specify the weights of the inputs to neurons in a single-layer neural network. The least squares method which was firstly published by Adrien-Marie Legendre in 1805 is one of the first algorithms to optimize the target function coefficients to reach the target value (Merriman, 1877). Using the delta rule as the learning rule was the beginning of the advanced feed forward neural network. When the first deficiency of perceptron was revealed by the failure to solve the XOR problem, various innovations followed each other such as Widrow's and Hoff's adaptive pattern classification machine (Widrow et al., 1960). It is also called Adaline. Adaline uses delta rule for learning algorithm. Even tough, it is too similar to perceptron, the main difference comes from the learning rule. In 1990, Widrow and Lehr came with a new concept of Multi Adaline (Madaline) which consist of two or more Adaline unit (Widrow et al., 1960). As mentioned before perceptron uses step function as an activation function this creates some undesired boundaries which do not allow the flex the solution environment to reach the best solution for objective like in the XOR problem. In 1969 Minsky, examined the perceptron and the activation functions (Marvin et al., 1969). They claimed that the perceptron and similar models could not find solutions to non-linear models. This explanation caused a damping on the research studies on artificial neural networks since the real-world problems consist of completely non-linear behaviors. Thanks to Rumelhart et al. with developed multilayer perceptron, XOR problem is finally solved (Rumelhart et al., 1986). Moreover, they mentioned the back propagation algorithm as learning rule. This affects positively to the neural network studies and be a milestone of the deep learning world.

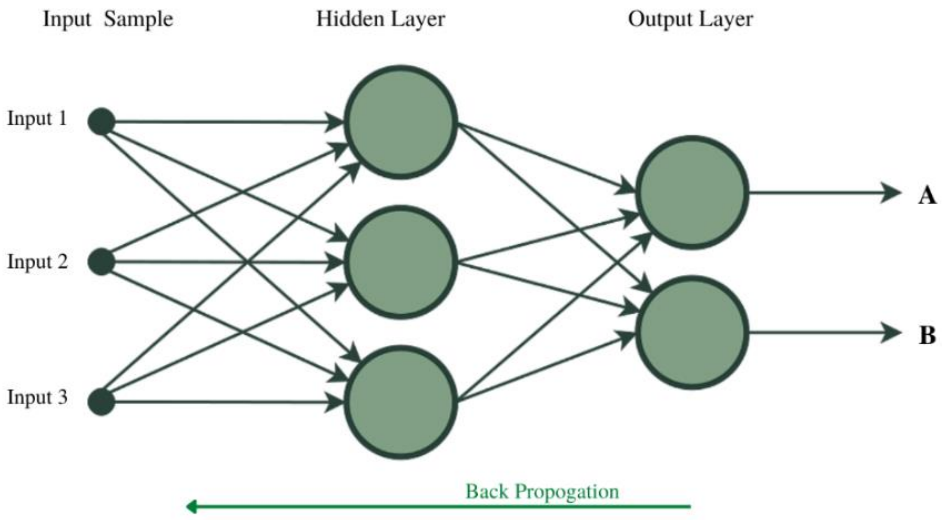


Figure 3. 3. Multi Layer Neural Network

A multilayer neural network consists of a hierarchy of neuron units which are organized in a successive manner. As it can be seen in Figure 3.3, a multi-layer neural network has more than two layers. The first layer is also called as input layer. It acts as a receiving layer for the input values given to the network. The last layer is also known as output layer. The results of the final decision are taken in output layer. Between these two layers, there are layers called hidden layers. When numbers of hidden layers start to increase, and ANN becomes deeper.

In ANNs, each neuron forms a weighted sum of inputs which are coming into the corresponding layer. After, neuron progress the resulting scalar value to a function which is called activation function. In order to speed up the training and overcome the generalization problem non-linear activation functions are used. In Figure 3.4, most used activation functions can be found. After that Equation (3.1) expresses Sigmoid activation, Equation (3.2) expresses Hyperbolic tangent activation, Equation (3.3) expresses Gaussian activation, Equation (3.4) expresses Rectified linear unit activation, Equation (3.5) expresses Leaky rectified linear unit activation and lastly Equation (3.6) expresses Sigmoid linear unit activation.

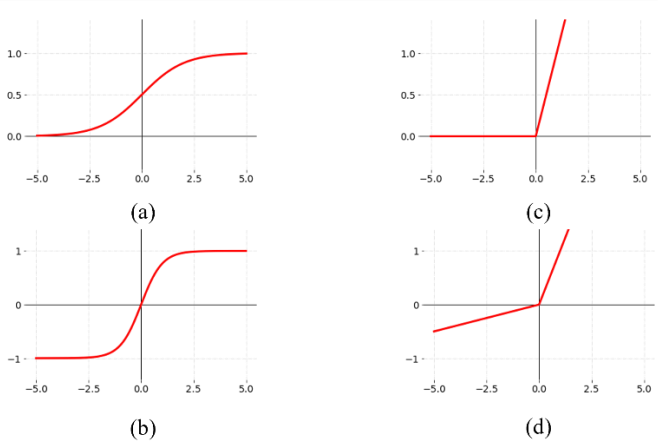


Figure 3. 4. Different Non-Linear Activation Funtions. (a) Sigmoid, (b) Hyperbolic Tangent, (c) Rectified Linear Unit, (d) Leaky Rectified Linear Unit

Sigmoid activation function is expressed as,

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{3.1}$$

Hyperbolic tangent (i.e., tanh) activation function is expressed as,

$$\tanh(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \tag{3.2}$$

Gaussian activation function is expressed as,

$$gaussian(x) = e^{-x^2} \tag{3.3}$$

Rectified linear unit (ReLu) activation function is expressed as,

$$ReLu(x) = \begin{cases} 0 & if\ x < 0 \\ x & if\ x \geq 0 \end{cases} \tag{3.4}$$

Leaky rectified linear unit (Leaky ReLu) activation function is expressed as,

$$Leaky\ ReLu(x) = \begin{cases} 0.01x & if\ x < 0 \\ x & if\ x \geq 0 \end{cases} \tag{3.5}$$

Sigmoid linear unit (Swish ReLu) activation function is expressed as,

$$Swish\ ReLu(x) = \frac{x}{1 + e^{-x}} \tag{3.6}$$

In ANNs, the main goal is to minimize the error in the algorithm by defining a lost function. A lost function states the distance between expected outcome and actual outcome of the ANN. The result of the lost function is used to calculate the gradients. The weights in neural network will be updated during the back propagation step of the training according to these gradient values by using gradient descent algorithms.

Cross-entropy-based loss functions which utilize the difference between two probability distributions mostly used in classification tasks. If the outcome layer has only two classes binary cross-entropy loss can be used. If the outcome layer has more than 2 classes, categorical cross-entropy loss can be used. Rather than classification tasks, if the main operation is regression, then mean squared error loss or squared error loss can be used as a loss function. Some of the loss function formulas are given below. In Equation (3.7) binary cross-entropy loss is expressed, in Equation (3.8) categorical cross-entropy loss is expressed and in Equation (3.9) mean squared error loss is expressed.

Binary cross-entropy loss,

$$L = -(y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)) \tag{3.7}$$

Categorical cross-entropy loss,

$$L = \sum_{i=1}^{M} y_i \log(\hat{y}_i) \tag{3.8}$$

Mean squared error loss,

$$L = (y_i - \hat{y}_i)^2 \tag{3.9}$$

Where; y_i is actual, $\hat{y}_i$ is predicted value. M states the total number of classes and L states the loss function.

Convolution is an operation of syndicating two signals to construct another signal in signal processing. In ML convolutional neural network (CNN) , is a feed-forward neural network which uses 2D filters or grid shape topology to analyze the input. As shown in Figure 3.5, a CNN has many feature-extractor filters. The main goal is to find the best filters to minimize the lost function. Hence, the network learns the filters which are able

to extract the features to achieve the desired task. Pooling layers is used to reduce the parameter size after the convolution layers. The first CNN architecture in the literature is known as LeNet-5 (LeChun, 2015).

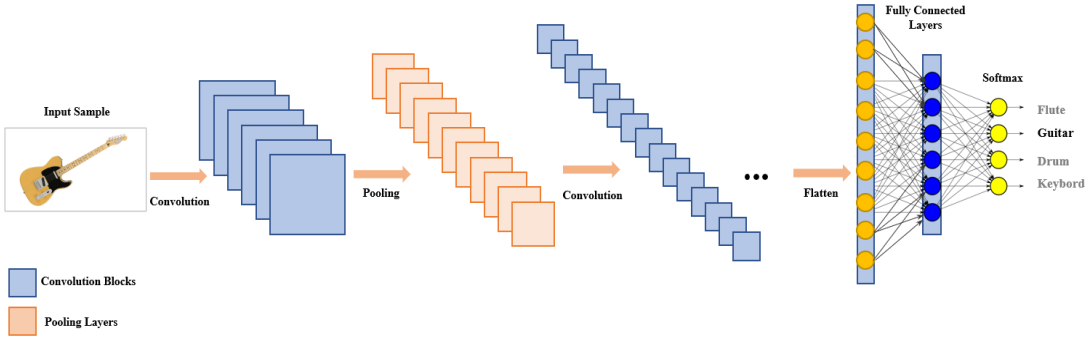


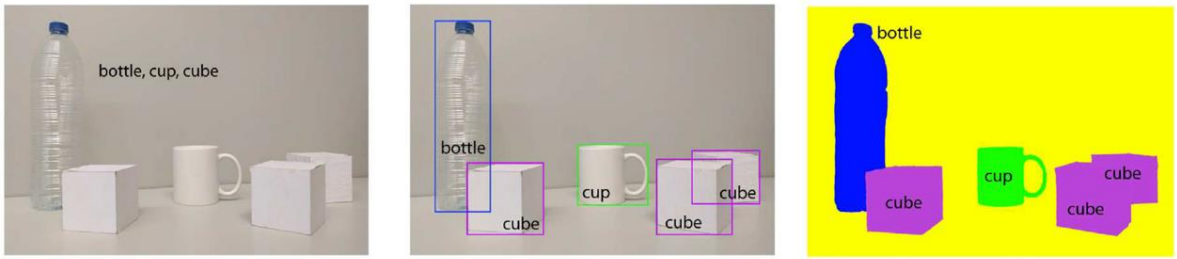
Figure 3. 5. Example of Convolutional Neural Network Structure

In 2012, Krizhevsky et al. proposed AlexNet which is a Deep Convolutional Neural Network (DCNN) with over 8 layers (first 5 are convolutional layers and last 3 are fully connected layers) with ReLu activation function (Krizhevsky et al., 2012). Its ILSVRC-2012 score is at top-5 error rates of 15.3% in the ImageNet competition. AlexNet was a great inspire of research world and encourage the people to dive into deep convolutional neural networks. In other words, deep learning era has been started after AlexNet published. After the breakthrough of AlexNet, CNN based architectures started to rise. GoogleNet (Szegedy et al., 2015) introduced the inception module by increasing accuracy. Then in 2015, ResNet (He et al., 2015) outperformed the competitors with residual connection architecture. ResNet brought solution to the exploding gradient and vanishing gradient problem which occurs due to derivative-based optimization algorithms in deep networks with ReLu activation function and skip connection mechanism.

3.3. Computer Vision Tasks

Computer vision is very popular due to its capacity of bearing efficient solutions to classical problems. It can be expressed as it is an interdisciplinary research area that deals with how machines can perform several tasks in order to extract higher-level meaning from digital images or videos. There are many different research topics in this field, but

the main areas used for this study are summarized as necessary in this chapter. Fundamentally, classification, object detection and segmentation are the main computer vision tasks. Even though all of them aim to recognize certain object in a given image or video, technical depth of the tasks increases from classification to segmentation. In Figure 9, it can be seen that classification only states the objects in a scene whereas object detection gives bounding boxes contains the particular object in the image with the corresponding class. Moreover, segmentation examines each pixel and decides the class it belongs to. In this way, it both finds the class of the object and detects the pixels in the image where the object is located.



(a) (b) (c)

Figure 3. 6. Different Computer Vision Tasks. (a) Classification, (b) Object Detection, (c) Segmentation (Garcia-Garcia et al., 2018)

3.3.1. Classification

Classification is one of the main tasks of the computer vision. Classification task tries to give an answer to the question of does the image contain any instance of a particular pre-defined object class. In deep learning CNN architectures are used in classification task successfully. Above mentioned architectures such as LeNet-5, AlexNet, GoogleNet and ResNet could be used for this task.

