MARMARA UNIVERSITY

INSTITUTE OF SCIENCE AND SCIENCE

WITH DEEP LEARNING METHODS IMAGE PROCESSING APPLICATIONS

ABDUL SA METAKTA ÿ

MASTER'S THESIS

Department of Computer Engineering

Computer Engineering Master's Program

ADVISOR

Dr. Lecturer Önder DEMÿR

CO-ADVISER

Dr. Lecturer Buket DOÿAN

ISTANBUL, 2020

MARMARA UNIVERSITY

INSTITUTE OF SCIENCE AND SCIENCE

WITH DEEP LEARNING METHODS IMAGE PROCESSING APPLICATIONS

ABDUL SA METAKTA ÿ

5 2 3 6 1 6 0 2 0

MASTER'S THESIS

Department of Computer Engineering

Computer Engineering Master's Program

ADVISOR Dr.

Lecturer Önder DEMÿR

CO-ADVISER Dr.

Lecturer Buket DOÿAN

ISTANBUL, 2020

PREFACE

I would like to thank my dear professors Dr., who provided their support, guidance, time, help and patience throughout my thesis work. Lecturer Member Buket Doÿan and Dr. Lecturer Member Önder Demir,

To my precious Mother Ayla Aktaÿ and Father ÿbrahim Aktaÿ, who have been with me at every moment of my life with their belief in me and unconditional love,

My dear brothers and sisters who make me feel like they are always with me with their support, respect and love,

To my sister and her family,

To my valuable colleagues who stood by me with their help and support,

To my dear students who work with me part-time for their help,

Marmara University Marmara Scientific Research Projects unit, which provided the Deep Learning Computer that I used during my thesis study,

I owe a debt of gratitude.

January, 2020 Abdulsamet AKTAÿ

CONTENTS

SUMMARY

IMAGE PROCESSING WITH DEEP LEARNING METHODS APPLICATIONS

Implementing image processing applications in real-time operating systems has become a trend recently. It has become a very popular topic. One of the sub-branches of the field of artificial intelligence deep learning methods used in the field of object detection from images By using image processing algorithms together, autonomous cars, autonomous unmanned vehicles aerial vehicles, assistive robot technologies, assistants for disabled and elderly individuals Applications are being developed in many areas such as technologies. This was carried out deep learning as an assistive technology for visually impaired individuals in the study It is aimed to detect tactile parquet surfaces using methods.

Unlike traditional image processing algorithms, in this study deep learning methods and image processing algorithms were used together. object detection YOLO-V3 model, one of the best models in its field, with its DenseNet model YOLOV3-Dense model was created by combining them. YOLO-V2, YOLO-V3 and YOLOV3Dense models are created by us and labeled 4580. separately on the Marmara Tactile Parquet Surface (MDPY) dataset with visual After training, their performances were compared with each other on the test data set. YOLOV3-Dense with 89% F1-score, 92% average sensitivity and 81% IoU values model is better than other models in detecting tactile parquet surfaces. has been observed. With a working speed of 60 frames per second, the YOLOV3-Dense model is real It can also be used in time-operated systems.

January, 2020 **Abdulsamet AKTAÿ**

ABSTRACT

IMAGE PROCESSING APPLICATIONS WITH DEEP LEARNING METHODS

Image processing applications in real-time systems have become a popular topic in recent times Deep learning methods, which are one of the sub-branches of artificial intelligence, and image processing algorithms are used together to develop applications in many fields such as autonomous automobiles, autonomous unmanned aerial vehicles, assistive robots technologies, assistant technologies for disabled and elderly individuals. This study aims to detect the tactile paving surfaces with deep learning methods in order to design an assistive technology system that can be used by visually impaired individuals, autonomous vehicles and robots.

Contrary to traditional image processing algorithms, deep learning methods and image processing algorithms are used together in this study. The YOLO-V3 model, which is one of the best methods of object detection, is combined with the DenseNet model to create the YOLOV3-Dense model. YOLO-V2, YOLO-V3 and YOLOV3Dense models were trained on the Marmara Tactile Paving Surface (MDPY) dataset, which was created by the researchers and included 4580 images, and their performances were compared with each other on the test dataset. It was observed that YOLOV3-Dense model is better than Other models in detecting tactile paving surface with 89% F1-score, 92% mean average Precision(mAP) and 81% IoU values. At 60 frames per second, the YOLOV3-Dense The model can also be used in real-time systems.

January, 2020 Abdulsamet AKTAÿ

SYMBOLS

X: X coordinate of the midpoint of the labeled data.

Y: Y coordinate of the midpoint of the labeled data.

W : Width of labeled data

H: Height of labeled data

dW : Image width

dH : Image height

: Neuron weight value

µ : Average

ÿ : Constant Coefficient

ÿ : Total

1 object : object finding status

1 noobj : Object not found

P : Probability

ABBREVIATIONS

CNN: Convolutional Neural Networks

Dpy: Tactile Parquet Surface

Esa: Convolutional Neural Networks

Iou : Intersection Over Union

fp : False-Positive

Fn : False-Negative

Map: Mean Average Precision

Mdpy : Marmara Tactile Parquet Surface

Ssd: Single Shot Detector

Tn: True-Negative

Tp: True-Positive

Yolo: You Only Look Once

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

Approximately 1.3 billion people worldwide have visual impairment. This Approximately 36 million of people have completely lost their vision [1]. visually impaired Individuals with disabilities use aids such as guide dogs and canes to move more easily. They use elements. In addition to these auxiliary elements, both outdoor and to the floor in order to enable them to move more easily both indoors and outdoors. There are tiled embossed surfaces. Shapes, colors and names of these surfaces It may vary depending on the countries [2].

This is commonly called tactile parquet surface (DPY) and is yellow in color. Although surfaces provide ease of movement for visually impaired individuals, there is no certain They could not find a solution. Over time, these surfaces wear out, leaving them as they should be. cannot be positioned on or around surfaces in a way that prevents movement. Problems such as obstacles may occur. These problems are visually impaired It is not always possible for people to recognize and solve the problem. Visually impaired individuals can be helped by using today's technologies to solve problems. Applications that can help are being developed. In addition to visually impaired individuals Robots and vehicles that can move autonomously also include these auxiliary applications. It needs. Image processing algorithms in developed auxiliary applications Methods such as [3-5], ultrasonic sensors [6] and light coding technology [7] have been used. By using image processing algorithms and DPY's color and shape attributes, With ultrasonic sensors and light coding technology, auxiliary elements are connected to each other. DPY determination was made by ensuring communication and information transfer.

With the influence of developments in information storage technologies in recent years The amount of unprocessed raw data has increased considerably [8]. With this increase, parallel computing methods and high processing power for areas such as machine learning and image processing has become indispensable. With all these developments, a screen with high processing power cards have become producible. Thanks to this development, parallel support is possible on graphics cards. New libraries that allow calculations have been created. With the created libraries Performing arithmetic operations on graphics cards becomes very easy and useful. has arrived. This development requires a lot of cost, time and high processing power. It is possible to perform calculations in a shorter time and in a more convenient way. made it. This created calculation and processing power is used as a deep network in artificial neural networks. It made it possible to work with multilayer networks called This deep created Thanks to networks, great advances have been made in the field of image processing. A lot Layered neural networks are concerned with the classification and detection of objects in images. Much better results than normal machine learning methods in competitions It has proven to be the best in its field [9]. 2012 by Alex et al [10] ImageNet [11] used multilayer convolutional neural networks in the classification competition. and with an error rate of 16.42%, deep learning methods have a better performance than other shallow networks. It has been proven to perform. Created using deep learning methods The success received from networks has gradually increased. The best human at classifying ImageNet data The error level is 5.1% [12]. PReLU-Net achieved a 4.94% error rate, improving human accuracy. It has become the first artificial neural network to exceed the level [13]. This value will increase in the following years. It shrank to 2.9% in 2019 with EfficientNEt [14].

1.1. Artificial intelligence

The concept of artificial intelligence was first officially introduced in 1955 in Honever, New Hampshire. By John McCarthy at a conference at Dartmouth College has been used [15]. Artificial intelligence, self-learning, reasoning and It aims to adapt human skills, such as making logical decisions, to machines. is the field of science. When we look at the historical development of artificial intelligence, we see that 1970 was a turning point. It appears to be a point. Many studies were conducted before this date and in 1969 Research stopped in 2019 due to the failure to solve the XOR problem. can be seen. After 1970, the studies of a limited number of researchers Interest in artificial neural networks has increased as a result of their maintenance and solving the XOR problem.

flared up again. These studies focus on both artificial intelligence and hardware technology. It was also supported by developments. Everyone now accepts that computers can also learn. and wants to benefit from this technology [16]. 10 years after 1970 Different new models have been developed. One of these models machine learning, simply using an algorithm to parse data, learning the parsed data and then making predictions about what the new incoming data is It is a trained machine with skills such as performing execution operations. Machine With learning, different information processing techniques have emerged. Most of these techniques One of the most important ones is artificial neural networks, which were developed based on the neural networks found in the human brain. The neural network in the human brain has a much less complex structure and it is realized in a computer environment in a shallow layered way. artificial nerve modeling networks in many more layers rather than shallow layers It is possible. Developments in computer graphics cards, the ability to program these cards and its ability to work in parallel eliminates the load on the processors, making it very It has made it possible to work with layered artificial neural networks. Multi-layered and complex Systems can be designed that can perform parallel calculations on graphics cards and New artificial intelligence models have been designed that can complete transactions in a much shorter time. This new model is called deep learning, a sub-branch of machine learning. was given. Deep learning is multi-processing to learn attributes and properties of data. It provides the opportunity to create computational models consisting of layers. This models speech recognition, visual object recognition, object detection and drug discovery, and genome They have made great contributions to studies in many other fields such as science. Deep Convolutional networks have led to major advances in image, video, speech and audio processing. Recurrent networks are a common feature in sequential data such as text and speech. It is used in this way [17]. See the historical development process of artificial intelligence in Figure 1.1 It is possible.

1.1.1. Deep Learning

Deep learning uses artificial neural networks algorithms in multi-layered architectures.

It is a sub-branch of the field of machine learning that provides the opportunity to work with multidimensional data.

Thanks to deep learning methods, natural language processing, image processing, visual object detection,

drug discovery, etc. The success rate in these areas has increased significantly [17]. deep learning, back

Explores the complex structure of multidimensional data sets using the propagation algorithm.

This process is done by obtaining the values of the parameters calculated in each layer in the previous layer.