3.3.2. Object Detection

Object detection is one another favorite computer vision task. The main problem is to localize the object with the right class of it. Algorithms try to locate the presence of objects within a bounding box and indicate their classes in a given image.

After the great success of AlexNet in 2012, image processing-based methods have left their place to deep learning-based object detection methods. DCNN based object detection yields more precise results than traditional image processing-based methods. Object detection algorithms can be divided into 3 sub-categories as single stage, two stage and anchor free. Two stages object detection algorithms, firstly make region proposal for the object sought and in the second stage, it makes the classification. On the other hand, one stage detection algorithms perform the localization and the classification simultaneously. Lastly, anchor free detectors do not use anchors to search the objects. They detect an object bounding box as a pair of key points. These are left corner of top, right corner of bottom, or the center with using a single or several convolution neural network. Having detected the objects as a group of key points, anchor free detectors abolish the requirement for defining bounding boxes mostly used in single-stage object detection algorithms. Recently, vision transformer-based object detection methods are also used in computer several vision tasks (Khan et al., 2021). In Figure 3.7, object detection evolution milestones are shown in chronical order and sub-categories.

Pascal VOC dataset (Everingham et al., 2010) is the most popular benchmark for object detection. It is a traditional data set and organization that happens every year from 2005 up to present. Microsoft Common Objects in Context dataset (i.e., MS COCO) contains huge scale images with many objects in scenes for object detection challenges (Lin et al., 2014). MS COCO has 80 object classes and almost 1.5 million object instances. Considering the short extraction times, ease of training, and substantial performance of one-stage detectors, it was thought that it would be the right choice for solving the problem that is the subject of this thesis.

Unlikely two stage detectors such as Faster-RCNN (Ren et al., 2015), detection is one stage in YOLO (You Only Look Once) object detection algorithm (Redmon et al., 2016). YOLO family is thriving and famous object detection algorithm due to its low inference time, high accuracy and ease of use and deployment features. It detects all the objects in a single pass of the image. This mechanism yields reducing the computation time. The image is broken down into smaller grids of N by N, and if the central point of any object is detected then that particular grid is the deciding grid for the object detection. By doing so, YOLO detection algorithms yield faster than two stage detection algorithms. You only

look once (YOLO) overcomes this problem and enable to make all detection job in one stage. Redmon et al. published the study in 2016 and take into account object detection as just a single regression problem. The algorithm examines the regression problem thoroughly from image pixels to bounding box coordinates and class probabilities. As in the name of the algorithm states, there is only one stage to detect what objects in an image are.

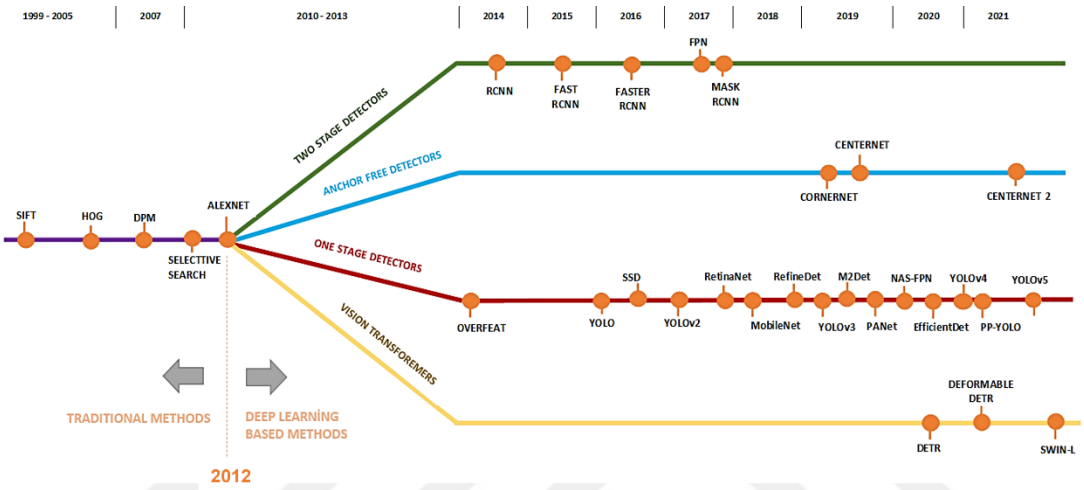


Figure 3. 7. Object Detection Taxonomy

YOLO9000 or YOLOv2 is published by Rodmen et al on 2016 (Redmon et al., 2016). YOLO9000's inference time was shorter than other object detection algorithms for a period of time on different datasets. It uses a joint optimization technique which is called WordTree to combine data from different domains to train instantaneously on famous data sets. Its name comes from the property of the algorithm which is able to detect for more than 9000 classes in real time. More, it uses batch normalization. The usual YOLO perform the training with a network at 224 × 224 and then up the image size to 448 × 448 for detection.

In 2018, Rodmen et al. published YOLOv3. Firstly, by the help of the traditional logistic regression, YOLOv3 predicts an objectness score for each bounding box. For classification, softmax is changed by logistic classifiers because of the claim that softmax's attribution to performance is insufficient. In the training phase, developers prefered binary cross-entropy loss for the class predictions. The logic behind the

bounding box prediction is to find the highest Intersection Over Union (IoU) belongs to object. (Redmon et al., 2018) The architecture of YOLOv3 is shown in Figure 3.8.

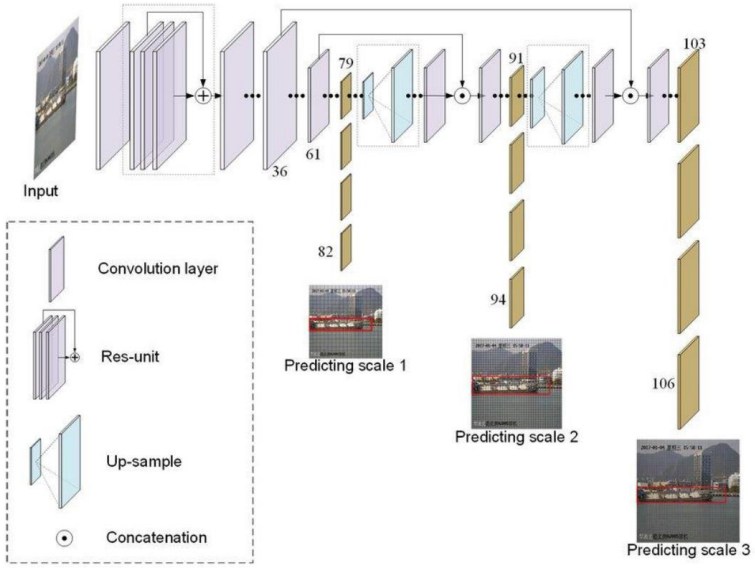


Figure 3. 8. Architecture of YOLOv3 (Redmon et al., 2018)

In 2020, YOLOv4 is published (Bochkovskiy et al., 2020). Authors introduce a new method of data augmentation, and Self-Adversarial Training (SAT). Optimal hyper-parameters are selected as applying genetic algorithms. Also, developers modified several existing methods to be ensure that if the design is enough for effective training and inference. They are namely modified SAM, modified PAN, and Cross Mini-Batch Normalization (CmBN).

In 2020, YOLOv5 models pretrained on the COCO dataset published by Ultralytics as open source (Ultralytics, 2020). YOLOv5 architecture mainly consists of three parts. It uses, like other DCNN based object detectors, a backbone algorithm. For YOLOv5, CSPDarknet is used for backbone. For the neck, PANet (Wang et al., 2019) has been used and for the head specified Yolo Layer has been used. The architecture can be seen in Figure 3.9. For feature extraction input data passes through the CSPDarknet and suddenly goes to the PANet for feature fusion. At the end, object class, object location and detection score are calculated in the Yolo Layer.

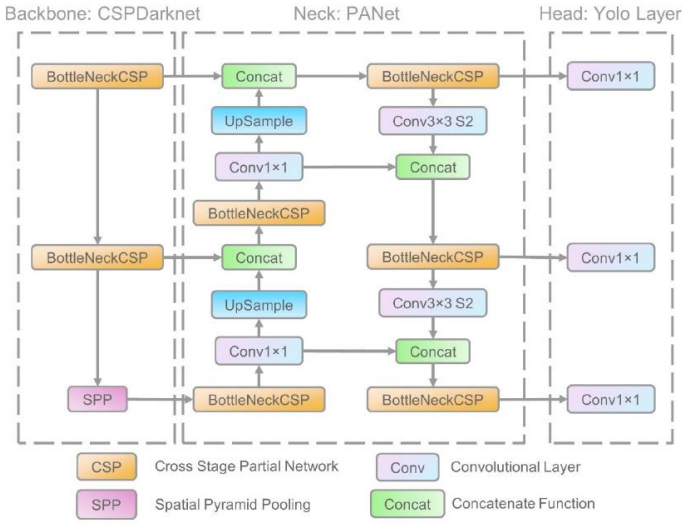


Figure 3. 9. Architecture of YOLOv5 (Xu et al., 2021)

In Figure 3.10, YOLO family evolution has been summarized in the chronological order with the key points of them separately.

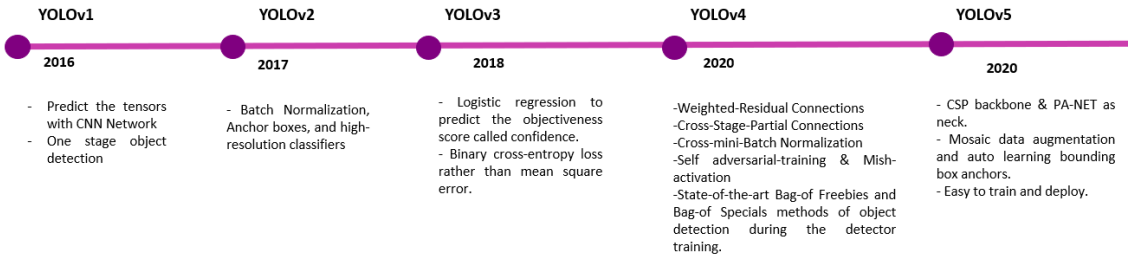


Figure 3. 10. YOLO Family Evolution Timeline

3.3.3. Segmentation

Segmentation which is a vital task of computer vision involves partitioning images or video frames into multiple objects and segments (Szeliski, 2010). In computer vision, image segmentation is the process of clustering the pixels in a given digital image. In order to state the location of the objects and their outlines for a given scene, image segmentation is utilized, and objects are specified as lines, curves etc. Segmentation can be divided into 3 groups as the nature of the detection: semantic segmentation, instance segmentation and panoptic segmentation. Semantic segmentation is detecting class of the

object for every pixel. Instance segmentation is performing a segmentation for every pixel in given image and classify the pixel to a belonging instance of the object. Panoptic segmentation is a mixture of instance and semantic segmentation. Furthermore, this computer vision technique is especially used in healthcare industry, aerospace applications, pedestrian detection, autonomous cars, augmented reality, and film industry.

In the literature there are many image segmentation algorithms both based on classical image processing, machine learning and deep learning methods. For traditional image segmentation algorithms, thresholding (Otsu, 1979), k-means clustering (Dhanachandra et al., 2015) and watershed methods (Najman et al., 1994) can be given as examples. During the era of DCNN, deep learning-based segmentation models started to appear. In 2015, Fully Convolutional Networks (FCNs) are proposed. It was a milestone in deep learning based semantic image segmentation models because of its high performance (Long et al., 2015). FCNs provides only convolutional layers, which yields to a clustering output with dimensions of the input sample. Still FCNs model is computationally expensive for real-time applications. Thus, encoder-decoder based models became popular. In 2017, SegNet is proposed (Badrinarayanan et al., 2017). It is an efficient segmentation algorithm that contains fully convolutional encoder-decoder architecture. In figure 3.11, SegNet's convolutional encoder-decoder architecture is illustrated. The loss of fine-grained image information is one of the challenging barriers of encoder-decoder based models. The reason behind is nothing but the loss of information through the encoding process as resolution.