Comparing it with the obtained values and deciding what changes should be made

does. The basis of deep learning is multilayer artificial neural networks. Simple

It is possible to see an artificial neural network model in Figure 1.3 [18]. found in Figure 1.3

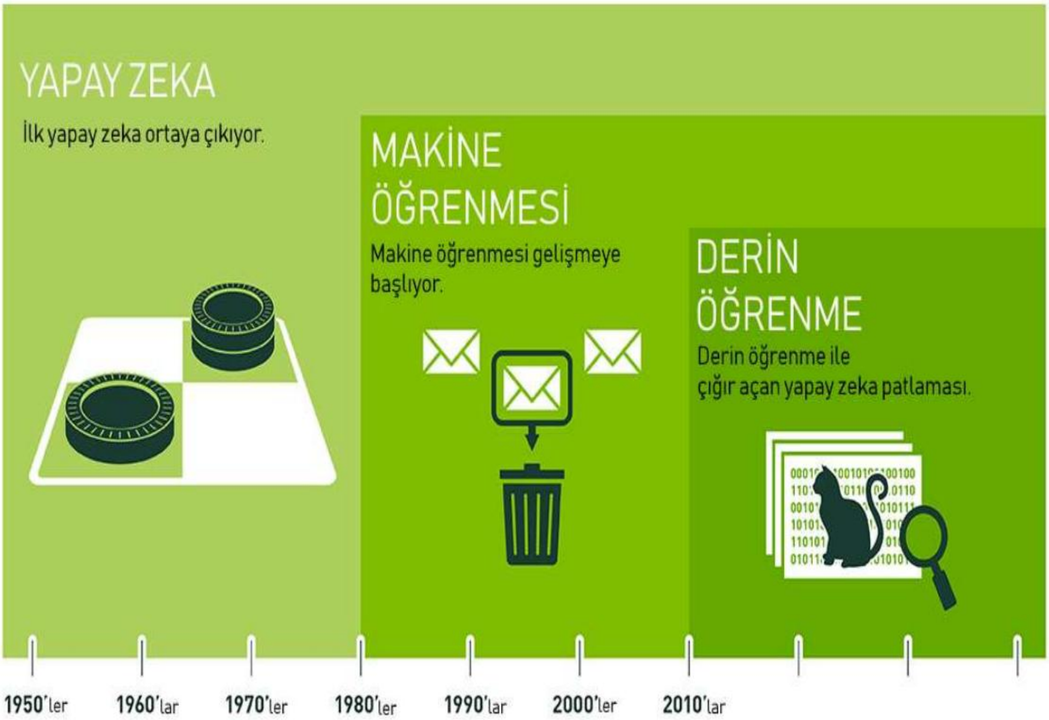


Figure 1.1 Artificial Intelligence Historical Development Process.

each connection between neurons , , is related to their weight. In Figure 1.2 It is possible to see the mathematical model of a neuron in a neural network [19].

There are two main types of deep learning methods: supervised learning and unsupervised learning. are divided into groups [20]. In unsupervised learning, there are no labels related to the data in the data set. The learning process is done with unlabeled data. Unlike unsupervised learning, in supervised learning, the data is processed appropriately according to the des

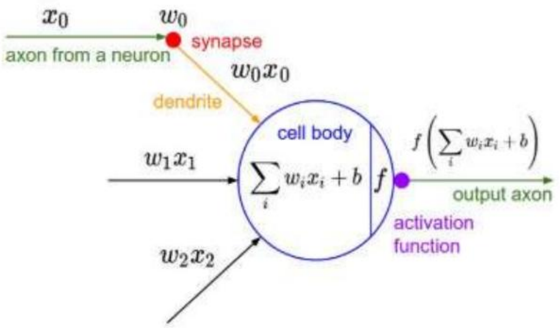


Figure 1.3. Neuron Mathematical Model.[19]

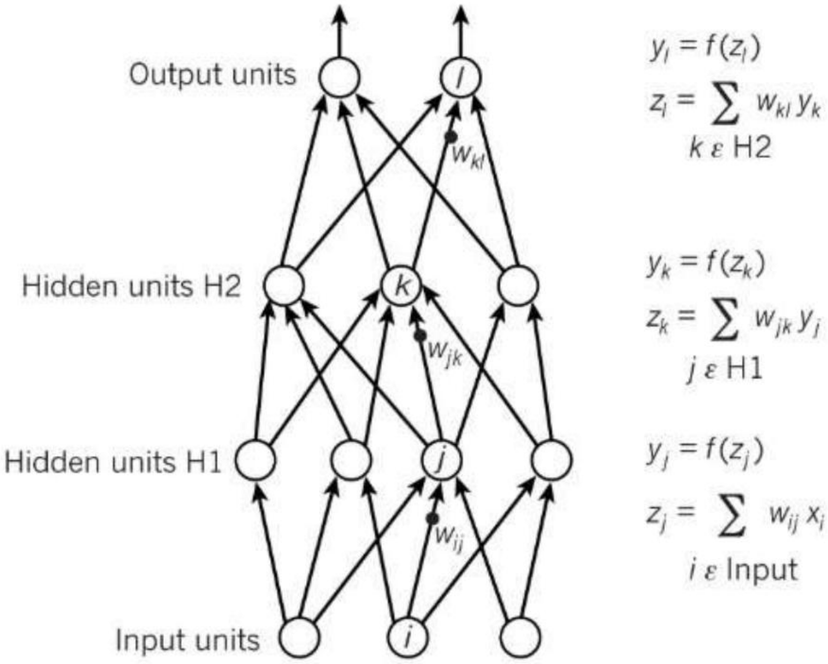


Figure 1.2. Artificial Neural Network Structure.[18]

is labeled and given to the system in a labeled manner. In the study carried out in depth Among the learning methods, the supervised learning method was used. Researchers The data collected by was labeled and training and testing procedures were carried out.

1.1.2. Convolutional Neural Networks (ESA)

Convolutional Neural Networks can learn weight values It is quite similar to ordinary artificial neural networks. Each neuron in the neural network takes data as input, processes the data and non-linearly calculates the result transfers it to the next neuron. Each neuron in the network is based on the raw image given at the beginning of the network. It calculates the score value of a class at the end of the network by processing pixels. The calculated value is last processed in the fully connected layer. A prediction value is produced that gives information about which class the image belongs to. in 1998 It was published by Yann LeCun et al. [21] and gave the first successful result. It is possible to see the layered artificial neural network model in Figure 1.4. postal numbers and It was developed for classifying and reading numbers on bank checks.

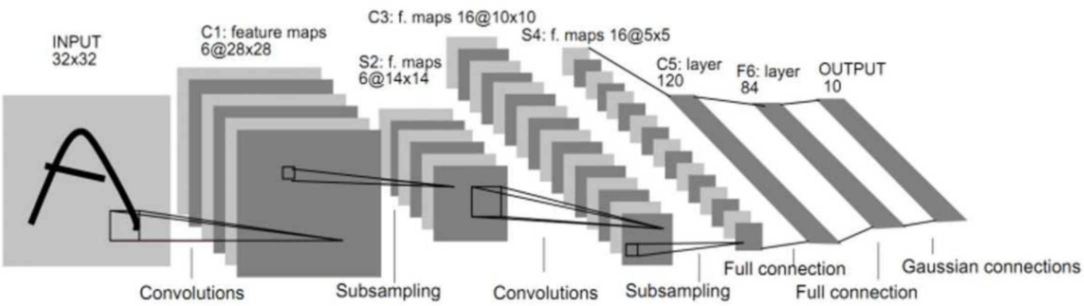


Figure 1.4. LeNET-5 Convolutional Neural Network Architecture.[21]

As seen in Figure 1.4, LeNET-5 has weight parameters, each of which can be trained It consists of a total of 7 layers. Pixel values of the given input image It is normalized and the background of the image is made white and the foreground is made black. The pixel values of the given input images are reduced to 0 and 1 levels. In this way

The learning time is increased by setting the input value to approximately 0 and the variance value to approximately 1. has been accelerated.

1.1.3. You Only Looked Once (YOLO)

The YOLO-V3 [22]network was developed using the YOLO[23] and YOLO-V2 [24]networks. As its working principle, YOLO network transforms the detection problem into the regression problem. It transforms. Unlike Fast R-CNN [25], which is one of the traditional methods, YOLO It does not require possible areas where the object can be found to detect an object. Thanks to regression, bounding box is created for each class in the image. It can create coordinates and probabilities simultaneously. For each class Instead of processing the image separately, the image is looked at only once and all Bounding box coordinates and probabilities of the classes are created. In this way, the image what objects are on it and exactly where the objects are in the image. Learning that it is done very quickly. This situation is traditional Compared to systems that detect objects, it significantly shortens the detection time. The working principle and configuration of YOLO are detailed in Chapter 2. explained.

1.2. Tactile Parquet Surfaces

Tactile parquet surfaces can be used both in interior spaces (subway, shopping mall, airport, visually impaired people both in libraries) and in outdoor areas (pedestrian sidewalks, parks and gardens). They are embossed surfaces laid on the floor so that individuals can move more easily from one place to another. First invented in Japan in 1965 Tactile parquet surfaces are widely used all over the world today. [26]. It is possible to see the tactile parquet surfaces in Figure 1.1.

In different countries, tactile parquet, cut dome, detectable warning, tactile floor
It has different names such as surface indicator and tactile guide path [2].

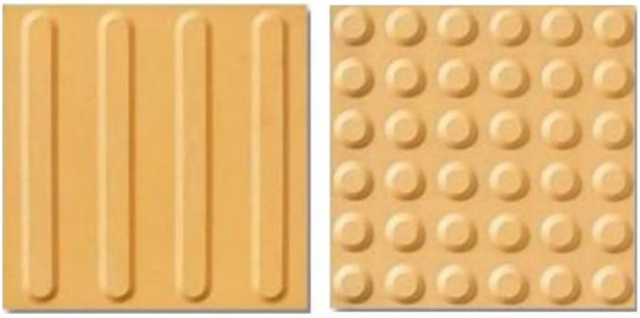


Figure 1.5. Tactile Parquet Surfaces.[26]

Figure 1.6 shows tactile parquet surfaces used in different countries [26].
Tactile parquets are suitable for people who have partially or completely lost their vision.
uses. On walking paths both indoors and outdoors
Can be installed at a low cost without the need for any extra changes
Thanks to this, tactile parquet surfaces are the most effective in guiding visually impaired individuals.
It is accepted as a system [26]. As seen in Figure 1.5: danger areas or
warning blocks indicating the location of destinations and directional blocks indicating the direction of movement
There are two types of tactile parquet surfaces. Tactile parquet surfaces, vision

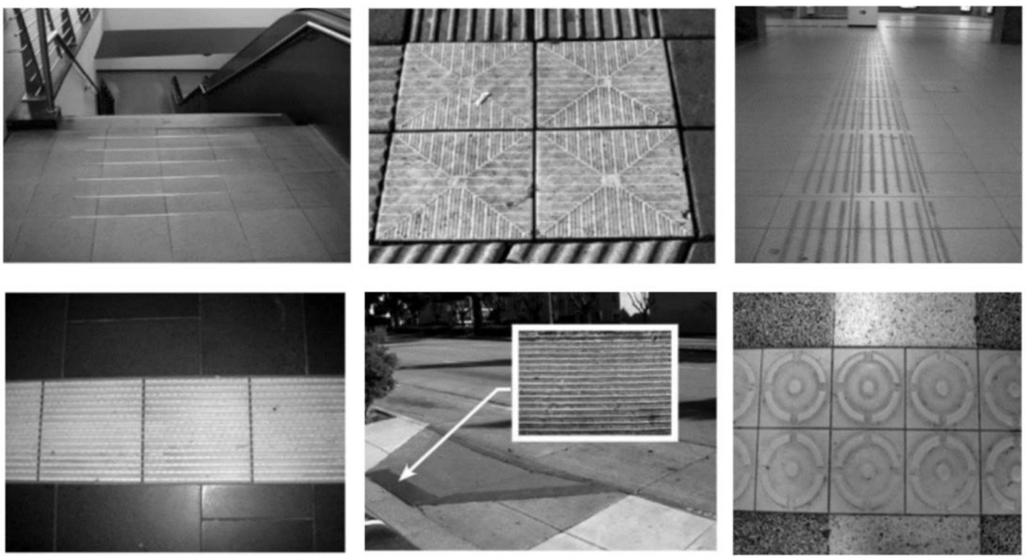


Figure 1.6. Tactile Parquet Surfaces According to Different Countries.[26]

disabled people's own position, danger position and walking direction
It is designed to help them understand. Therefore, tactile parquet surfaces
It is vital that it is installed correctly and safely. Nevertheless,
Incorrect and dangerous tactile parquet surface installations made in Turkey for the visually impaired
It poses a great obstacle and danger to individuals. Such situations also occur abroad.
It is possible to see [2]. Images of incorrect DPY installations in Figure 1.7
Due to this type of incorrect installation, visually impaired individuals
mobility is restricted and misinformation may be given. this type
detecting errors in a more controlled manner and reporting them to the relevant units,
It is very important for correction. Visually impaired individuals are aware of these types of errors.
It is very difficult for them to accurately report their location to the authorities. this type
errors are detected autonomously and reported to authorized persons according to map location information.
notification, prevention of possible accidents and correct and safe communication of visually impaired individuals.