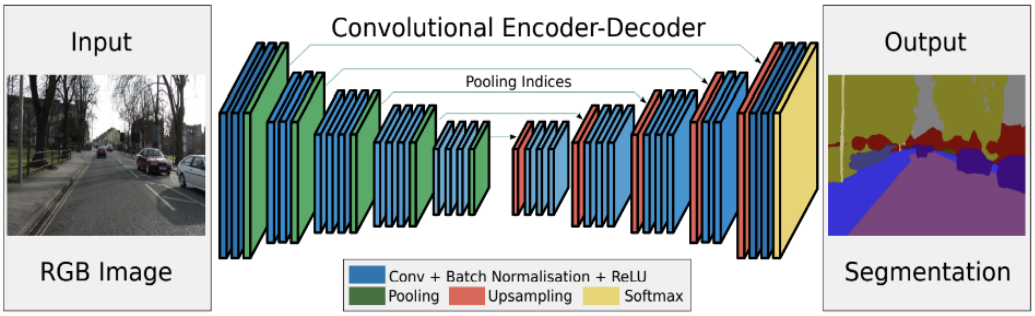


Figure 3. 11. The Architecture of SegNet Image Segmentation Model (Badrinarayanan et al., 2017)

For medical uses, U-Net was initially developed for biomedical image segmentation (Ronneberger et al., 2015). U-Net model's structure can be defined as two separate parts. They are a incubating flow to seizure the context, and an equal growing flow that make successful localization possible. Although U-Net is recommended for biomedical image segmentation, it is used in many different fields with its simple and effective structure. Therefore, the U-Net model was preferred for this thesis.

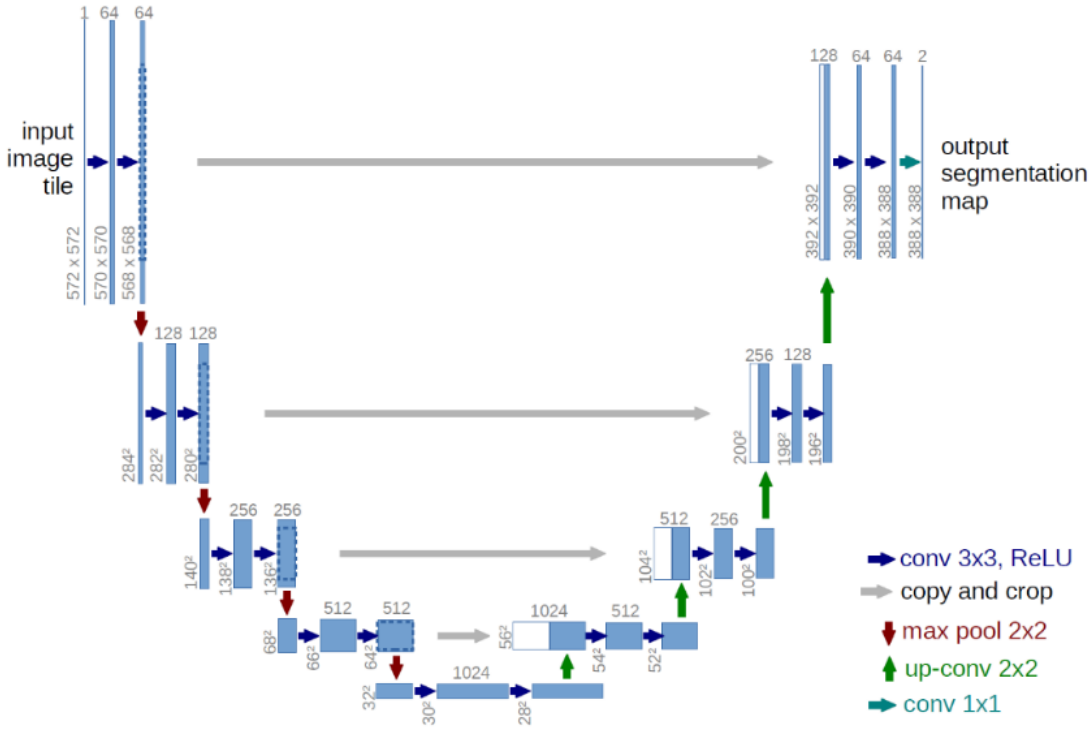


Figure 3. 12. The Architecture of U-Net (Ronneberger et al., 2015)

The network architecture of U-Net is illustrated in Figure 3.12. Its left stream is contracting path and right stream is a growing path. The downstream is just the usual structure of a CNN with 3×3 convolutions and max pooling operation with stride of 2. Its activation function is ReLu. Every down-sampling step, the number of feature channels are doubled. At each step in the expansive path an up-sampling activity takes place. 2×2 convolution that halves the number of feature channels is applied with a concatenation. At the last layer convolution is 1×1 in order to address each 64-length feature vector to the wanted class number. There are 23 convolutional layers in the whole architecture (Ronneberger et al., 2015).

4. DATA SET AND PERFORMANCE METRICS

In this chapter, data set creation methodology to develop the proposed system is explained. Next, used tools, frameworks and computer environments are given. Lastly, the performance metrics are provided.

4.1. Data Set

With the intention of create an uninterrupted customer experience, it is aimed to develop a system where users can easily and effortlessly access the suitable tires for themselves. In line with this goal, the problem is clearly defined as "customers cannot find the necessary information to buy tires from the sidewall of their tires" according to the statistics which have been obtained from the Brisa’s call centers. The system mainly detects the tire size and brand information with a pipeline that contains image processing, image segmentation and deep learning-based object detection algorithms. The computer vision system developed to find a solution to this problem is principally based on supervised learning algorithms. Therefore, to train and test these algorithms, a large enough data set that represents all real-life situations is needed. Before that, it is necessary to analyze a large data set in order to recognize the problem universe and to prevent possible errors. For the reason that a data set consisting of images under real-life conditions needs to be created. In this case, it is necessary to consider different conditions such as shooting angles and lightening, as well as the features of the users' mobile devices. In addition, it should be ensured that the system should be user-friendly by considering all kinds of images that users can take. For instance, the user can take a photo without a tire, take an image with an obstacle in front of it, or shoot an image that includes an irrelevant part of the tire. In such cases, a comprehensive data set consisting of images from real users has been created in order to make appropriate referrals. In short, the data set should be created with the images of the users who will use the application when such a system is implemented, rather than the researchers involved in the study, so that it reflects the user's behavior and preferences. In Figure 4.1, unwanted image conditions are given in 9 different categories: not a tire in the image, unbalanced light and sun flakes, an object in front of the tire, distant to camera, horizontal disorientation, tread view of tire, dual tires in the image, vertical disorientation, not whole tire in the image.

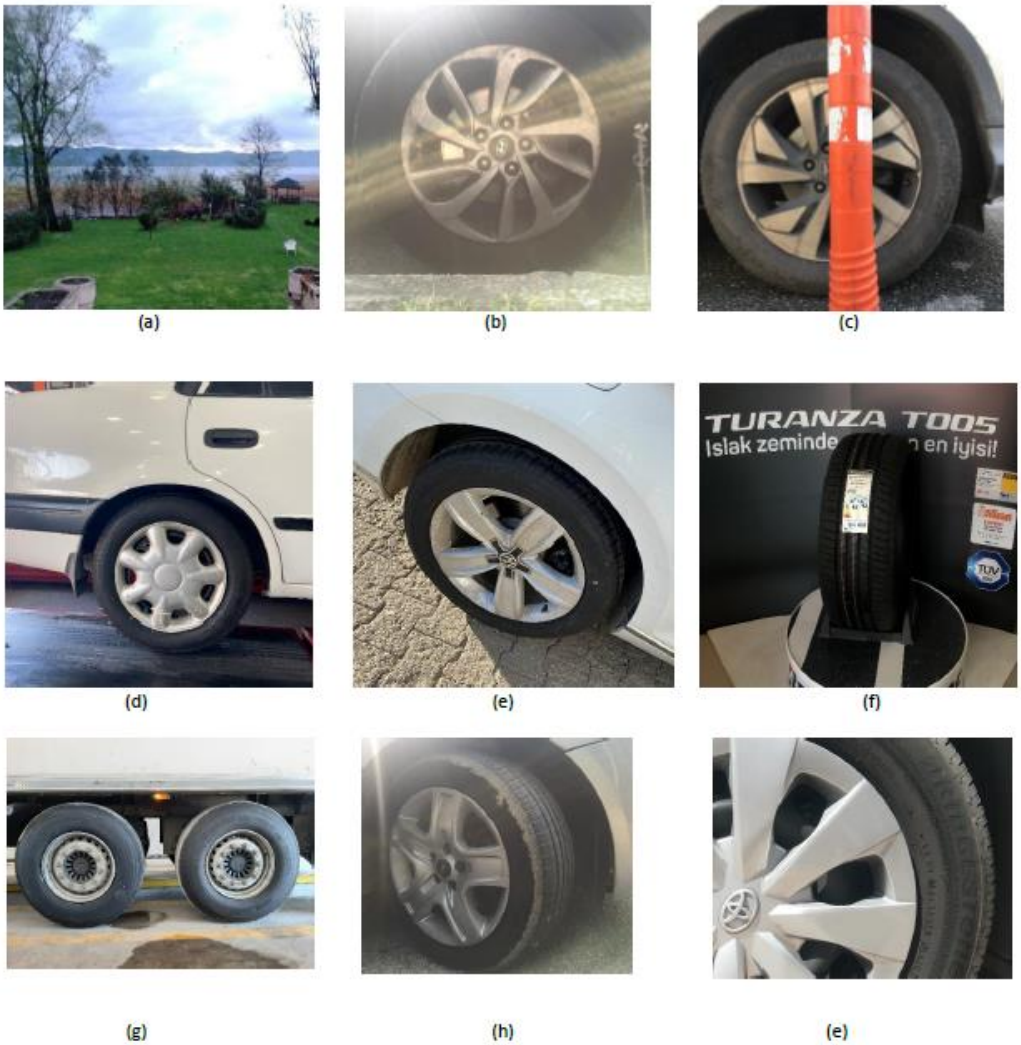


Figure 4. 1. Types of the Unwanted Images. (a) No Tire in the Image, (b) Unbalanced Light, (c) Blocked Object, (d) Distant Tire, (e) Horizontal Disorientation, (f) Tread View of Tire, (g) Dual Tires, (h) Vertical Disorientation, (e) Part of a Tire

The appearance of the tire varies due to usage conditions, the compound formulation preferred by the tire manufacturer, environmental conditions, raining, snowing, mud, dust, wear, stretches, ozone cracks and geographical conditions. On top of these, when the design preferences of different brands (font type, font size, decoration, locations of the markings) and the angle and environmental values of the image that customers will take with their mobile devices are considered, the complexity of the problem that is being tried to be solved increases considerably. In Figure 4.2, it can be seen easily that tire appearance, surface cleanness, color and ageing differ for every example.



Figure 4. 2. Tire Images in Different Conditions

So as to the algorithms to work more effectively, a large and unique data set has been created to cover tire diversity and complexity. In this direction, a project spanning 12 months with the support of more than 100 people from 20 different cities, exemplifying all regions of Turkey, in order to cover geographical changes, seasonal environmental conditions differences, customer's shooting preferences, and tire pattern and design diversity as summer pattern, winter pattern and all-seasons patterns.

4.1.1. Data Set Collecting

To create the data set, images were collected from 100 different users from 20 different cities for 12 months, covering all geographical regions of Turkey. During the period, approximately 50000 images were collected. From the stack of images, the ones belonging to 9 undesirable situations described above were removed. In this way, the data set to be used for training and testing was cleaned to cover all variations of the desired situations.

A gamification activity was organized among the users to create the data set for selected users. On this page, users both sent images and competed in a competition with pleasant

prizes as a result. In order to collect images from this gamification activity, the upload area was updated on the website used by the users. This upload system allows the multiple and simultaneous data upload. Selected users from all over the Turkey is able to upload the tire photos to train the algorithm from this system. System saves the photos to the directories with dealers' id number, date, and time. The important improvement here is to rename the files with MD5 hashing which gives a unique id to the image due to its original name and information. This allows to detect the fake and multiple file upload which is a serious threat to create such a data set. Because one can upload an image many times and this led the data set size to grow. Data is stored in the Microsoft Azure blob storage. A service has been developed to load the uploaded photos to the Microsoft Azure blob storage. This system gives a flexibility and sustainable data managing since dealing the data is crucial since the collection, labeling and training is done by different sources. System also backups the data for emergency cases. In Figure 4.3, data collection interface is given.



Figure 4. 3. Data Collection Interface

4.1.2. Data Set Preparation

Creating a dataset is one of the critical steps in the field of machine learning. An inclusive dataset that fully represents the problem and has a homogeneous distribution of situations is undoubtedly more efficient than a limited and biased dataset of unbalanced class samples. In this direction, a detailed data set creation strategy was followed in order to solve the complex problem, which is the subject of the thesis. The data set creation strategy is discussed under 3 main headings. The first is the elimination of unwanted images and the cleaning of the data set. Secondly, augmentation of the data set with the augmentation methods come after. The third is to label the data set in accordance with the

segmentation and object detection algorithms that will later be used in the pipeline model. Note that the third step not only includes bounding box labeling for object detection but also creating masks for the segmentation models.

In order to clean up the data set, the images in above mentioned unwanted conditions were deleted manually. Approximately, 5000 images are deleted. Afterwards, images with low resolution were removed. Afterwards, this data set was increased up to 300000 with data augmentation methods to cover the variations encountered in real situations. In this context, artificial data augmentation methods such as adding noise, cutting parts, elastic transform, rain effect, sun flake addition, snow effect and changing color channels in the Albuminations library (Buslaev et al., 2020) are used as given in Figure 4.4. This augmentation strategy is used to overcome the overfitting problem. The augmentation techniques that used are zoom out, rotation, drop out (i.e., cut out), adding random rain effect, adding random snow effect, adding random sun flake and light effect, adding random noise and HSV shift. The applied augmentation techniques can ensure that they reflect the diversity of images that may occur in real-life situations such as seasonal conditions, geographical conditions, usage habits and camera shooting angles at different times of the year. Augmentation methods were applied on the randomly selected images from the data set in such a way that the parameters are determined randomly by an algorithm.



Figure 4. 4. Data Augmentation Techniques

This dataset is labeled in 3 different ways for segmentation, brand detection and size detection models. The related labeling study was carried out by a team of 20 people with

1000 man-hours (62 man-days) effort, excluding controls. The first labeling strategy is labeling the rim and tire with circles. To do so, the masks for the segmentation model which is used for segmenting rim, tire and background prepared. In Figure 4.5, examples of first 2 strategies are given.



(a) (b)

Figure 4. 5. Tire and Rim Labeling for Segmentation Model and Creating Masks. (a) An illustration for Tire-Rim Labeling, (b) Several Examples of Masks for Segmentation Model After Labeling

The second labeling strategy is to enclose the regions with the size and the brand on the tire image, which is flattened by locating the rim center, in a rectangular box. In this way, the regions where the size and brand are located will be labeled, which will help the object detection algorithms to find these regions as if they are an object. Then, with the help of known coordinates, region of interest (ROI) will be taken from the images and input to the next classification or object detection algorithm will be created. The third labeling strategy is performed on the ROI's that are yielded as a result of the output of the object detection algorithm trained by the data stack formed as a result of the second strategy. Here, the alphanumeric characters that make up the tire size are individually labeled as objects. There are 12 total classes, including "0-9" characters, "R", "/". Figure 4.6 provides examples of these two labeling strategies. All tags, including the first strategy, are stored in XML format, and backed up.

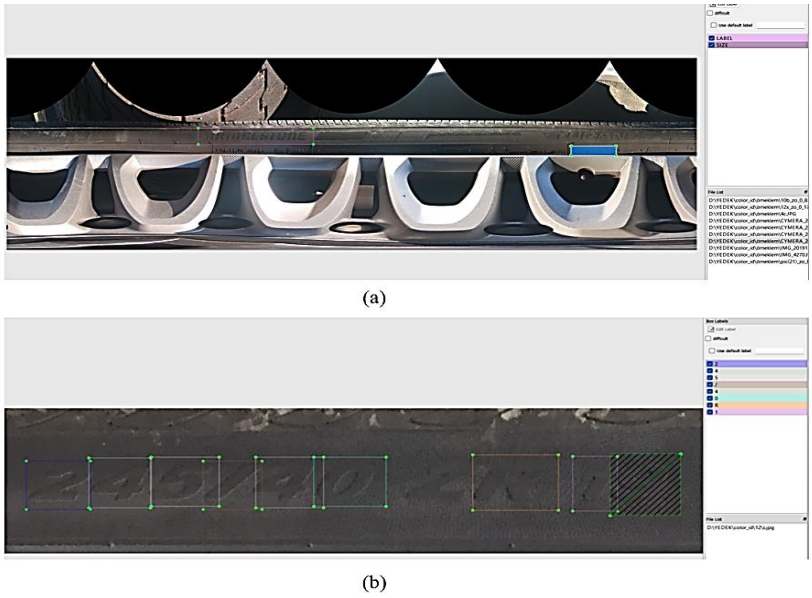


Figure 4. 6. Labeling for Object Detection Methods. (a) Brand and Size Location Labeling, (b) Character Labeling for Size Detection

4.2. Computing Environment and Tools

To train and test the algorithms powerful computers are used since computing power was a limitation for the studies in the literature. Computing environment consists of two different configuration CPU and GPU and two different framework software namely OpenCV-DNN and TensorFlow. The inference times of each step in the model flow and the total extraction time, without including the image loading times, are reported in Chapter 5. Afterwards, tests were carried out on 4 different hardware to evaluate its performance on different GPUs. In this way, it has been determined in which hardware and configuration the inference time works optimally for the developed system. The extraction times and total extraction times of each step in the model flow, without including the image loading times, are analyzed on 4 different GPUs, K80, P100, V100 and GTX 2080 TI.

4.3. Performance Metrics

Accuracy, precision, recall, F1 score are used for performance assessment. In order to understand these metrics, it is necessary to define true positive (TP), true negative (TN), false negative (FN), and false positive (FP). TP means correctly predicted positive values,

TN means correctly predicted negative values, FP means wrongly signifies the existence of a condition and FN means signifies the nonappearance of a condition when it is actually exists. To analyze the results, confusion matrix which maps these metrics can be formed. Accuracy means the percentage of true predictions over all data. In Equation (4.1) accuracy is formulated, in Equation (4.2) precision is expressed, in Equation (4.3) recall is expressed and in Equation (4.4) F1 Score is given.

Accuracy can be calculated as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4.1}$$

Precision states the percentage of the positive detected states are correctly detected. Recall, on the other hand, is a metric that shows how many of the actually positive situations we predict positively. Precision and Recall can be calculated as follows, respectively:

$$Precision = \frac{TP}{TP + FP} \tag{4.2}$$

$$Recall = \frac{TP}{TP + FN} \tag{4.3}$$

F1 score is harmonic average of precision and recall which can be expressed as following:

$$F1\ Score = \frac{2 * TP}{2 * TP + FP + FN} \tag{4.4}$$

Additionally, mean average precision (mAP) is calculated for the classification performance. mAP is defined by calculating the average precision of the classes and average precision is defined by calculating the area under the P-R curves. AP calculated as follows:

$$AP = \int_0^1 P(R)dR \tag{4.5}$$

5. PROPOSED SYSTEM FOR DETECTION OF BRAND AND SIZE OF VEHICLE TIRES

In chapter 5, it is mentioned about the proposed system that will detect the brand and size information required to purchase tires through an image taken with mobile devices. Word of detection is used for both localizing and classification for the entire system. First of all, the model pipeline, which includes all techniques, is given in general. Afterwards, the techniques and algorithms used in each step are detailed one by one.

5.1. Proposed Pipeline of The System

A pipeline has been studied in order to find the 9-character size information and the brand of the tire required from a randomly taken tire image. There are several reasons why the solution is a pipeline. These reasons can be listed as:

- Due to the round structure of the tire, the markings are scattered on the tire sidewall in the form of arc distribution. In other words, the objects (brand and size) searched according to the position of the vehicle and the tire can be found at any angle on a 360-degree arc. In this case, the object can be completely inverted, or it can be angled both vertically and horizontally, depending on the tire position and the shooting angle.
- The color of the tire is black due to the carbon black chemical in its compound, and the markings on the tire are found from the same material by methods such as embossing or engraving. In other words, black colored writings on black surface led to the low contrast in captured images.
- Since the developed system is planned as a customer application, it should have a user-friendly interface (UI) and excellent customer experience (CX). Thus, inference times and response times should be within what a customer can expect. In this case, instead of waiting for a single algorithm that will take a long time, it would be more accurate to have a structure that makes step-by-step checks and, if there is an inappropriate situation, it can give messages that will guide the customer correctly without running the rest of the algorithm pipeline.

In other words, there is not an algorithm or model that can extract and detect the desired information (size and brand) at once, on the contrary, several algorithms should be used

to achieve desired detection goal in a user-friendly way. To bear an efficient solution to this need, computer vision and deep learning methods are used together. The algorithm flowchart (i.e., model pipeline) is shown in Figure 5.1. The proposed model basically consists of 5 parts: preprocessing steps, tire and rim segmentation model, tire transform and background subtraction, tire brand and size location detection models and size detection model. The model processes a tire image taken with any mobile device as input. As a result of the necessary controls and determinations, it gives the tire brand class and 9-character size information as output. Since the usage areas of the model are mobile applications and websites, the time between uploading images into the model and returning results should be optimum. Therefore, performance of the whole system depends on both detection accuracy and detection time. Each part of the pipeline works sequentially, following each other. If an error is encountered in the previous step, it saves time and cost by not running the next steps.

As shown in Figure 5.1, the pipeline begins with an image entering the system. The first step, preprocess steps, checks the appropriateness of the input. These controls include checking the resolution of the image and checking for blurriness. At this stage, a callback is made with the corresponding error code and message for the unappropriated inputs and the rest of the algorithm is not run. Then, in the second stage, the rim, tire, and background are segmented with a deep learning-based image segmentation model. Thanks to this step, the rim center and radius values required for the transforming of the tire markings to be performed in the next step are determined. Also, the tire existence and camera angle control occur in second part of the pipeline. In the third step, the tire is flattened using the values from the previous step. At this stage, the tire is converted from the polar coordinate system to the cartesian coordinate system, so that the texts are always straight. Afterwards, parts of the rim and background are clipped from the image and discarded. Afterwards, the brand and size text locations on the obtained flat tire image are found with deep learning-based object detection algorithms. During this determination, the classification of the tire brand is carried out, and the tire brand is obtained as the first output. However, as mentioned before, using a classification algorithm for size determination will not be efficient. That’s because there are hundreds of combinations available on the market as suggested by ETRTO. It made more sense to train a separate object detection algorithm that detects alphanumeric characters for 9-character size

detection. The size area found in this step is cropped as region of interest (ROI) from the image and extracted as an input to the next object detection algorithm which is size detection algorithm. Finally, a specially trained deep learning-based object detection algorithm is run to find the tire size. Detection takes place on the ROI obtained from the previous step. Then, a 9-character size text is created by sorting over the coordinates of the bounding boxes. The estimated tire size is presented as the second output and pipeline is completed.

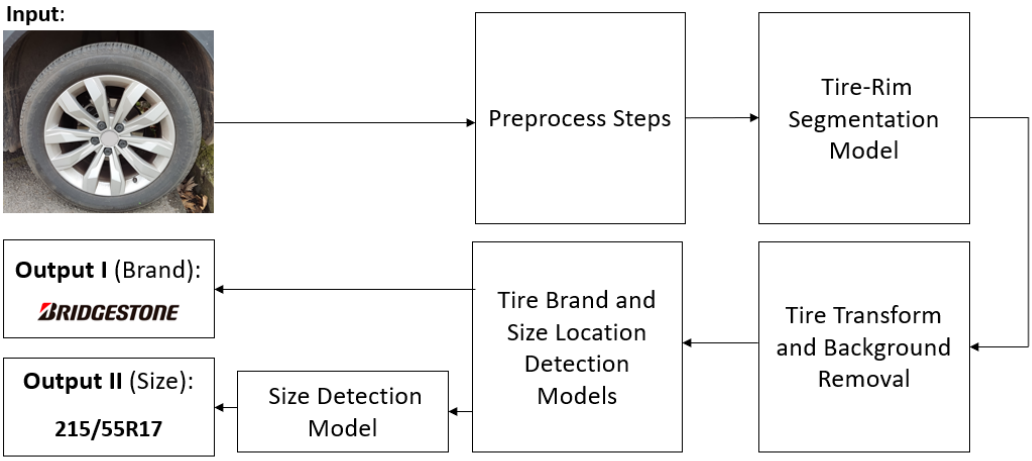


Figure 5. 1. Tire Size and Brand Detection Algortihm Flow Chart

5.2. Preprocessing

In the preprocessing step, the suitability of the image is controlled. These controls are file extension control, resolution control and blur check step, respectively. If the input image does not pass these checks, the rest of the model steps do not work. In this case, the user is informed and directed to get the right image to work in the model.

The first check is whether the file that will enter the system has an image extension. At this stage, it is checked that the file sent as input belongs to one of the following extensions: Joint Photographic Experts Group (.JPEG or .JPG), Portable Network Graphics (.PNG), tagged image file (.TIFF), bitmap image file (.BMP) and high efficiency image file format (.HEIC).