Fig.1.7. Incorrectly Installed Tactile Parquet Surfaces.[26]

It is very important for them to travel wherever they want. Both this purpose and visually impaired individuals can travel autonomously The basics of the system required for the study planned to be carried out for this purpose are presented in this thesis. was included in his work.

1.3. Purpose and Importance

In this study, assistive technologies were used for individuals with visual impairments. It can be used in vehicles that can be used both autonomously and in vehicles that can move autonomously. tactile parquet in images, videos and real-time operating systems It is aimed to develop a model that can detect surfaces. deep learning Considering the applicability and suitability of methods in the field of image processing The study is based on deep learning methods and image processing. algorithms were used. When it comes to object detection based on deep learning There are three basic object recognition methods encountered. These; Region-Convolutional Neural Network (R-CNN), Single Shot MultiBox Detector (SSD) and You Only Look Once(YOLO) stop[23,25,27]. In the study, YOLO method and convolutional neural networks [28] was used. As a data set, it was used by researchers within the scope of this study. consisting of 4580 labeled images created and made available to other researchers. MDPY (Marmara Tactile Parquet Surface) data set was used. on the data set With 89% F1-Score, 92% average sensitivity and 81% IoU value in the experiments. Tactile parquet surfaces were detected.

Contributions Made:

- In the study carried out; which is a new and large data set in this field Marmara Tactile Parquet Surfaces (MDPY) were created and this data set depending on a particular camera angle, using the YOLO method on Tactile parquet surfaces were detected without any traces.

- A new model with high performance values has been developed.

- Deep learning methods in the field of tactile parquet surface detection
Its usability has been proven.

- High speed and speed in tactile parquet surface detection
High accuracy values are recommended for real-time systems.
provided the opportunity to use the system.

1.4. Literature research

It is difficult for visually impaired individuals to perceive and protect themselves from obstacles while traveling. With the development of technology, we can detect obstacles in advance and protect disabled individuals. guiding systems have been designed. Under this heading, deep learning methods image processing studies and obstacle detection and visually impaired important studies in the literature on assistive systems designed for individuals is mentioned.

1.4.1. Studies on Image Processing

Deep learning is a very popular field of study today. natural language processing, such as image processing, sound processing, object recognition and interpretation, pharmaceutical production, astronomy It is widely used in areas.

Jun Xiao et al [29] imaging problems in electromagnetic tomography In their study on two different deep learning based algorithms They have developed. The first of these is the algorithm created by combining sparse autoencoder (SAE) and radial basis function (RBF) networks. Secondly, optimize using batch normalization and activation function It is an algorithm based on fully connected network structure. These two algorithms developed and traditional imaging algorithms were applied on the samples and the results compared. Algorithms developed according to the results were used on GPU. Since they work much faster than traditional methods, they also They provide a better image.

Quasars are supermassive black holes at the dynamic centers of host galaxies. It is the formation of high brightnesses in a wide frequency range [30]. Pasquet-Itam, J and Pasquet, J [30] used convolutional methods to detect quasars in their study. used neural networks. at the Apache Point observatory in New Mexico Multi-filter imaging and spectroscopic red light obtained with a 2.5-meter telescope They used data obtained from redshift research. on this data Photometric analysis of quasars using convolutional neural network. Estimating, diagnosing and classifying radiations have been carried out. Obtained The results can be compared to other classification and clustering methods such as random forest. classifier(random forest classifier), k-nearest-neighbors, support vector machine (support vector machine) and gaussian process classifier (Gaussian process classifier). Convolutional neural networks compare to other classification methods. It gave much higher accuracy results than the previous ones and detected 175 new quasar candidates.

Katsumi Hagita et al.[31] developed Super Resolution with Generative Adversarial Networks that can operate based on deep learning. Using the Networks SRGAN) model, you can view the image with higher and equal resolution. They made it possible to view it. with focused ion beam to acquire data Scanning electron microscopy equipped with a focused ion beam FIB-SEM) was used. In the working logic of FIB-SEM, the sample its surface is stripped with an ion beam and examined by an SEM positioned orthogonal to the FIB. is displayed[31]. While lateral resolution is governed by SEM, depth resolution is that is, the FIB milling direction is determined by the thickness of the peeled thin layer [31]. Generally Lateral resolution is better than depth resolution. This situation is 3D When imaging is performed, an asymmetric resolution may occur [31]. To fix this problem, Katsumi Hagita et al.[31] Super Resolution with Generative Networks Using the Adversarial Networks (SRGAN) model, you can make the image appear higher and equal. provided high resolution imaging. As a result of the study, SRGAN method, bilinear and bicubic methods, which are traditional methods

compared to . In the study conducted on 8-bit images, SRGAN method is better and this method is followed by bicubic and two-line methods, respectively. has been observed. When working on 2-bit images, traditional It has been observed that the methods are better and reverse the order.

Biomedical images are one of the most valuable visual resources and references. Physicians and biomedical experts in the fields of biomedical analysis and diagnosis and treatment of diseases It plays an important role for researchers[32,33]. At the same time, biomedical Due to the wide application areas of imaging technology [34], X-ray, ultrasound, many biomedical techniques such as magnetic resonance imaging and computed tomography A wide variety of images are produced by the imaging device and displayed in different formats. is stored. To store all these data in different formats and It is a very difficult task to get it back again. For non-biomedical images traditional biomedical image acquisition methods and content-based image retrieval methods (content-based image retrieval CBIR) uses pixels and low-level features (color, shape, etc.) to define or It does not use deep features to describe images [35]. But biomedical There are much more efficient methods of storing and retrieving images this way. It is possible to develop. Shuchao Pang et al. [35] this is the traditional Unlike other approaches, biomedical images are used with deep preference learning methods. They designed an indexing and retrieval system. In this designed system, biomedical stacked autonomous encoders to display distinctive features of images denoising autoencoders SDAE) and convolutional neural networks CNN) is used. At the same time, biomedical image database using a training and teaching model to extract the similarity coefficients of images. employee preferential learning technologies by Shuchao Pang et al. [35] used. With their study, for the first time, preferential use was made for biomedical imaging. learning technology was used [35]. Created two powerful algorithms, other popular with biomedical image indexing approaches and existing image acquisition methods. in detail on various image databases that are publicly accessible.

compared. As a result of the comparison, it can be seen that Shuchao Pang and [35], state-of-the-art biomedical image It has even surpassed indexing techniques.

Classifying human behavior in videos is a very popular topic in the field of computer vision [36]. Especially in recent years, many studies have been carried out on detecting abnormal human behavior in automatic video surveillance [37,38]. Manassés Ribeiro and his colleagues [39] worked in this field and found abnormal features in videos. They aimed to detect human behavior. Recently, convolutional neural networks have become the latest technology in the field of object recognition [40,41]. Image with human help Compared to discriminators, convolutional neural networks automatically analyze both image features. It has superior discriminatory power in image display as it can learn both It is preferred because it is the latest technology and maintains its status as the latest technology [42,43]. But Convolutional neural networks are trained in a supervised manner. Only normal behaviors are They can detect abnormal behavior if they are trained with a data set that includes It's getting harder. To deal with this problem, auto-encoders come into play. Although auto-encoders initially worked on image acquisition, they have recently They also sometimes work in the field of video anomaly detection [38,43]. Auto-encoders They are concerned with problems of classifying only a single class. With just one class They can be trained. But auto-encoders use a one-dimensional vector as input data. Since they receive data, they cannot capture two-dimensional sequences in image and video sequences. To address this problem, Manassés Ribeiro et al. [39] used convolutional auto-detectors. They used encoders. UCSD pedestrian dataset [44] and Avenue dataset [45] used. The UCSD data set contains 2 subdata. The first subdata of these 5500 normal and 3400 abnormal video frames in the set There are. In the second subdataset, there are approximately 346 normal and 1652 There is an abnormal video frame. View and appearance with original video frames combinations of motion features as input data to convolutional autoencoders. has been given. Regularization of reconstruction errors reconstruction errors, RRE.), to measure the abnormality level of frames used. Classification performance, ROC curve at different threshold values It was evaluated using the area under the ROC curve, AUC. As a result of the evaluation, the highest AUC value obtained in the UCSD-2 pedestrian data set was 0.847 and the highest AUC value obtained in the Avenue pedestrian data set was 0.772. has been made. Kolmogorov [46]complexity, a measure of video spatial complexity was used as. Experiments to validate the methods, open source data mentioned Used on sets. As a result of the study, Manassés Ribeiro et al. [39] distinguishes normality/abnormality in a more general and formal way than its own algorithms. They claim that there is no method that can define it.

Jiaquan Shen and colleagues [47] in their study on deep convolutional networks detecting vehicles in aerial images using hyper feature map They intended. The basis of the algorithm is the Faster R-CNN[48] algorithm. constitutes. This algorithm basically detects objects in natural scenes. It does. Suitable for detecting small objects in aerial images It is not. Jiaquan Shen and his friends spotted the small vehicles in aerial footage. hyperactive Eltwise model and Concat model to detect accurately and effectively It offers a new approach by combining it with a feature map network. The designed algorithm detection performance on the Munich data set and their own data set. They evaluated. The system has 82.2% sensitivity (Recall) and 90.2% accuracy. It has a rate of .

1.4.2. Studies on Tactile Parquet Surface Detection

Marcelo C. Ghilardi et al [4] helping visually impaired individuals tactile detection of parquet surfaces in real time for the purpose of They have carried out. In their study, edge detection methods such as Canny, Laplace and Sobel They obtained a new approach by combining algorithms and decision trees. There are 521 images in the data set created by the researchers. The images are placed at a height of one meter and at a 45-degree angle with the ground, positioned around the person's waist. Obtained with a smartphone connected to the region. In 320 of these images

While there is a tactile parquet surface, 201 of them do not have a tactile parquet surface. In the study, the system achieved approximately 16.27 fps. works with. The system made The accuracy rate is approximately 88.48%.

In a similar study, Takuya Asami and Kengo Ohnishi [3] found visually impaired They implemented a system that helps individuals find pedestrian crossings. System a USB camera integrated into the glasses, a vibrating bracelet worn on the wrist, vision A notification warning in the visually impaired alphabet integrated into the disabled cane Being a computer used to process the data coming from the system and the USB camera It consists of four parts. In this study, images taken from a USB camera By detecting the pedestrian crossing with the normalized cross-correlation method, A guidance system has been developed. When a pedestrian crossing is detected, the user The wristband vibrates, allowing the user to stop in front of the pedestrian crossing. Developed The system detects pedestrian crossings with 86.7% accuracy.

Daniel Centeno Einloft et al. [49] In the study conducted for the detection of DPY [4] Similarly, in indoor spaces such as shopping malls, metro and bus terminals. They aimed to make a usable DPY detection system. In his studies They used the gray level co-occurrence matrix method used in image classification and support vector machines, which are one of the supervised learning methods. first results Although promising, no numerical values regarding operating accuracy are available. not given.

Xu Jie et al.[50] used image processing algorithms in their study. They aimed to detect tactile parquet surfaces using In the system made multicolor image segmentation and kirsch edge detection algorithm were used. Obtain the tactile parquet surface as a strip using Hough transformation orientation. They did. The system, which has a working speed of 2 fps and can be used portable, No numerical values regarding operating accuracy are mentioned.

AM Kassim et al. [51] visually impaired individuals' tactile parquet They designed a new cane to detect surfaces. Designed cane

Images are taken with the added web camera and sent to the computer via the Arduino Xbee transmitter. is sent. The image is processed by passing through different filters with the Matlab program in the computer environment. The “A” area value obtained from the processed images is $m = 4$ / The m metric value is calculated using the formula. According to the calculated m value It is decided whether the image taken is circular or linear. $0.15<m<0.30$ The image taken in the range is a linear image, that is, the user The image taken in the range of $0.85<m<1.0$ is circular, that is, warning parquet It is understood that there is a surface and the user must stop. Calculated m metrics The results obtained according to the value are sent back to the cane with the Arduino Xbee transmitter. According to the incoming data, the user is informed through the headset integrated into the system. Since the implemented system requires having certain hardware equipment It has no practical use.

As a result of the literature review, the use of previously made systems in daily life It is understood that its usability is quite difficult. These difficulties; to low fps systems [3,4,50] cannot be used in real time, both indoors and outdoors. In outdoor areas, tactile parquet surfaces can be detected with only one system [4,49]. inaccessibility, shopping mall, subway, etc. Separate for interior spaces such as walkways, streets Creating a separate system for outdoor spaces such as [3,51] Extra equipment such as a computer, walking stick and wristband in the backpack for use [49] studies cannot be completed due to lack of data, etc. It consists of problems. Unlike the studies found in the literature, this study The system was processed in real time at 60 frames per second and the cane, both indoors and outdoors without the need for extra equipment such as wristbands or computers. Tactile parquet surfaces were determined in outdoor areas. by researchers Thanks to the data set created and studied, the problems have been prevented.

Table 1.1 shows the main studies on tactile parquet surface detection. See the methods used, accuracy values, advantages and disadvantages It is possible.

Table 1.1. Studies Found in the Literature Related to Tactile Parquet Surface Detection

Study	Method Used	Truth	fps	Advantages	Disadvantages
[4]	Canny, Laplace, Sobel Edge Specification Algorithms and Decision trees	88.48%	16.27	Traditional image of processing algorithms examples are available.	Low fps value real time in systems Suitable It is not.
[49]	Gray level co-occurrence matrices and Support Vector Machines	-	-	Consulted learning examples of methods available.	System outputs any related to One information not given.
[50]	Multicolor Image Segmentation and Kirsch Edge Determination Algorithm.	-	2	Traditional image processing algorithms and one of the hardware parts pertaining to occasional use examples are available.	Low fps has value. In daily life use of quite costly and it's hard One is the system.

2. MATERIALS AND METHODS

The completed thesis study basically consists of 3 steps, as seen in Figure 2.1. is formed. In the first step, sub-processes related to the preparation of the data set are carried out in the second step. Creation and configuration of the convolutional neural network to be used for deep learning, In the third step, the system is trained and convolutional data is used to be used in other systems. Neural network weights were obtained.

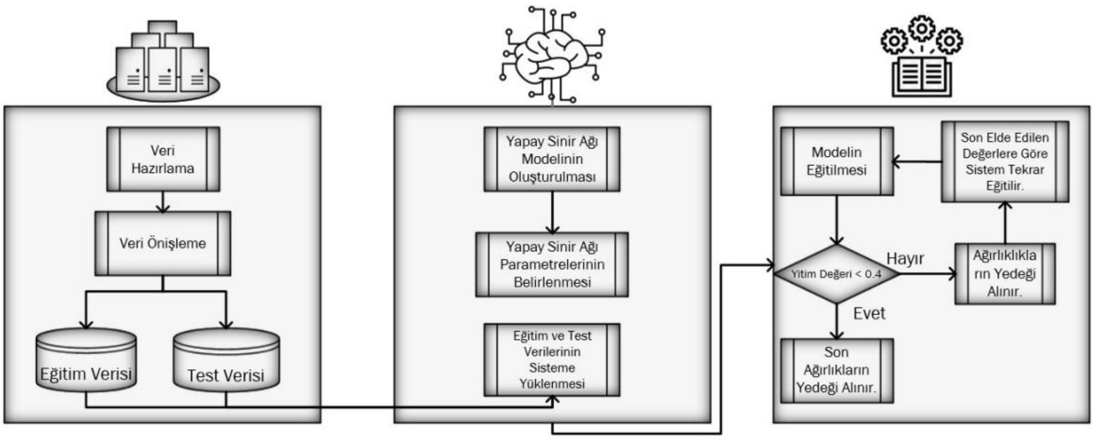


Figure 2.1 System Architecture.

2.1. Data Set Used

In the study, the data set created by the researchers was used. Data During the preparation and preprocessing phase, visual data such as images and videos taken in indoor and outdoor environments are The sources were transformed into a labeled data set by going through certain processes.

Figure 2.2. Sample Images in the Data Set of Interior Spaces.

Figure 2.3. Sample Images in the Outdoor Data Set.

Images of libraries, subways, parks, in various districts and regions of Istanbul. It is obtained from places frequently used in daily life, such as walking paths and streets. has been made. While the images were provided, GoPro Hero 7 action camera and Sony HDR-X3000R-4K action camera was used. During image acquisition, any The person's private life has not been violated. The interior images obtained are shown in Figure 1.9, It is possible to see outdoor images in Figure 1.10.

2.1.1. Data Multiplexing

In order to use the obtained images much faster and in accordance with the system It was resized to 1024x768. convolutional In supervised learning with neural networks, a large number of labeled data is needed, which is quite costly to collect and prepare [52]. To obtain the needed data Data augmentation using Python programming language and OpenCV library for process has been carried out. Data augmentation to a dataset that initially contained 1200 images By applying these operations, the number of images was increased to 4580. Data augmentation methods mirroring, whitening, rotating, etc. methods were used [53]. your images are big

Figure 2.4. Data Multiplexing Sample Images.

Having more than one DPY in some of them further increases the number of labeled data. Shape It is possible to see sample images with data duplication in 2.4.

2.1.2. Data Cleansing

Repeating itself by checking the data obtained after the data multiplexing process, Data that we cannot encounter in daily life are not included in the data set. Self Having repetitive data may affect the data set given during the training of the artificial neural network model. When any data other than the data set is given by overfitting, the detection process causes it not to be done. To prevent this, the data set is checked and Inappropriate images were not included in the data set.

(a) (b)

Figure 2.5. (a) Sample image included in the dataset. (b) Obtained after multiplexing sample image taken.

Figure 2.5 (a) shows an image included in the data set. Data multiplexing in (b) The image obtained after applying the mirroring technique is shown in the process. This There is no big difference between the two images. During training of the data set To avoid repetition, relevant images were identified and not included in the data set.

2.1.3. Data Labeling

To use the obtained data in artificial neural networks with the supervised learning method. It is necessary to label. In which parts of the images in the data set is DPY detected? A labeled data set was created by labeling the data. Do the labeling

LabelTool made using python programming language and pillow library for program was used [54].

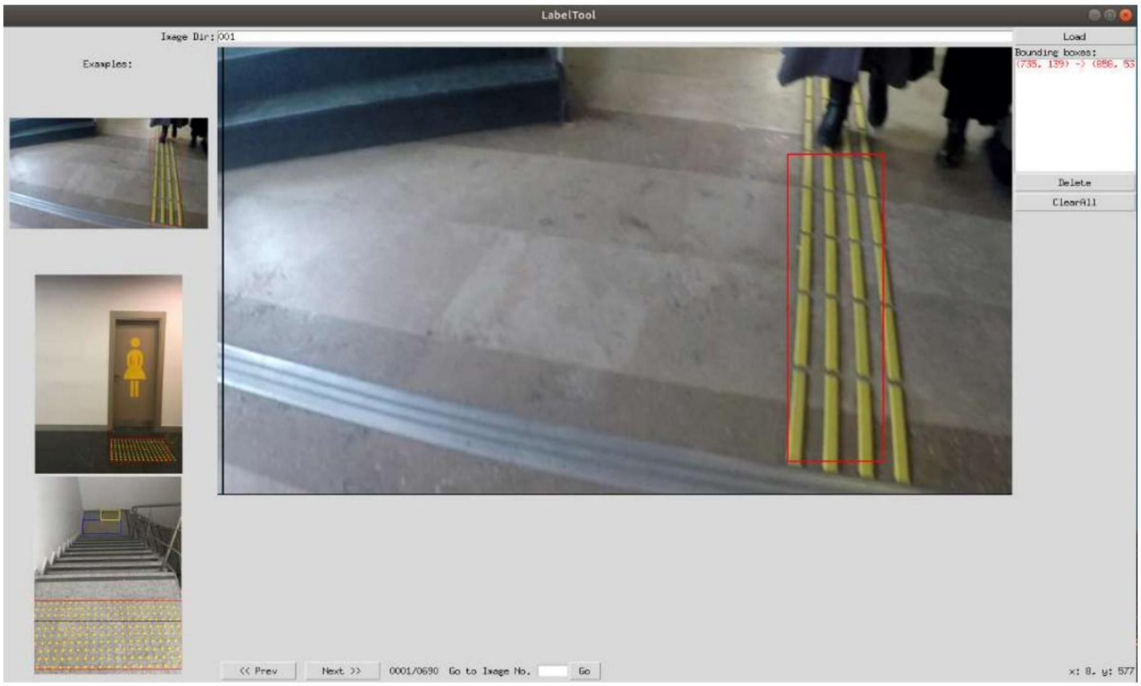


Figure 2.6. LabelTool Data Labeling Program[54]

As seen in Figure 2.6, the images are labeled with bounding boxes. by creating a text file with the same name as the image name where DPY is on it. DPY's coordinates are recorded. This process includes each of the 4580 data in the data set. Made separately for each other.

2.1.4. Normalizing the Data Set

The length(dH) and width(dW) of the labeled images are giving information about the exact location of the image (, , ,) points and the information on how many DPYs are on the image is recorded in the text file. By applying this process to all images, a labeled data set was created. Adapting the created data set to the YOLO [23] architecture used in the system Normalization is done for . With the normalization process (, , ,) coordinates are reduced to the range 0 and 1. With this process, the labeled text in the image

DPY's center point coordinates (X, Y), height (H) and width (W) information are obtained. is done. The formula of the normalization process is given in Equation 2.1.