The resolution control checks that the image entering the model has the minimum number of pixels that will run the model successfully. The control mechanism warns the user if the image size is less than 1 megapixel and does not run the rest of the model. The main reason for this step is that the images coming from other applications that apply image spatial size reduction and compression techniques greatly reduce the performance in tire brand and size detection. The system is designed to work on images that are not over compressed. This control is based on file size for optimal runtime.

The next control is the blur control. At this stage, while the image is being taken, blurry images are detected due to various reasons and a warning is given because the algorithms in the following stages cannot perform accurate and sufficient inferences from blurry images. In this section, images taken under very light or very dark environmental conditions are considered in this category too. For blur detection approach, the input image filtered with a Laplacian kernel to find edges and then the variance in the filtered image is computed. The Laplacian operator moves the Laplacian kernel over the image to find edges on the x and y axes. By taking the variance of the output filtered image, the amount of edge transitions is checked. If the variance is lower than the threshold value, the image is considered as blurred. Variance threshold value is taken as 50 for this step due to the experiments on the dataset. In Figure 5.2, blur and darkness control is demonstrated. It can be seen that for the normal image the variance value is above the threshold but for blurred and dark image the variance value is lower than the threshold value.

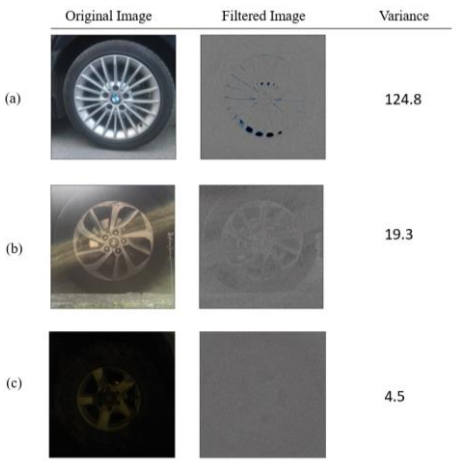


Figure 5. 2. Blur and Darkness Detection. (a) Standard tire image, (b) Blurred tire Image, (c) Dark tire Image

5.3. Tire Segmentation Model

At this step, tire-rim segmentation is performed. In addition, due to the results of the tire-rim segmentation, tire presence control has been done in this step. Due to the flexible structure of the tire mounted under the vehicle, it can never be displayed in a full circle form. The area in contact with the ground due to the load is narrower. When the user's shooting angle is added to this situation, the tire takes the shape of an ellipse geometrically. Therefore, if the Hough transform approach is used to determine the tire and rim boundaries, the center of the circle to be drawn around the tire does not coincide with the exact center of the rim, and the desired smoothness cannot be obtained in the tire image smoothing process to be done in the next step. It has been experienced that the flattened image versions obtained as a result of the transformation of the images from the center of the rim to the cartesian coordinates by inverse polar transformation (Zhang et al., 2008) by determining the center of the rigid rim in the wheel consisting of the assembly of the rim and tire, affect the algorithm performance positively. For this reason, the segmentation model for rim and tire is preferred.

With this model, the rim, tire sidewall part and the background image outside the tire are decomposed at the pixel level in the tire image. In this way, information such as rim radius, tire radius, rim center point is obtained. Furthermore, these outputs become inputs for the next step which is tire transformation and background subtraction. For the segmentation model, a deep learning-based segmentation algorithm which segments at the pixel level and has a reasonable inference time is used. The U-Net method has been preferred because it has many advantages such as combining both local and global features on the image, processing the image as a whole, and able to train high-resolution images. Hence, rather than advanced deep learning-based image segmentation models, a shallow and easy to use model selected. In the U-NET model used, the dimensions of the input images are set to be 256×256 pixel and single channel. The output is a 256×256-pixel image consisting of 3 classes of tire, rim, and background.

For training and testing half of the dataset is used. The reason behind that is to optimize the labeling cost and time. 20000 images have been used for training and 5000 images has been used for testing. In Figure 5.3, mask results of several examples are given.

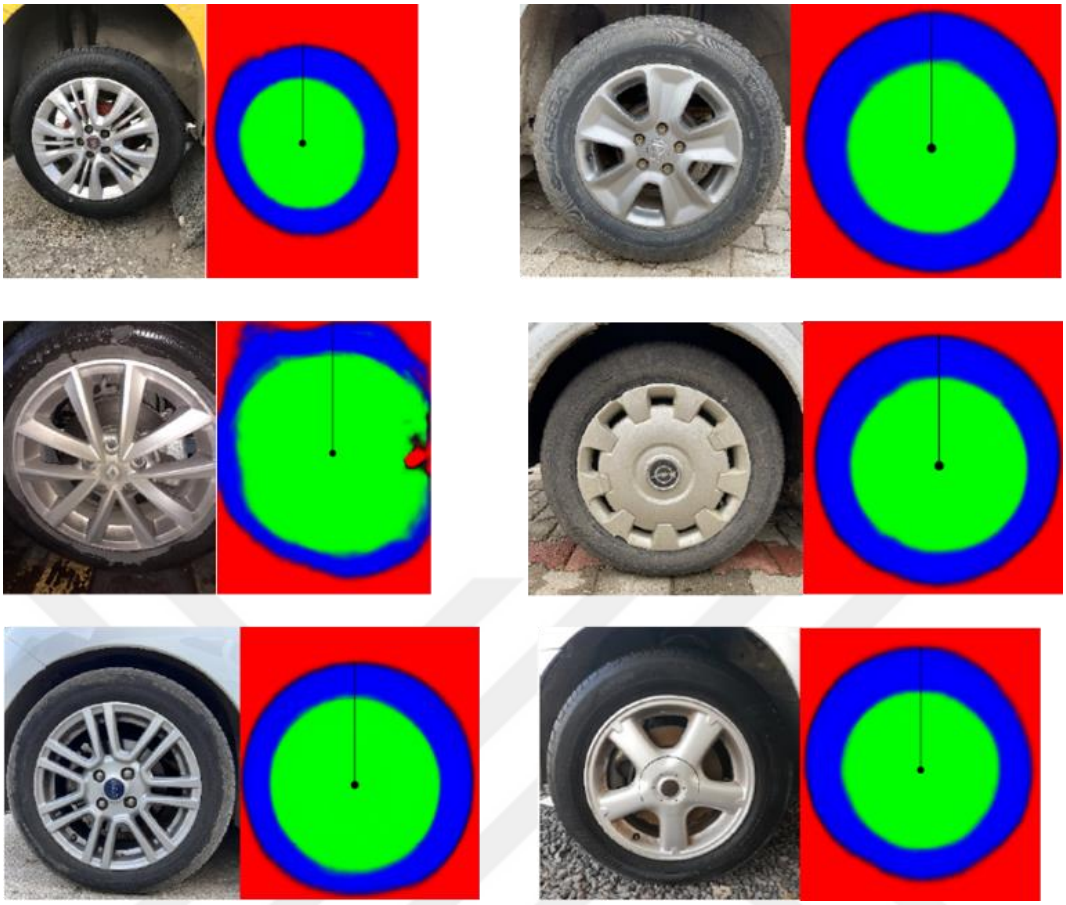


Figure 5. 3. Result Examples of Tire - Rim Segmentation Model

It can be seen in the result masks, green pixels belong to rim, blue pixels belong to tire and red pixels belong to the background. After gathering the result masks, rim center, tire radius and rim radius are found. In this calculation, there is a center of gravity of the pixels belonging to the rim compartment on the segmented output mask. Then, the points of the nearest and most remote tire compartment to this center of gravity are detected and selected as two radii. The nearest one is accepted as the rim radius and the distant one is tire radius. In figure 5.4, the output mask and rim center, rim radius and tire radius are given. c_{rim} states the center of gravity of the rim pixels after segmentation, in other words rim center. Having found the rim center location, the distance between rim center and furthest rim pixel to the center is found as r_{rim} (rim radius). Then, the distance between rim center and furthest tire pixel to the center is called r_{tire} (tire radius).



Figure 5. 4. Calculated Parameters from Rire-Rim Segmentation Model. (a) Input Image, (b) Output Mask with Center Location, Rim Radius and Tire Radius

The most important step at this stage is to check whether there is a tire in the image. If there is no tire in the image, the rest of the algorithm should not be run. Therefore, making this check in the first steps is necessary both for the CX design and for the correct operation of the system. It has been observed that performing a check on the already segmented image is also more efficient than detecting with a different detection or classification algorithm. In the first stages of the thesis, a simple CNN classification algorithm consisting of 2 classes, with tires and without tires, was used. However, it was decided that it would be more efficient to perform this analysis on the segmented image. At this stage, the algorithm basically compares the pixels belonging to the tire segmentation in the total image to the total pixels. It can be seen in Figure 5.5 that if there is a tire in the image, then at least 10% of the segmentation output mask belongs to the tire. Otherwise, there is no tire in the input image.

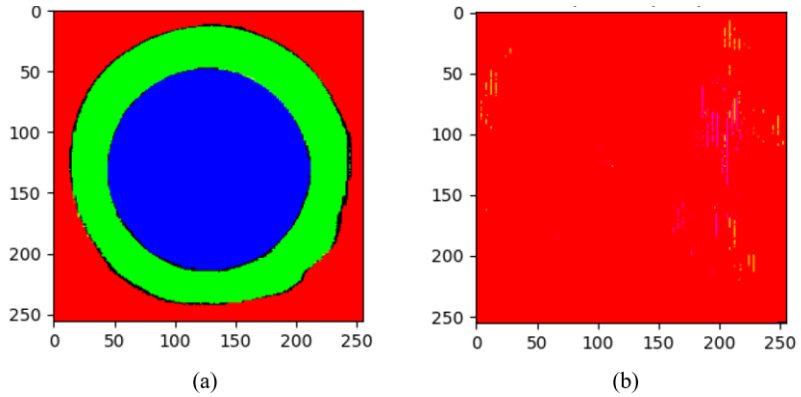


Figure 5. 5. Tire Existence Check on Segmentation Result Mask. (a) Mask Result of Tire Existance, (b) Mask Result of No Tire Existance

5.4. Tire Transform and Background Removal

One of the important challenges is the possibility of tire marking coming in at any angle in any part of the tire which is making it difficult both to identify and to label individual characters for training the model. The solution for this is the inverse polar transform method.

The rim center coordinates, and rim and tire radii calculated with reference to the outputs from the segmentation model are used to flatten the arc-curved structure of the marking in the tire images. Having known these tree inputs, inverse polar transform is performed. As a result of the inverse polar transformation, brand or size information may be cut in the middle since the transformation starts from an arbitrary point. In such a case, the searched object will not be found as a single piece in the image. Therefore, especially for the size determination model, not including the entire size problem may arise. In order to prevent such a situation, 10% of the total image was added from the opposite side of the left and right ends of the flattened image. In other words, 10% of the left side of the picture is added to the right side, and 10% of the right side is added to the left side of the picture. In the experiments, since the area of the size object does not exceed 10% of the area of the total image, the searched objects are not cut by adding this amount to the right and left margins. The parameter references of source image and transformed image is given in Figure 5.6 with some of the formulation. Inverse polar transform can be formulated as:

$$dst(\rho(x), \phi(x)) = source(x, y) \tag{5.1}$$

In Equation (5.1), “source” is source image and “dst” is destination image and x, y are Cartesian coordinates and ρ, ϕ are the polar coordinates.

$$\vec{I} = (x - x_c, y - y_c) \tag{5.2}$$

$$\phi(x) = Kangle\,.\,|\vec{I}| \tag{5.3}$$

$$\rho(x) = Klin\,.\,|\vec{I}| \tag{5.4}$$

$$Kphi = dsize\ height\ /\ 2PI \tag{5.5}$$

$$Klin = dsize\ width\ /\ Max\ Radius \tag{5.6}$$

Where x_c and y_c are transformation center coordinates. “Max Radius” is the radius of the outer circle to transform. “dsize height” and “dsize width” are destination image sizes. $\vec{I}$ is the distance vector. “Kphi” and “Klin” are the transformed lengths. In Equation (5.2) distance vector is formulated, in Equation (5.3) and Equation (5.4) are express polar coordinates. Lastly, in Equation (5.5) and Equation (5.6) state the transformed lengths.