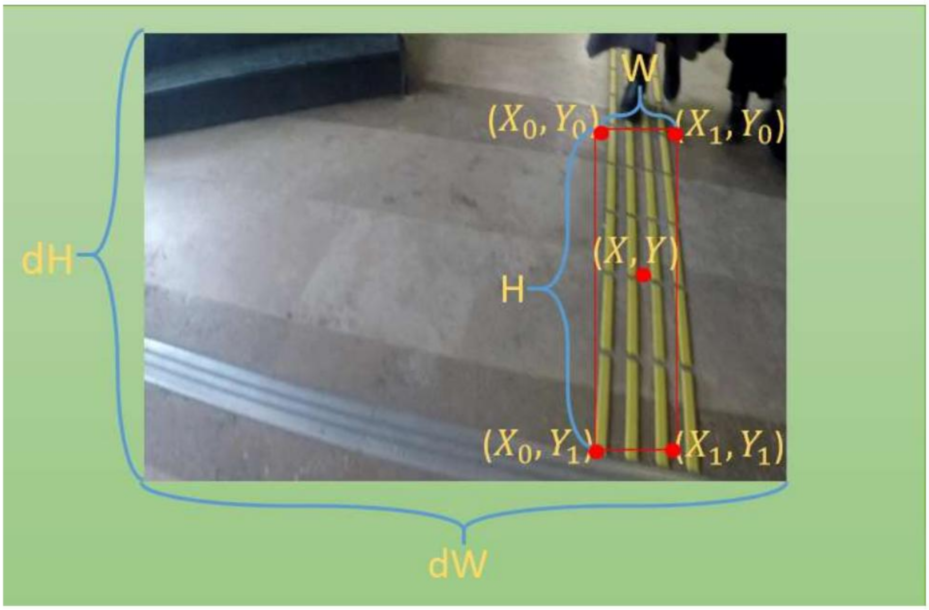


Figure 2.7. Coordinates of Tagged Data.[23]

$$= \frac{+}{2} \quad \frac{\text{one}}{}$$

$$= \frac{+}{2} \quad \frac{\text{one}}{}$$

$$= (\ddot{y}) \qquad \frac{\text{one}}{}$$

$$= (\ddot{y}) \qquad \frac{}{} \tag{2.1}$$

X: X coordinate of the midpoint of the labeled data.

Y: Y coordinate of the midpoint of the labeled data.

W : Width of labeled data

H: Height of labeled data

2.1.5. Determination of Test and Training Classes

The new coordinate calculated according to the equations in Equation 2.1, height and The width values (X,Y,W,H) are saved in a new text file with the same name as the image. After the pre-processing is completed, the data set is 65% training and 35% testing. The data was randomly divided into two.

2.2. Methods Used

The implemented system is YOLOV3-Dense, YOLO-V3 and YOLO-V2 models. It can be operated using three different models. YOLOV3-Dense model While creating the YOLO-V3 model and the DenseNet network, which is a densely connected neural network, was made by combining.

2.2.1. YOLO(You Only Look Once)

By looking at images, people can instantly understand which objects are in the images, their locations and interactions with each other. The human visual system is very fast and is working correctly. As seen in Figure 2.8, with a very small amount of attention It can even perform complex tasks such as driving. Fast and easy to detect objects High-accuracy algorithms enable computers to drive cars without special sensors. auxiliary devices to transmit real-time scene information to users.

Figure 2.8. A person who can multitask.

made possible [23]. The object detected in traditional object detection algorithms An extra classification process is carried out to classify the data. Algorithms are primarily It determines the areas in an image where objects are likely to be found. determined Convolutional neural networks designed as classifiers in areas are used separately for each region. The object is detected by running it. Although the systems produced provide good results, Since the image is processed in two separate processes, the resulting parameter numbers and needs The required processing power is increasing [55]. For this reason, the systems made work in real time. It is not possible to use it in systems. One of the newest approaches designed in this way In one of the R-CNN [48] algorithms, it first creates potential bounding boxes that can be found in an image and then develops a classifier on these suggested boxes. Region-based methods are used to operate it. After classification, improving the confidence values of the generated bounding boxes, allowing the same object to be used more than once. eliminate detection of times and reorder boxes based on other boxes in the image Operations such as sorting are performed [25]. Each of these complex operations is

Since it is trained, the system is quite slow and difficult to optimize. R-CNN in Figure 2.9 It is possible to see the architecture.

Unlike traditional object detection methods, YOLO uses bounding boxes, class calculation of probabilities and all other operations are treated as a single regression problem. and brought a new view to the field of object detection. Which image on the image with YOLO? looking at the image only once to determine where objects are is enough(YOLO). Predicting multiple bounding boxes simultaneously with a single convolutional network and class probabilities are estimated for each class. This combined model It has several benefits over traditional object detection methods.

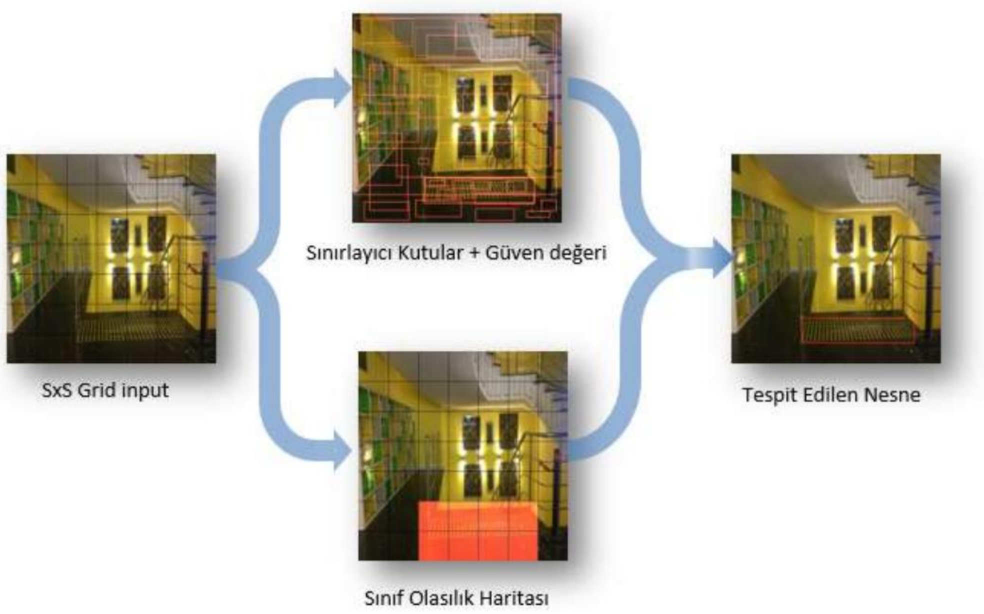


Figure 2.10. YOLO Detection Scheme.

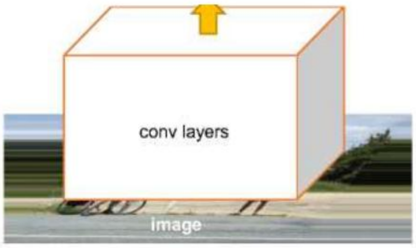


Figure 2.9. Fast R-CNN Architecture.[48]

YOLO detection model uses each image in the training set as shown in Fig.6.
SxS(S=7) divides it into square cells (grid). Any of the objects to be detected
If the center position of one of them is within any of the dividing cells, the center position
The cell in which it is located is responsible for detecting that object. Each cell B bounding
It generates bounding boxes and estimates the confidence score for these bounding boxes.
Confidence score measures how confident the model is that the bounding box it produces contains an object.
and how accurate the probability of producing the generated bounding box is.
reflects. Confidence score = Pr(Object)* , as Pr(Object)ÿ (0,1)
is calculated. If there is no object in the dividing cells, those cells
The confidence score is 0, if present it is 1. Each cell is also in equation 2.2
As can be seen, C class conditional probabilities
Pr(|), predicts. These possibilities are linked to cells containing
objects. class per cell regardless of the number of bounding boxes B generated
Only one of the possibilities is predicted. Conditional class probabilities during testing
and class-based confidence scores for each bin by multiplying the individual confidence score estimates.
It is calculated [23].

Pr(|) ÿ Pr( ) ÿ =Pr( ) ÿ (2.2)

2.2.2. YoloV3

YOLO-V2 has 19 convolutional layers and an extra 11 added for object detection.
Deep network architecture called darknet-19, which has one layer, is used. 30
YOLO-V2, which has a layered architecture, generally detects small objects.
is experiencing problems. This is because the inputs are subsampled in layers.
Attributes of small objects are lost in the detection phase by losing their attributes.
It cannot be processed. First to capture low level features in YOLO-V2
combining feature maps in layers with low-level feature maps
ID matching method was used. But there are still residual blocks in YOLO-V2.
Features such as blocks), skip connection, and upsampling
There is no. YOLO-V3 includes all this.

The YOLO-V3 network was developed using the YOLO and YOLO-V2 networks. in YOLO-V3 Darknet-53 architecture, one of the derivatives of Darknet like YOLO-V2, is used. Darknet-53 consists of 53 layers trained on ImageNet. Identification By using an extra 53 layers to perform the process, a total of 106 completely YOLO-V3 architecture was created with the convolutional layer. Therefore, it runs much slower than YOLO-V2. To perform feature extraction, sequential From the idea of 3x3 and 1x1 convolutional layers and Residual Networks [56] is used. There are 5 residual blocks in YOLO-V3. Each residual block

It consists of more than one residual unit. See the Darknet-53 architecture in Figure 2.11

It is possible.

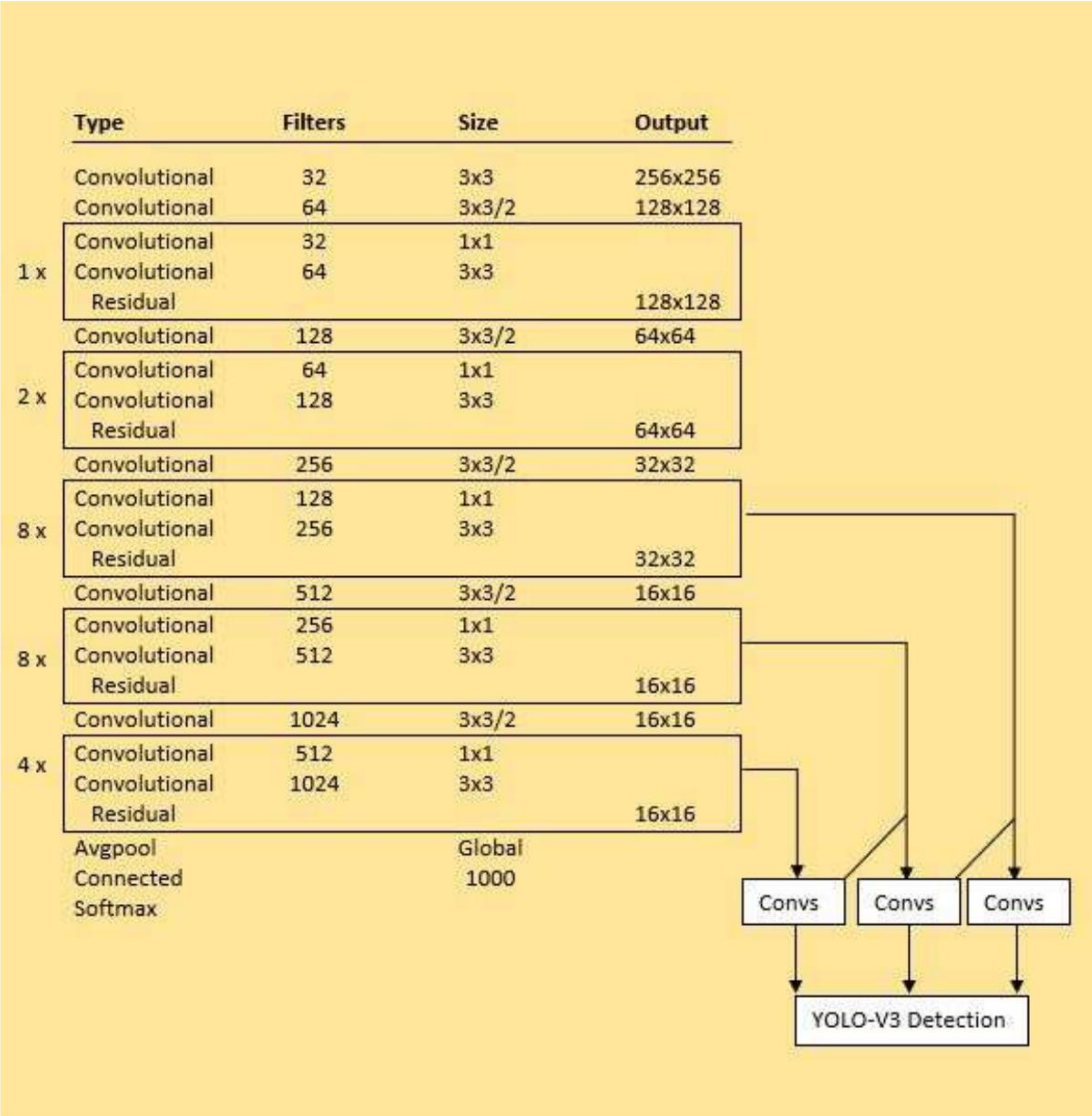


Figure 2.11. Darknet-53 architecture.[23]

The most distinctive feature of YOLO-V3 is that it detects at three different scales. YOLO totally

It is a convolutional network. The output given as a result of the detection process is placed on the feature map.

According to the result of the applied 1x1 detection kernels

is determined. As seen in Figure 2.11, in YOLO-V3 this process is done three times in three different places.

It is obtained by applying 1x1 detection kernel to the image at different scales. Perception structure of the nucleus 1 1 (5 +). by feature map within a cell is the number of predicted bounding boxes, '5' is an object confidence value and four bounding represents the box parameter and the number of classes in the system. In Figure 2.12 [57] It is possible to see the YOLO-V3 network structure.

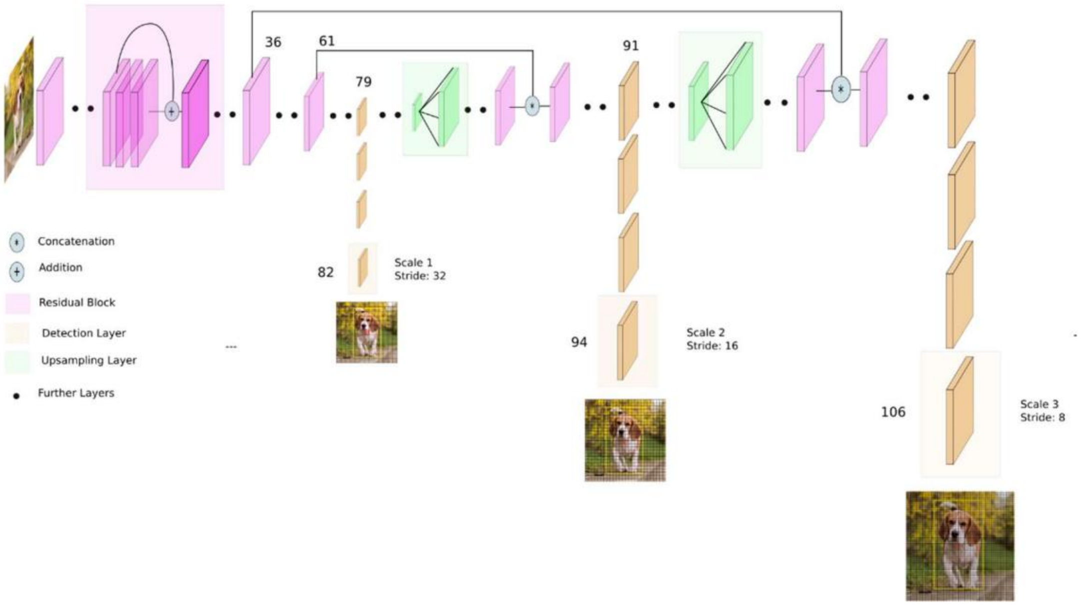


Figure 2.12. YOLO-V3 Network Structure.[57]

The given image is downsampled five times. In Figure 2.12 As can be seen, detection is performed at different scales in the last three subsampling layers. The first detection process is done at the 82nd layer. View the image in the first 81 layers Subsampling is done. 82. A 32-step stripe operation on the layer By doing this, large objects are detected. Then in the 79th layer By applying upsampling process to the feature map, it is connected with the feature map in the 61st layer. Made in the first detection process applications with different step shift sizes in layers 94 and then 106. By applying it, medium-sized and small-sized objects are detected, respectively. Deep semantic information with small size feature maps and large size feature maps attribute maps have finer grained information of the targets.

fusion is done. YOLO-V3 provides better results by upsampling. feature by resizing the feature maps of layers used in depth carries out the integration. Feature maps at different scales in YOLO-V3 After being brought to the same size, thanks to the concatenation method, deep It combines the features in the first layer with the features obtained in the first layers. This Thus, YOLO-V3 is good at detecting both large targets and small targets. It provides good performance [58]. In Figure 2.13, YOLO-V3 has other features that are good in its field. The performance of object detection algorithms on the COCO dataset It is possible to see the comparison.

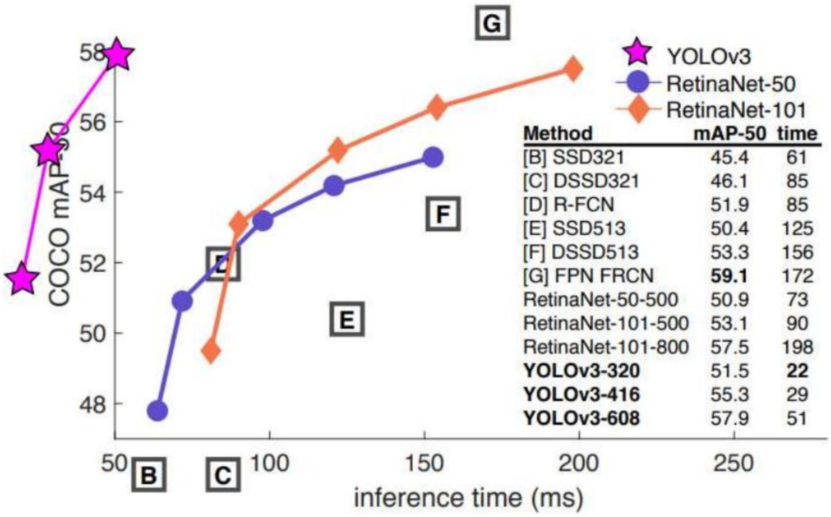


Figure 2.13. of object detection methods with an IoU value of 0.5 on the COCO data set. Comparison of mAP values and detection times.[58]

2.2.3. Densely Connected Neural Networks (DenseNet)

While training neural networks, feature changes occur due to convolution and sub-sampling processes. There is a decrease in the maps. At the same time, transitions between layers There is loss of image quality. More effective use of image feature information Huang et al. [59] developed the DenseNet system for its use. made In the system, each layer is connected to other layers in a feedforward manner. This In this way, any layer l has the feature information of all previous layers.

can access. The general network architecture of densely connected neural networks is shown in Figure 2.14 [59].

It is possible to see.

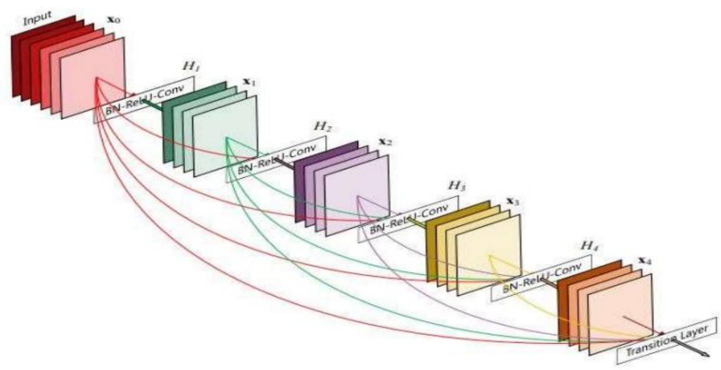


Figure 2.14. Densely Connected Neural Networks.[59]

$$X_l = H_l[X_0, X_1, X_2, \ldots, X_{l-1}] \tag{2.3}$$

$[X_1, X_2, X_3, \ldots, X_{l-1}]$ layers are combinations of feature information. H_l, feature information is the transfer function used to process. In this way, the rate of spread of features to the network increases, reuse becomes easier and the number of parameters decreases significantly.

is happening.

2.2.4. Recommended Algorithm YOLO-V3 Dense

YOLOV3-Dense network parameters used in the study are shown in Figure 8.

like this. To better handle high-resolution images, the images were first resized to 512x512. In the following layers the images are 32x32 and

Features were extracted by passing them through 16x16 subsampling filters.

The transfer function used to process attribute information in the study is H_l

It uses the BN-LReLU-Conv(1X1) function. BN (Batch Normalization) batch normalization, LReLU(Leaky Rectified Linear Units) and Conv, convolution

is a combination of features. H_l nonlinear $[X_0, X_1, X_2, \ldots, X_{l-1}]$ layers produces. Each of these layers contains 64 features with a resolution of 32x32.

It consists of layers. H_1 x_0 BN-LReLU-Conv(1X1) function on

is implementing. Then use the BN-LReLU-Conv(3X3) function on the result.

It creates the feature map of the image by applying Above H_1 by

The results are obtained by applying the operations performed by and on. in hand

made [x_0 x_1, x_2] results are given as input to the function [, x_0 x_1, x_2, x_3] results

obtained. This process produces 16x16x1024 attributes as seen in Figure 2.15.

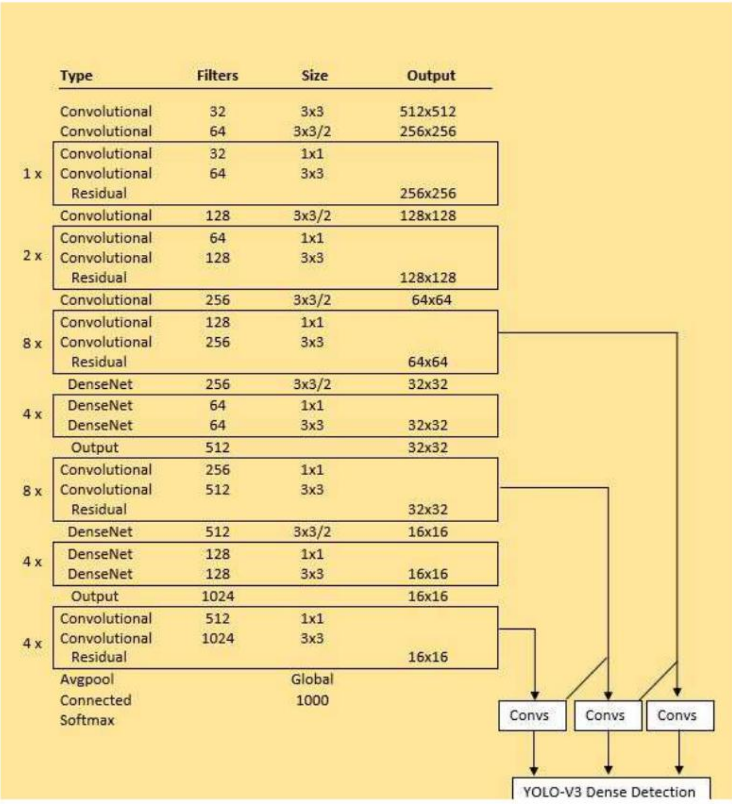
It progresses iteratively up to the next layer. In this way, image attributes are

The features taken before the image was distorted were also used in the layers

that worked on low-resolution images by transferring them from one layer to another. attribute

loss is reduced, feature usage rate is increased and a better result is obtained.

provided.