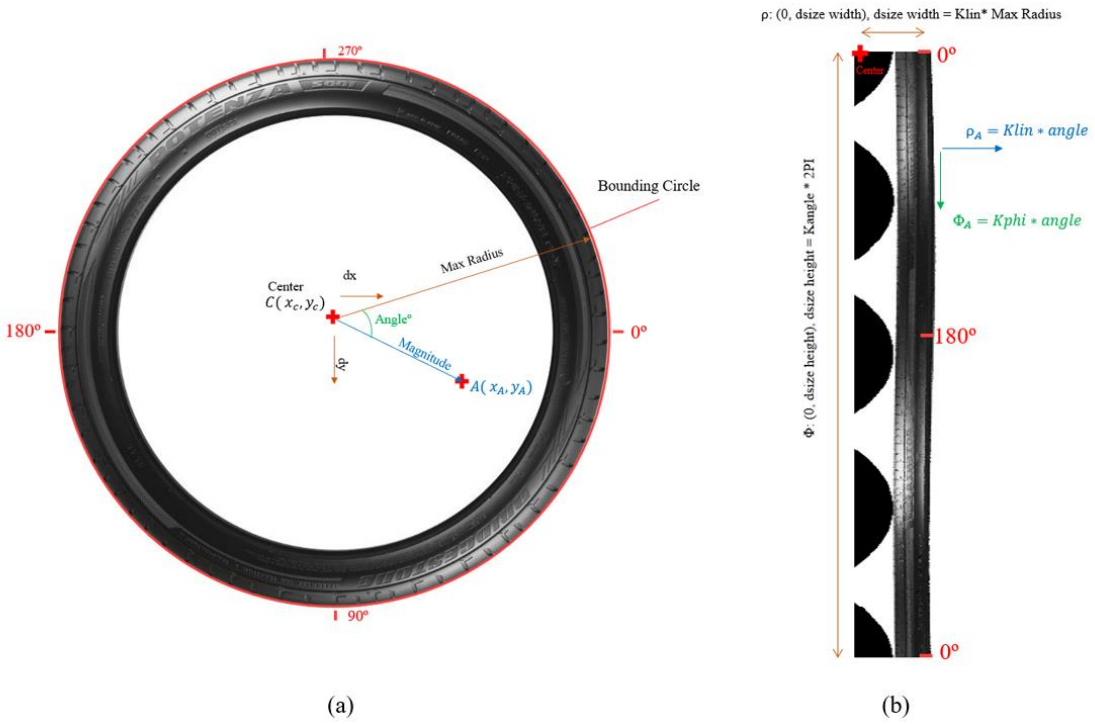


Figure 5. 6. Inverse Polar Remaps References. (a) Parameters References on Unwrapped Tire Image, (b) Transformed Tire Image and Parameters

Thanks to this transformation, the tire in an arc bend becomes a flat image. In addition, the radius values found in the previous step and the background and rim parts are clipped from the images. In this way, the writings are always come flat, and the rim and environment are extracted from the image to reduce the calculation load. In Figure 5.7, output example of transform is given.



Figure 5. 7. Output of the Inverse Polar Transform and Background Removal Example

5.5. Tire Brand and Size Location Detection Model

Hereby, in order to detect the brand and size texts, the positions on the flattened tire image must be determined. While determining the position of the tire brand, its class is also determined. However, an additional character detection algorithm is needed after the location is determined for size detection.

Before data augmentation for the brand detection model, 40000 labeled images were used for training, 9000 for validation and the remaining 1000 for testing. In the training process, the 9 most common tire brands in Europe and Turkey were selected as classes. For confidentiality concerns they will be named as brand 1, brand 2, brand 3 etc. All other brands were included in the "other" category as the tenth class.

For the detection model, YOLOv3 and YOLOv5 models were compared. During the test phase, the YOLOv5 model was preferred in brand and size detection models due to its higher performance scores and shorter extraction time. The performance comparison is given in chapter 5. The brand and size location detection model were trained on a dual RTX 5000 GPU system, 1000 epochs and 32 batch sizes, and took 22 hours. In Figure 5.8, brand detection outputs are given.



Figure 5. 8. Brand Detection Results Shown as Bounding Boxes

5.6. Size Detection Model

Size determination is made with a 2-stage object detection algorithm. In the first step, the locations where the size text on the flattened tire image are found with the YOLOv5 algorithm and cropped. As can be seen in Figure 5.9, size location is found on the flattened tire image as an object. Then, it is cropped according to bounding box's coordinates. Afterwards, size decision is given by the help of a second YOLOv5 algorithm trained with character-based training on the cropped image. In character-based training, there are 12 classes that represent numbers from 0 to 9 and characters of "R" and "/" . The entire original dataset of 50000 images was used as model training. There are 40000 images in the training set, 9000 images in the validation set, and 1000 images in the test set, and these numbers were multiplied by 6 thanks to data augmentation methods. The validation sets were used to optimize the hyperparameters in the learning models, and the test set was used for performance comparison.



Figure 5. 9. Size Location Detection Example

The size character detection model is trained with the early stopping criterion on dual RTX 5000 GPUs. Several results for both YOLOv3 and YOLOv5 detection results are given in Figure 5.10 below.

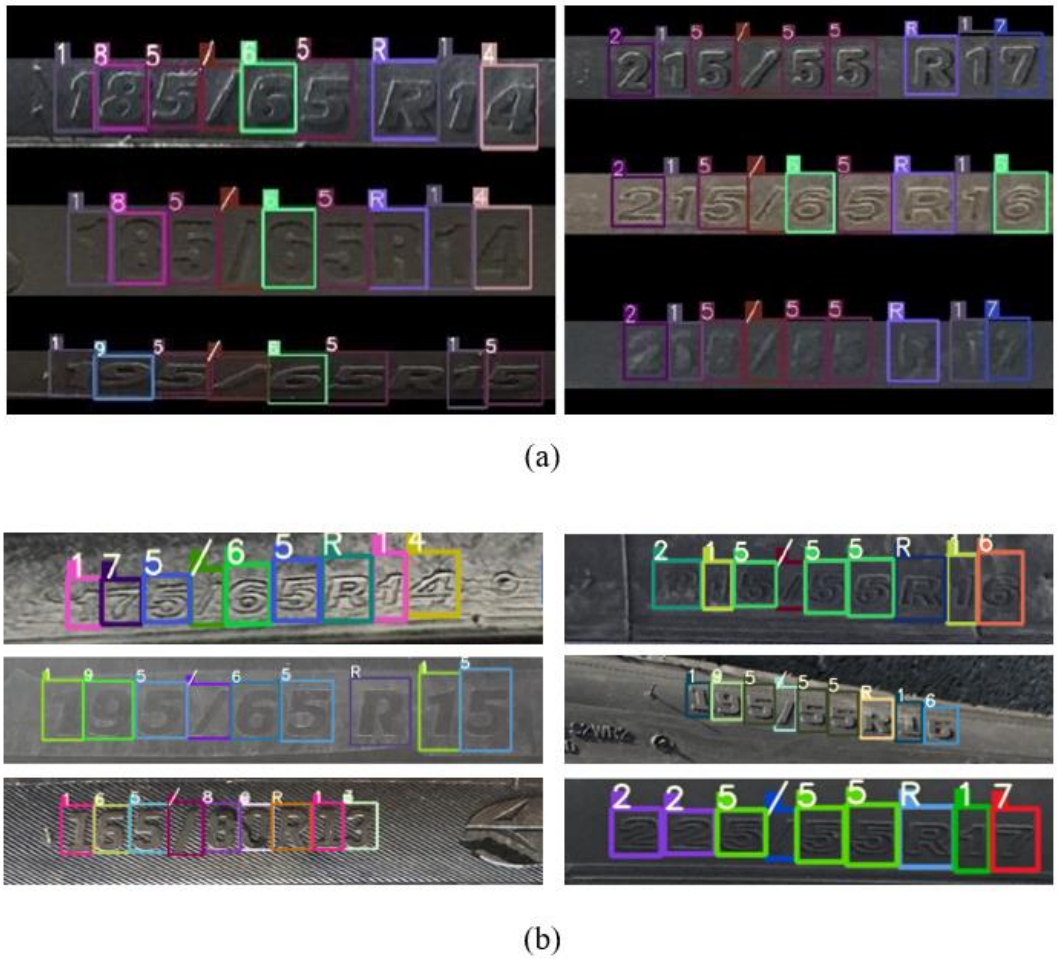


Figure 5. 10. Size detection Model Outputs. (a) YOLOv3 Results, (b) YOLOv5 Results

Finally, a system has been developed that detects the brand and size of the tire from a tire image with high performance. A flowchart showing the entire flowchart and its results is given in Figure 5.11.

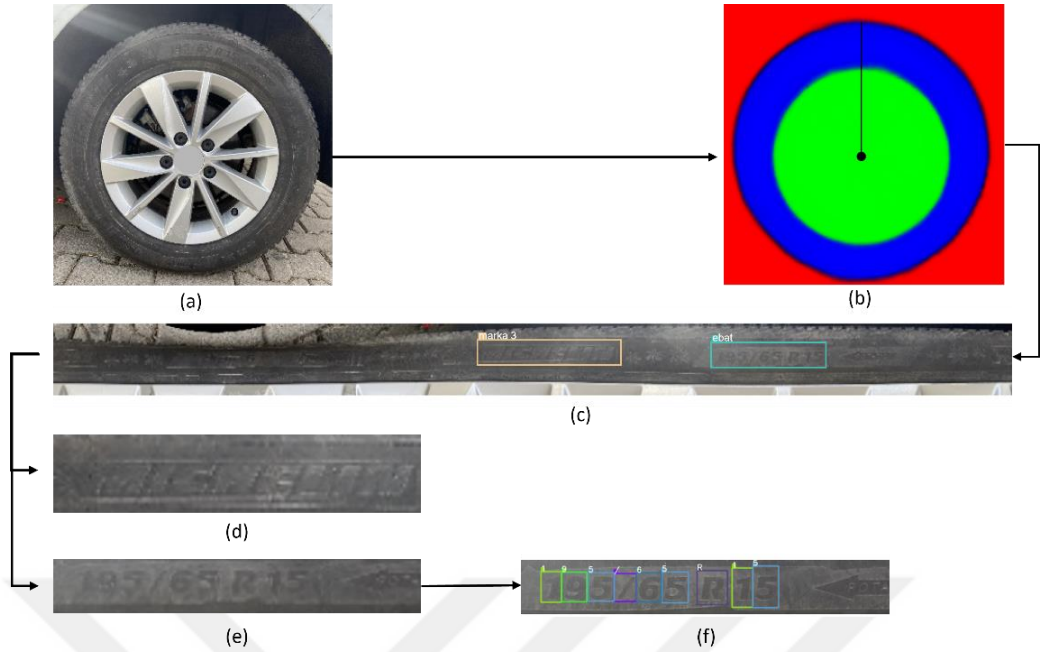


Figure 5. 11. Flow Chart of The System with Outputs. (a) Input, (b) Tire-Rim Segmentation Output Mask, (c) Brand Detection and Size Location Detection on Flatten Tire İmage, (d) Brand Detection, (e) Size Location Detection, (f) Size Character Recognition Results

6. RESULTS AND DISCUSSION

In Chapter 6, the performance results of the system described in the previous chapters will be discussed. The system, which is designed to detect tire brand and size information from the tire image, basically has two outputs. The first of these is the tire brand class, and the second is the tire size information. Performance comparisons are also grouped into 3 main groups: brand detection model results, size detection model results and inference time on different hardware's results. Afterwards, comments were made about the models compared according to the results, and the algorithms in the final model and the preferred hardware system were given. Lastly, all performance tests are reported on a test set of 1000 tire images that equally cover the data universe.

For the brand detection performance results, the YOLOv3 object detection algorithm and YOLOv5S object detection algorithms trained with the same hyperparameters on the same training data were compared on a test set consisting of 1000 images. Comparison of confusion matrices in Figure 6.1 shows that the performance results of YOLOv5 are higher for the detection algorithms given. Moreover, size location detection algorithm is selected as YOLOv5 according to its performance on brand detection. Another reason for selecting YOLOv5 as size location detection algorithm is not to make the system complex by selecting another algorithm. Size location detection algorithm has an accuracy of 98.6% with 0.6 IoU threshold on the same test set. IoU threshold hyperparameter for size location detection algorithm is decided through hyperparameter optimization tests. Our experiments show below 0.6 IoU threshold, multiple detection occurs and above 0.6 IoU threshold performance drops.

YOLOv3 has 97.3% accuracy, 96.9% recall, 96.7% precision and 96.8% F1 score performance metrics for brand classification. YOLOv5, on the other hand, has 98.6% accuracy, 97.9% recall, 97.9% precision and 97.9% F1 score performance metrics on the same test set. On the system with YOLOv3 dual RTX 5000 GPU the highest inference for brand detection time is 1.25 seconds (and 1 seconds for size location detection), while on the same system YOLOv5 has 0.19 seconds inference time under the same measurement conditions. In Table 6.1, these performance metrics are given in tabular form.