	Type	Filters	Size	Output
	Convolutional	32	3x3	512x512
	Convolutional	64	3x3/2	256x256
1 x	Convolutional	32	1x1	
	Convolutional	64	3x3	
	Residual			256x256
	Convolutional	128	3x3/2	128x128
2 x	Convolutional	64	1x1	
	Convolutional	128	3x3	
	Residual			128x128
	Convolutional	256	3x3/2	64x64
8 x	Convolutional	128	1x1	
	Convolutional	256	3x3	
	Residual			64x64
	DenseNet	256	3x3/2	32x32
4 x	DenseNet	64	1x1	
	DenseNet	64	3x3	32x32
	Output	512		32x32
8 x	Convolutional	256	1x1	
	Convolutional	512	3x3	
	Residual			32x32
	DenseNet	512	3x3/2	16x16
4 x	DenseNet	128	1x1	
	DenseNet	128	3x3	16x16
	Output	1024		16x16
4 x	Convolutional	512	1x1	
	Convolutional	1024	3x3	
	Residual			16x16
	Avgpool		Global	
	Connected		1000	
	Softmax			

Figure 2.15. YOLOV3-Dense Network parameters.

Basic network architecture of Darknet-53 architecture in YOLOV3-Dense architecture in Figure 2.16
It shows how to use it. Low resolution found in Darknet-53
instead of the original transfer layers with
DenseNet for reusing and combining feature maps
used.

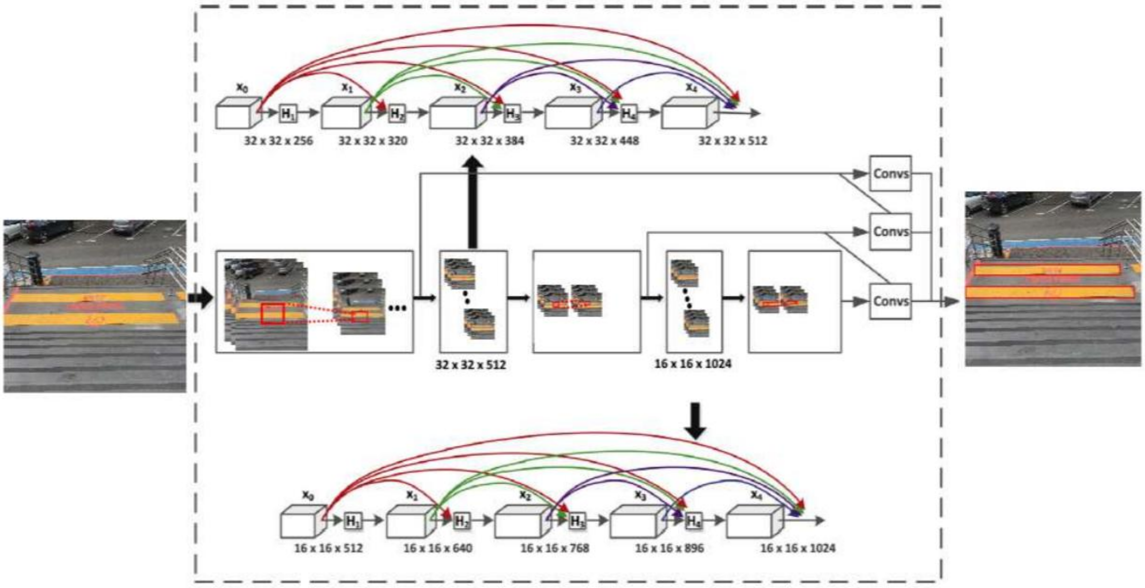


Figure 2.16. YOLOV3-Dense ANN architecture.

2.2.4.1. Used technologies

The data set used in the study consists of images obtained by us. This GoPro Hero 7 action camera and Sony HDR-X3000R-4K while capturing images action camera was used. To multiplex and tag the resulting data Python programming language was used. To run the system on the graphics card Created by Nvidia and available as open source, on-board graphics card Version 9 of the CUDA library, which offers parallel computing, was used. OpenCv, which can be used as an open source for operations performed on images library was used.

3. FINDINGS AND DISCUSSION

The training and testing processes of the artificial neural network model designed in the study were carried out using NVIDIA GeForce It was performed on the GTX1080 Ti 11 GB graphics card. the system properly momentum required for operation, initial learning rate, weight Weight decay regularization and other parameters are kept the same as in the original YOLO-V3 model. The model was trained for approximately 160 hours. first 40000 After 1 step, the learning rate decreases to 0.0001, then after 50000 steps, it decreases to 0.00001. has been dropped. The training process lasted 70000 epochs and every 10000 steps weights are backed up. These backed up weights are then applied to the test data. average precision values of the weight by running Mean Average Precision was compared. Performance of the trained model To evaluate the system's precision, sensitivity, loss, F1 score, Average sensitivity (mAP) and intersection over union (IoU) values were examined.

3.1. Metric Methods Used

3.1.1. Precision, Sensitivity (Recall), F1 score and mAP

All systems that use the probability model divide the examples of the system under four separate headings. collects [60]. These are true positive (TP), false positive (FP), true negative (TN) and are false negative (FN) values. These values are based on data tagged by the user. whether the label value and the prediction result made by the trained system match It occurs because it does not match. Assuming the labeled data is positive, the prediction If the result is positive, the TP value is formed, and if it is negative, the FN value is formed. In the same way Assuming the labeled data is negative, if the prediction result is positive, FP is negative. TN value is formed. It is possible to see the complexity matrix in table.1.

Table 3.1. Complexity Matrix.

		Tagged Status	
	Total Situation	Label Positive	Label Negative
	Guess Positive	True Positive (TP)	False Positive (FP)
	Guess Negative	False Negative (FN)	True Negative (TN)

Precision: The accuracy of true positive results in predicted samples. It is calculated as the ratio of all predicted results. Formula of precision value It can be seen in Equation 3.1.

$$= \underline{\qquad} \tag{3.1}$$

Sensitivity (Recall): Presence of true positive results in predicted samples. It is calculated by the ratio of all positive results required. Formula of sensitivity value It can be seen in Equation 3.2.

$$= \underline{\qquad} \tag{3.2.}$$

F1-score: As the harmonic mean of the values given above is calculated. The formula for the F-1 score can be seen in Equation 3.3.

$$= 2 \quad \underline{\quad} \tag{3.3}$$

Mean Average Precision (mAP): A single measurement of values such as precision, sensitivity, F1-score. It is a metric system designed to be evaluated from a single point. mAP value The formula can be seen in Equation 3.4.

$$= ÿ\ (\ ) \tag{3.4}$$

3.1.2. Loss Function

In the YOLO model, each cell (grid) produces more than one bounding box. TRUE bounding boxes generated to calculate the attenuation value in cases where it is positive The box with the highest IoU value should be selected. In this way it will come In generations, bounding boxes with low IoU values are eliminated and higher IoU values are obtained. Bounding boxes with the value are produced. Calculating YOLO loss value , the class between the bounding boxes, which is the predicted and actual reference value, is replaced by It takes the sum of the square errors of the detection and confidence values. To these values; These are called loss of classification, loss of localization and loss of confidence [23].

3.1.2.1. Loss of Classification

Conditional class for each class when object is detected at loss of classification It calculates the square error of the probabilities. Equation of loss of classification It is possible to see it in 3.5.

ÿ 1 ÿ ÿ ( ) ÿ ÿ( ) (3.5)

:u 0 .

( ): ÿü ç s g .

3.1.2.2. Loss of Locating

The localization loss is the loss related to the location and size of the estimated bounding box. calculates the value. It is calculated only for the box detecting the object. It is possible to see the localization loss equation in Equation 3.6.

ÿ ÿ 1 [(ÿ) + (ÿ)]

+ ÿ ÿ 1 [— — — — + ÿ ÿ ÿ] (3.6)

: . ÿü . 1 g 0.

: g ÿ ÿ .

Machine Translated by Google

3.1.2.3. Loss of Trust

Loss of confidence is the confidence value of the detecting box when the object is detected. calculates. It is possible to see the loss of trust equation in Equation 3.7.

ÿ ÿ 1 (3.7)

: ÿü . u g .

one : . ÿü . 1 ð 0.

Loss of trust if no object is detected in the box:

ÿ ÿ 1 (3.8)

: ÿü . u g .

one : one g u .

: g ç .

Final loss value; It is the sum of classification, localization and loss of confidence values.

ÿ ÿ 1 [(ÿ) + (ÿ)]

+ ÿ ÿ 1 [— — — — + ÿ ÿ ÿ]

+ ÿ ÿ 1

+ ÿ ÿ 1

+ ÿ 1 ÿ ÿ ( ) ÿ ÿ( ) (3.9)

3.2. Comparison of Models

It is possible to see the loss values chart of the study in Figure 3.1.

To compare the performance of the models in tactile parquet surface detection, YOLO-V2, YOLO-V3 and YOLOV3-Dense models were used on the created data set. has been operated. The loss values used during the training process are recorded and presented below. They are compared with each other in the Subduction Values Table in Figure 3.1. last loss values are approximately 1.15 for the YOLO-V2 model, 0.82 for the YOLO-V3 model and

For YOLOV3-Dense it is 0.36. Between original YOLO-V3 and YOLOV3-Dense approx.

There is a difference of 0.46.