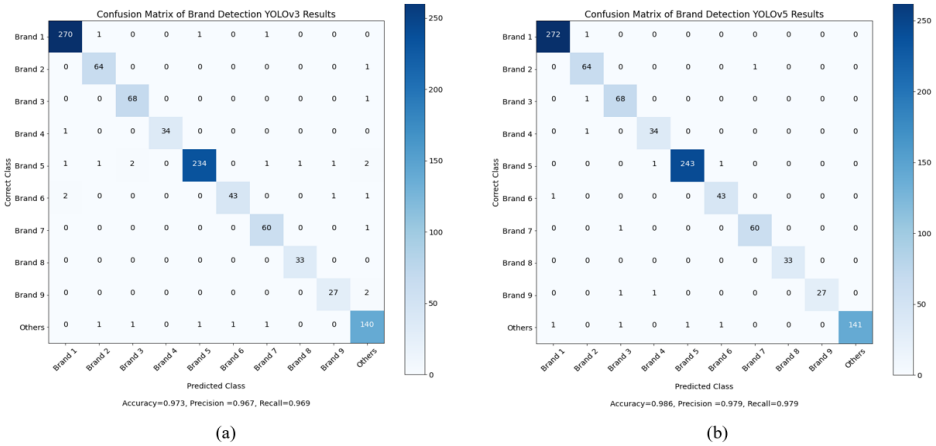


Figure 6. 1. Confusion Matirces of Brand Detection Algorithms on The Same Test Set. (a) YOLOv3, (b) YOLOv5

Table 6. 1. Performance Comparison of Brand Detection Model Algorithms

Metrics	YOLOv3	YOLOv5
Accuracy	97.3%	98.6%
Recall	96.9%	97.9%
Precision	96.7%	97.9%
F1 Score	96.8%	97.9%

A character recognition model is trained for the size detection model. YOLOv3 and YOLOv5 object detection algorithms are preferred for this model. For performance tests, a test set consisting of 1000 images used in brand detection performance tests was used. In total 9000 characters exist in the test set. The comparison of size detection performance needs to be examined under two groups: character detection performance for each character (0-9, "R", "/") and the final detected size value (for example: 185/55R15). The main reasons for examining it in two groups are that the studies in the literature have given the character detection results, and in fact, the correct comparison is whether the 9-character size information that is desired to be achieved can be determined correctly.

When YOLOv3 and YOLOv5 performances were compared for character detection, it was observed that YOLOv3's performance decreased especially in muddy, dusty, damaged and partially deleted characters. There are two hyperparameters has been decided with experiments on test set. They are confidence threshold and Intersection of unioun threshold. Confidence threshold is the minimum acceptable score of prediction to

consider the output as a TP. IoU threshold is the minimum acceptable union area between prediction boxes and ground truth for considering the output as a TP. In the tests, 0.50 was found suitable as the IoU threshold value for character detection with using an algorithm called non-maximum suppression to result best detected bounding box for a given object (Neubeck et al., 2006). When the IoU goes above 0.50, the detection performance of the system decreases. More, if IoU is selected below 0.50, there are multiple detections for the same object. Figure 6.2 is a graph prepared by taking the average of all class character classes for performance comparison with a IoU threshold of 0.5. In this graph, precision and recall values are plotted for both algorithms. As given in Figure 6.2, between 0.3 and 0.5 confidence thresholds at a 0.5 IoU value, YOLOv5 is superior to YOLOv3 in both precision and recall values. To maximize the F1 Score of YOLOv5, precision and recall results of YOLOv5 at 0.5 IoU are calculated. The confidence threshold that maximizes the F1 Score is calculated as 0.42. For this reason, confidence threshold has been selected as 0.42.

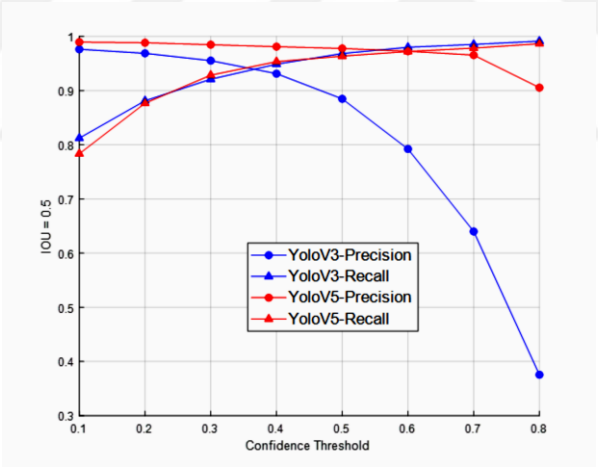


Figure 6. 2. YOLOv3 and YOLOv5 Performance Comparison for Character Recognition

Object detection method was preferred for character recognition of tire size consisting of a total of 12 classes, because the locations of the characters are also critical apart from classification. By ordering the centers of the bounding boxes coordinately, the tire size with 9 characters is obtained. As mentioned above, YOLOv3 and YOLOv5 were compared as object detection algorithms. Confusion matrices of both models are shown in Figure 6.3. If it is interpreted in general, it is seen that especially "1" and "7" characters, "5" and "6" characters and "9" and "8" characters are predicted incorrectly.

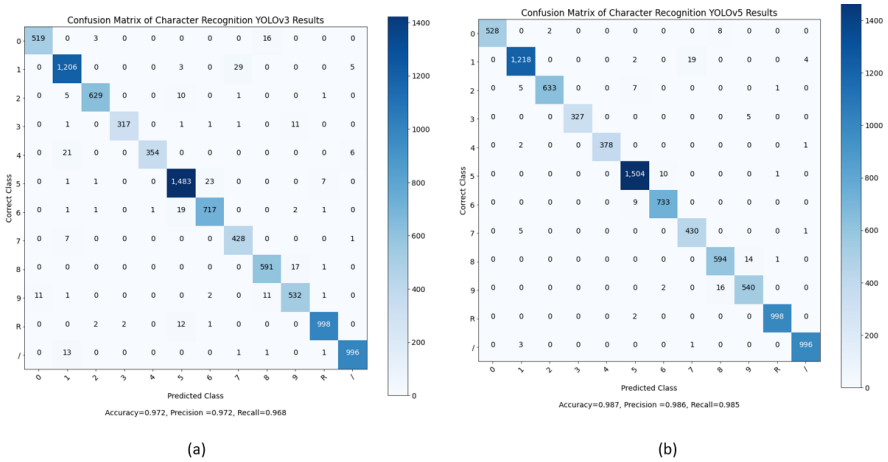


Figure 6. 3. Confusion matrices of size detection models for character recognition. (a) YOLOv3, (b) YOLOv5

YOLOv3 has 97.2% accuracy, 96.8% recall, 97.2% precision and 97.0% F1 score performance metrics for size character recognition. YOLOv5, on the other hand, has 98.7% accuracy, 98.6% recall, 98.5% precision and 98.5% F1 score performance metrics on the same test set. YOLOv5 has total 121 FP values. On the system with YOLOv3 dual RTX 5000 GPU the highest inference for brand detection time is 1.25 seconds (and 1 seconds for size location detection), while on the same system YOLOv5 has 0.18 seconds inference time under the same measurement conditions. In Table 6.2, these performance results are given in tabular form. YOLOv5 was preferred for the size character detection algorithm, as in the brand detection algorithm.

Table 6. 2. Performance Comparison of Character Recognition Model Algorithms

Metrics	YOLOv3	YOLOv5
Accuracy	97.2%	98.7%
Recall	96.8%	98.6%
Precision	97.2%	98.5%
F1 Score	97.0%	98.5%

Since the main purpose is to determine the size of the tire in the image, a performance comparison has been added in addition to the character detection performances. This comparison compares the ground truth of the size information of the tire in the image with the estimated size, as it would be in the real situation. Since there are more than 100 size

combinations in the dataset, instead of categorizing in this way, the distribution by tire rim size (13" - 22") is given in Figure 6.4.

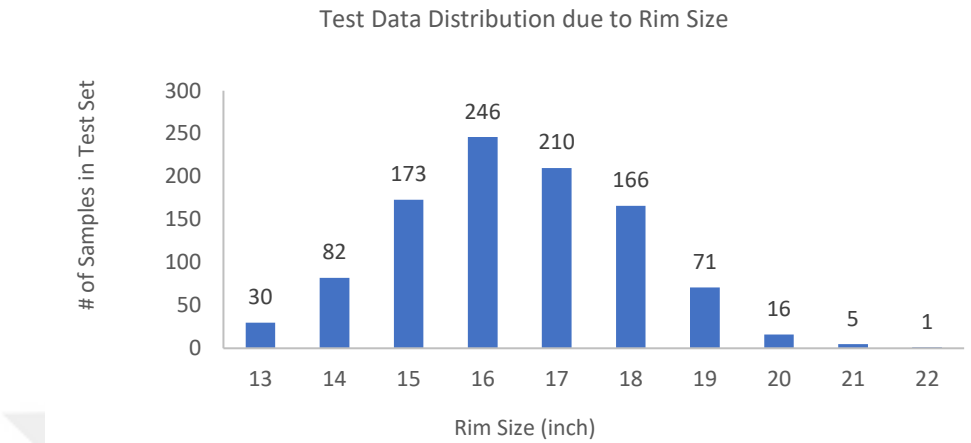


Figure 6. 4. Test Set Distribution due to Rim Size of the Tires in The Data Set

When the results of the test set consisting of 1000 images are examined, it is seen that the model using YOLOv5 correctly detects 965 out of 1000 total samples. In this case, the success of the model in detecting the correct size is 96.5%. Since most of the 121 False Positive values found in above mentioned character recognition model occur in the same images, the accuracy is satisfying. The reason why the character recognition performance seems higher is that the total number of samples is 9000 in character recognition test set and 121 errors in 9000 samples are proportionally small comparing to 35 errors in 1000 samples.

Next, the performance comparison according to inference time on different computing systems takes place. The system was tested on OpenCV-DNN and Tensorflow frameworks on two different hardware, CPU and GPU. The extraction times of each step in the model flow and the total extraction are broken down without including the image loading times. Intel I7-10750H is used as CPU and RTX 5000 is used as GPU. Other hardware features in the system are 64 GB DDR4 Ram and 5TB SSD disk. The comparison results are given in Table 6.3. As a conclusion, system with RTX 5000 GPU on TensorFlow has the most optimum inference time. Therefore, it was decided to run the system on the GPU.

Table 6. 3. Hardware (GPU vs CPU) and Framework (OpenCV-DNN vs TensorFlow) Inference Time Comparison on the Models

Hardware	Framework	Tire - Rim Segmentation Model Inference Time (secs)	Brand and Size Detection Model Inference Time (secs)	Total Inference Time (secs)
CPU	OpenCV-DNN	0.10	4.20	4.30
CPU	Tensorflow	0.15	4.42	4.57
GPU	OpenCV-DNN	0.14	2.23	2.37
GPU	Tensorflow	0.03	2.25	2.28

The system whose results are given serves as a web service that works with cloud computing using containers. The total inference times and inference time of each step in the model pipeline, without including the image loading times, are analyzed on 4 different GPUs, K80, P100, V100 and Dual RTX 5000 with YOLOv5 algorithms in Table 6.4. According to the table, Dual RTX 5000 GPU has shorter inference times than its competitors in preprocess, tire-rim segmentation, and tire transformation steps. The V100 GPU, on the other hand, outperformed the others in brand and size detection models. When the total inference time is evaluated, the Dual RTX 5000 outperforms all its competitors. Because of this performance, Dual RTX 5000 GPU configuration was used for the final system.

Table 6. 4. Inference Time Breakdown for All Steps of Pipeline

GPU	Preprocess Step (secs)	Segmentation Step (secs)	Tire Transform and ROI (secs)	Brand Detection (secs)	Size Location Detection (secs)	Size Character Detection (secs)	Total Inference Time (secs)
K80	0.244	0.236	0.580	2.044	2.087	0.332	5.523
P100	0.250	0.240	0.554	0.437	0.451	0.183	2.115
V100	0.237	0.243	0.564	0.215	0.213	0.162	1.634
Dual RTX 5000	0.142	0.103	0.226	0.189	0.219	0.176	1.056

7. CONCLUSION AND RECOMMENDATIONS

Within the scope of the study, a method and system has been developed that provides an end-to-end solution to the tire purchasing experience by detecting the brand and size information from the tire image that the users will take with their mobile devices in an optimum inference time with a high success rate. In developed system, segmentation and object detection models based on convolutional neural networks are used. A large data set covering many variations was collected and the final data set consisting of 300000 images was reached by using state-of-the-art data augmentation methods.

The final system uses U-NET architecture for segmentation, YOLOv5 for brand detection, and YOLOv5 for size location and character detection. The system works with an inference time of about 1 second on the system with Dual RTX 5000 GPU. This response time is more successful than the 2-second inference time reached on the GPU by similar studies in the literature.

The model showed an accuracy of 96.5% in brand detection, 98.7% in character recognition for size, and 96.5% in total size detection, outperforming the 94.7% and 95% models recommended in similar studies in the literature with an accuracy score of 96.5%. It should not be forgotten that these studies in the literature work on images taken from the straight angle of the tire with fixed camera systems. The study carried out within the scope of this thesis provides a solution to size detection by achieving better accuracy on images taken under more challenging and variable conditions compared to others. Moreover, in this study, the issue of brand detection which has not been found in previous studies has also been successfully addressed.

For the future, it is considered that the 9-character size layout, which is formed as the output of the algorithm, will be cross-checked with the size universe library in ETRTO. As a further study, it is aimed to increase the model performance by diversifying and increasing the number of images that make up the data set. For some models in the study, the entire data set could not be labeled due to time and cost. In order to use the data, set more efficiently, it is planned to complete the labeling for all models in further studies with semi-supervised methods.

REFERENCES

Asadi, A. R., & Hemadi, R. (2018). *Design and implementation of a chatbot for e-commerce*. Information Communication Technology and Doing Business, 1-10.

Badrinarayanan V., Kendall A., & Cipolla R. (2017). *SegNet: A deep convolutional encoder-decoder architecture for image segmentation*. IEEE Transactions on Pattern Analysis and Machine Intelligence, *39*(12), 2481-2495.

Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). *Yolov4: Optimal speed and accuracy of object detection*. arXiv preprint arXiv:2004.10934.

Buslaev, A., Iglovikov, V. I., Khvedchenya, E., Parinov, A., Druzhinin, M., & Kalinin, A. A. (2020). *Albumentations: Fast and Flexible Image Augmentations. Information*, *11*(2), 125. Doi:10.3390/info11020125

Chen, R. B., Zhang, J. J., Xiao, H. F., & Meng, F. H. (2013). *Methods of Dynamic Tire Pressure Judgment Based on Image Processing*. In Applied Mechanics and Materials, *239*, 980-985. Trans Tech Publications Ltd.

Cherkassky, V., & Mulier, F. M. (2007). *Learning from data: concepts, theory, and methods*. John Wiley & Sons.

Cui, L., Huang, S., Wei, F., Tan, C., Duan, C., & Zhou, M. (2017). *Superagent: A customer service chatbot for e-commerce websites*. In Proceedings of ACL 2017, system demonstrations).

Dhanachandra N., Manglem K., Chanu Y. J. (2015), *Image segmentation using K-means clustering algorithm and subtractive clustering algorithm*. Procedia Computer Science, *54*.

Donnet, J. B. (1998). *Black and white fillers and tire compound*. Rubber chemistry and technology, *71*(3), 323.

El Naqa, I., & Murphy, M. J. (2015). *What is machine learning?*. In machine learning in radiation oncology. Springer, Cham.

Everingham, M., Van Gool, L., Williams, C. K., Winn, J., & Zisserman, A. (2010). *The pascal visual object classes (voc) challenge. International journal of computer vision*, *88*(2), 303-338.

Freeman, W. T., & Roth, M. (1995, June). *Orientation histograms for hand gesture recognition*. In International workshop on automatic face and gesture recognition, *12*, 296-301.

Gao, F., Ge, Y., Lu, S., & Weng, L. (2021). V*ehicle Tire Text Reader: Text Spotting and Rectifying for Small, Curved, and Rotated Characters*. IEEE Transactions on Instrumentation and Measurement, *70*, 1-12.

Garcia-Garcia, A., Orts-Escolano, S., Oprea, S., Villena-Martinez, V., Martinez-Gonzalez, P., & Garcia-Rodriguez, J. (2018). *A survey on deep learning techniques for image and video semantic segmentation*. Applied Soft Computing, *70*, 41-65.

Gas, U. C. N. (2008). *Concerning The Adoption of Uniform Technical Prescriptions for Wheeled Vehicles, Equipment and Parts Which Can Be Fitted And/or Be Used on Wheeled Vehicles and The Conditions for Reciprocal Recognition of Approvals Granted on The Basis of These Prescriptions**.

Gent, A. N., & Walter, J. D. (2006). *Pneumatic tire*. University of Akron, Mechanical Engineering Department.

Gupta, S., Borkar, D., De Mello, C., & Patil, S. (2015). *An e-commerce website based chatbot*. International Journal of Computer Science and Information Technologies, *6*(2), 1483-1485.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). *Deep residual learning for image recognition*. In Proceedings of the IEEE conference on computer vision and pattern recognition, 770-778.

Henry, J. (2021). *Customers Want To Buy Everything Online – Tires Too, Bridgesone Says*. Forbes. Retrieved from: forbes.com/sites/jimhenry/2021/10/14/customers-want-to-buy-everything-online---tires-too-bridgestone-says/?sh=c50949d4a351. (Retrieved date: 02.03.2022)

Hough, Paul VC. (1962). *Method and means for recognizing complex patterns.* U.S. Patent No. 3,069,654.

Huang, Z., & Leng, J. (2010). A*nalysis of Hu's moment invariants on image scaling and rotation*. In 2010 2nd International Conference on Computer Engineering and Technology. *7*, V7-476. IEEE.

J. Long, E. Shelhamer, and T. Darrell (2015), *Fully convolutional networks for semantic segmentation*. IEEE Conference on Computer Vision and Pattern Recognition, 2015, 3431–3440.

Kazmi, W., Nabney, I., Vogiatzis, G., Rose, P., & Codd, A. (2019). *Vehicle tire (tyre) detection and text recognition using deep learning*. In 2019 IEEE 15th international conference on automation science and engineering (CASE), 1074-1079. IEEE.

Kazmi, W., Nabney, I., Vogiatzis, G., Rose, P., & Codd, A. (2020). *An efficient industrial system for vehicle tyre (tire) detection and text recognition using deep learning*. IEEE Transactions on Intelligent Transportation Systems, *22*(2), 1264-1275.

Khan, S., Naseer, M., Hayat, M., Zamir, S. W., Khan, F. S., & Shah, M. (2021). *Transformers in vision: A survey*. ACM Computing Surveys (CSUR).

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). *Imagenet classification with deep convolutional neural networks*. Advances in neural information processing systems, *25*.

LeCun, Y. (2015). *LeNet-5, convolutional neural networks*. *URL: http://yann. lecun. com/exdb/lenet*, *20*(5), 14.

Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., ... & Zitnick, C. L. (2014, September). *Microsoft coco: Common objects in context*. In European conference on computer vision, 740-755. Springer, Cham.

Marvin, M., & Seymour, A. P. (1969). *Perceptrons*. The MIT Press.

McPhee, L. (2016). *How SimpleTire Made a Success of E-commerce*. Retrieved from: tractionnews.com/does-ecommerce-work-in-tire-industry. (Retrieved date: 02.03.2022)

Merriman, M. (1877*). A List of Writings Relating to the Method of Least Squares: With Historical and Critical Notes, 4*. Academy.

Najman L., Schmitt M. (1994). *Watershed of a continuous function*. Signal Processing, *38*(1), 99–112.

National Museums Scothland, (2018). *Dunlop tyre*. Retrieved from: nms.ac.uk/explore-our-collections/stories/science-and-technology/made-in-scotland-changing-the-world/scottish-science-innovations/dunlop-tyre/ (Retrieved: 22.04.2022)

Neubeck, A., & Van Gool, L. (2006). *Efficient non-maximum suppression*. In *18th International Conference on Pattern Recognition (ICPR'06)*, *3*, 850-855. IEEE.

Otsu N.,(1979), *A threshold selection method from gray-level histograms*. IEEE Transactions on Systems, Man, and Cybernetics, *9*(1), 62–66, 1979.

Redmon, J., & Farhadi, A. (2017). *YOLO9000: better, faster, stronger*. In Proceedings of the IEEE conference on computer vision and pattern recognition, 7263-7271.

Redmon, J., & Farhadi, A. (2018). *Yolov3: An incremental improvement*. arXiv preprint arXiv:1804.02767.

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). *You only look once: Unified, real-time object detection*. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 779-788.

Ren, S., He, K., Girshick, R., & Sun, J. (2015). *Faster r-cnn: Towards real-time object detection with region proposal networks*. *Advances in neural information processing systems*, *28*.

Ronneberger O., Fischer P., & Brox T. (2015). *U-Net: Convolutional networks for biomedical image segmentation. International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2015, 234–241.

Rosenblatt, F. (1958). *The perceptron: A probabilistic model for information storage and organization in the brain*. Psychological Review, *65*(6), 386–408. doi:10.1037/h0042519

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). *Learning representations by back-propagating errors*. nature, *323*(6088), 533-536.

Smitters (2018). *The Future of E-Commerce for Tire Markets to 2023*.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015). *Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition*, 1-9.

Szeliski, R. (2010). *Computer vision: algorithms and applications*. Springer Science & Business Media. Chapter 5.

The European Tyre and Rim Technical Organization (2018), *ETRTO Standards Manual 2018*. Brussels.

Ultralytics. (2021). *Train Custom Data*. Retrieved from: github.com/ultralytics/yolov5/wiki/Train-Custom-Data. (Retrieved May 4, 2022),

Wang, K., Liew, J. H., Zou, Y., Zhou, D., & Feng, J. (2019). Panet: Few-shot image semantic segmentation with prototype alignment. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 9197-9206.

Weissman, S. L., Sackman, J. L., Gillen, D., & Monismith, C. (2003). *Extending the lifespan of tires. Sympletic Engineering Corporation*. Institute for Transportation Studies. University of California at Berkeley.

Widrow, B., & Hoff, M. E. (1960). *Adaptive switching circuits*. Stanford Univ Ca Stanford Electronics Labs.

Widrow, B., & Lehr, M. A. (1990). *30 years of adaptive neural networks: perceptron, madaline, and backpropagation*. Proceedings of the IEEE, *78*(9), 1415-1442.

Xia, Y., Panjie, G., & Liang, S. (2013). *Study on Tire sidewall marking recognition based on moments*. In 2013 6th International Conference on Intelligent Networks and Intelligent Systems (ICINIS), 21-24. IEEE.

Xu, R., Lin, H., Lu, K., Cao, L., & Liu, Y. (2021). *A forest fire detection system based on ensemble learning*. Forests, *12*(2), 217.

Zhang, Y., & Chu, H. (2008). *Inverse-polar ray projection for recovering projective transformations*. In 2008 IEEE Conference on Computer Vision and Pattern Recognition, 1-7. IEEE.

PUBLICATIONS AND WORKS

Özdemir V, Çalışkan A, Yiğit A. (2020). Machine learning based electric energy consumption prediction of a large-scaled production plant with small-scaled data . *Akıllı Sistemler ve Uygulamaları Dergisi (Journal of Intelligent Systems with Applications) 2020*; *3*(2): 84-89. DOI: 10.54856/jiswa.202012124

Ozdemir, V., Çalışkan, A., Özgen, M., Baytürk, E. & Özen, S. (2021). The performance comparison of optical character recognition and object detection methods on rubber products. *International Journal of Current Innovations in Interdisciplinary Scientific Studies*. *5*(1), 01-15.

Ozdemir V., Urhan O., Özgen M., Çalışkan A., Ozcan H. A., Eliöz R., Kara H. (2022). Tyre (Tire) Brand and Size Detection with Computer Vision. *30th IEEE Signal Processing and Communications Applications, SIU 2022*. (Under publishment)

Çalışkan A, **Özdemir V.** (2021). Augmented reality (AR) assisted smart glasses: case study for remote support between two distant production plants. *Akıllı Sistemler ve Uygulamaları Dergisi (Journal of Intelligent Systems with Applications) 2021*; *4*(1): 10-19. DOI: 10.54856/jiswa.202105141

BIOGRAPHY

Volkan Özdemir completed his high school education at Ankara Kırkonaklar Anatolian School. He graduated from Middle East Technical University in Aerospace Engineering in 2015. Then, he started his master's degree in 2019 at Kocaeli University in the Department of Electronics and Communication Engineering. He has studied signal processing, image processing and machine learning in his master's degree. He is currently entrepreneurship manager in BRISA BRIDGESTONE SABANCI Tire Manufacturing Inc.