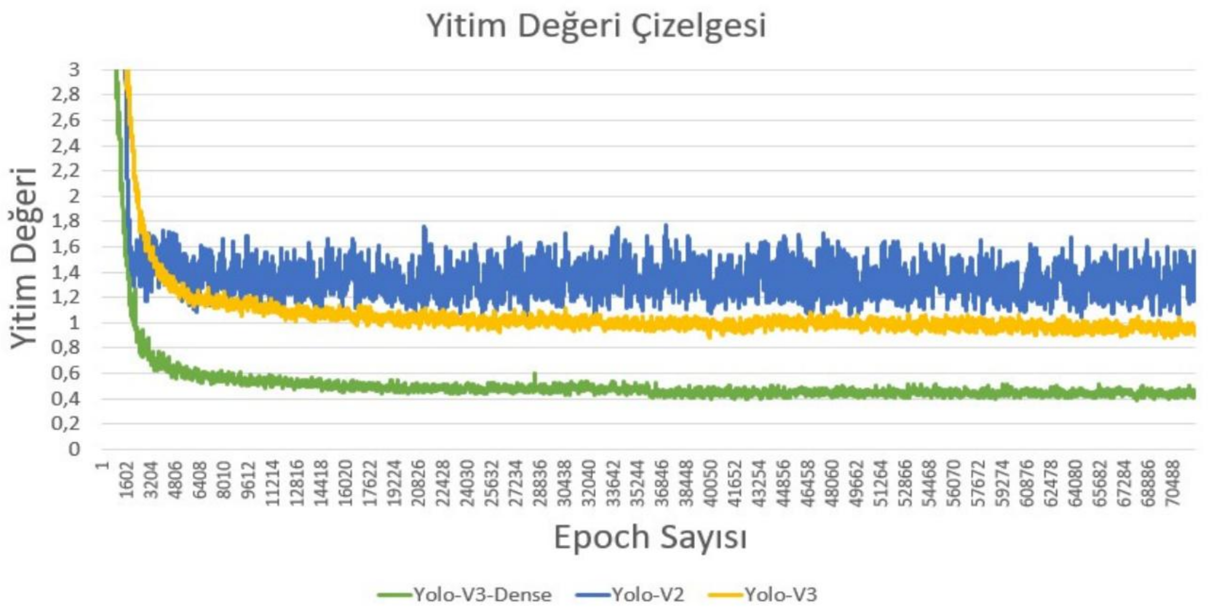


Figure 3.1. Loss values of the models.

The confidence score threshold value was taken as 0.25 and the average sensitivity values were Compared to the 81.2% value of the YOLOV3-Dense model and the original YOLO-V3 Figure 3.2 average sensitivity values, which are approximately 10% better than the model can be seen in the chart. F1- for YOLO-V2, YOLO-V3, YOLOV3-Dense models Score, IoU, mAP and Average detection time values are seen in Table 3.2. chart

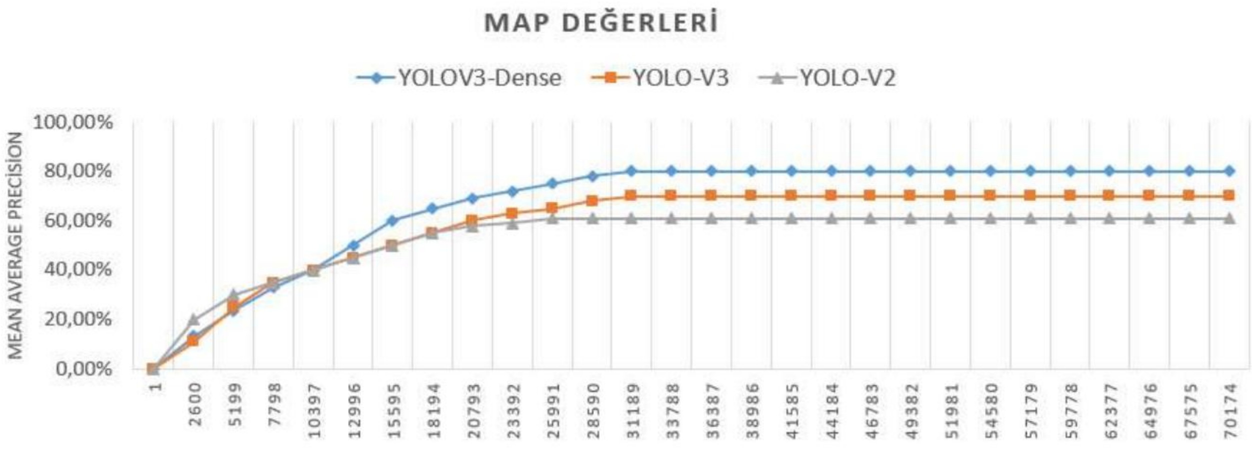


Figure 3.2. Table of Average Sensitivity Values, Confidence Score Threshold Value=0.25

and looking at the tables, in the bounding box detection process of the YOLOV3-Dense model It was found to be better than the other two models.

Table 3.2. F1-Score, mAP, IoU and detection time values of the models.

Models	F1-Score	mAP	IoU	Detection Time(s)	Confidence Score Threshold Value
YOLO-V2	0.701	0.60	0.50	0.201	0.25
YOLO-V3	0.721	0.71	0.61	0.232	0.25
YOLOV3-Dense	0.814	0.80	0.70	0.254	0.25

The confidence score threshold value was determined as 0.50, and bounding boxes below this value were These boxes are ignored and processed when calculating mAP, F1-score and IoU values. not taken. With this change in the confidence score threshold value, YOLO-V2, YOLO-V3 and YOLOV3-Dense models, mAP, F1-score and IoU values are approximately It can be seen in Table 3.3 that it performs 10% better.

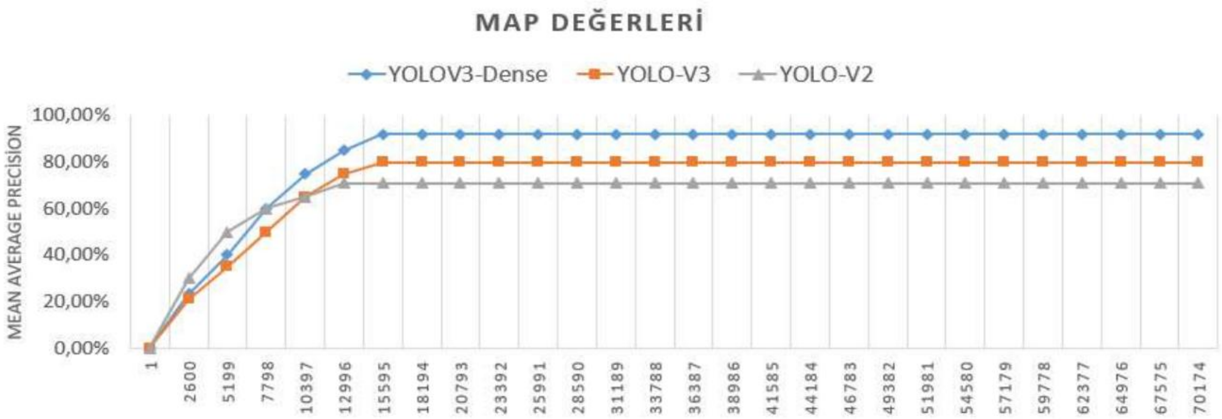


Figure 3.3. Table of Average Sensitivity Values, Confidence Score Threshold Value=0.50

Table 3.3. F1-Score, mAP, IoU and detection time values of the models.

Models	F1-Score	mAP	IoU	Detection Time(s)	Confidence Score Threshold Value
YOLO-V2	0.69	0.71	0.58	0.201	0.50
YOLO-V3	0.78	0.80	0.68	0.232	0.50
YOLOV3-Dense	0.89	0.92	0.81	0.254	0.50

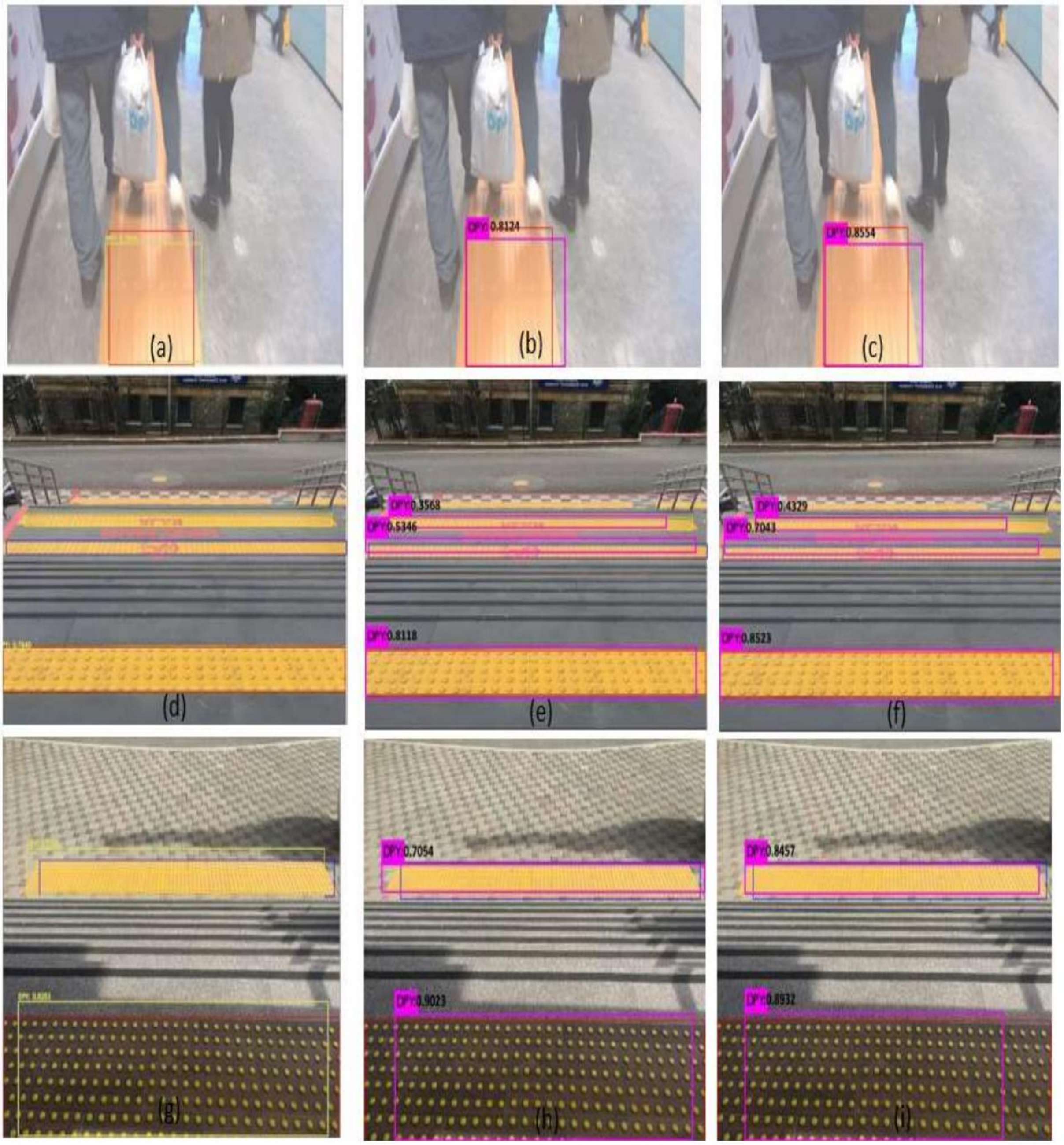


Figure 3.4. Sample Images with DPY Detected.

Three sample images with DPY detected are shown in Figure 3.4. Each image 3 The results of testing the model separately are collected in Table 3.4. a, d, g images b, e, h images with YOLO-V2 c, f, i images with YOLO-V3 images with YOLOV3-Dense has been detected.

Table 3.4. Confidence score values of the sample data set.

Image	Model Used	Trust Score
a	YOLO-V2	0.79
b	YOLO-V3	0.81
c	YOLOV3-Dense	0.85
D	YOLO-V2	0.78-0-0
to	YOLO-V3	0.85-0.53-0.35
f	YOLOV3-Dense	0.85-0.70-.0.43
g	YOLO-V2	0.82-0.80
h	YOLO-V3	0.90-0.70
I	YOLOV3-Dense	0.89-0.84

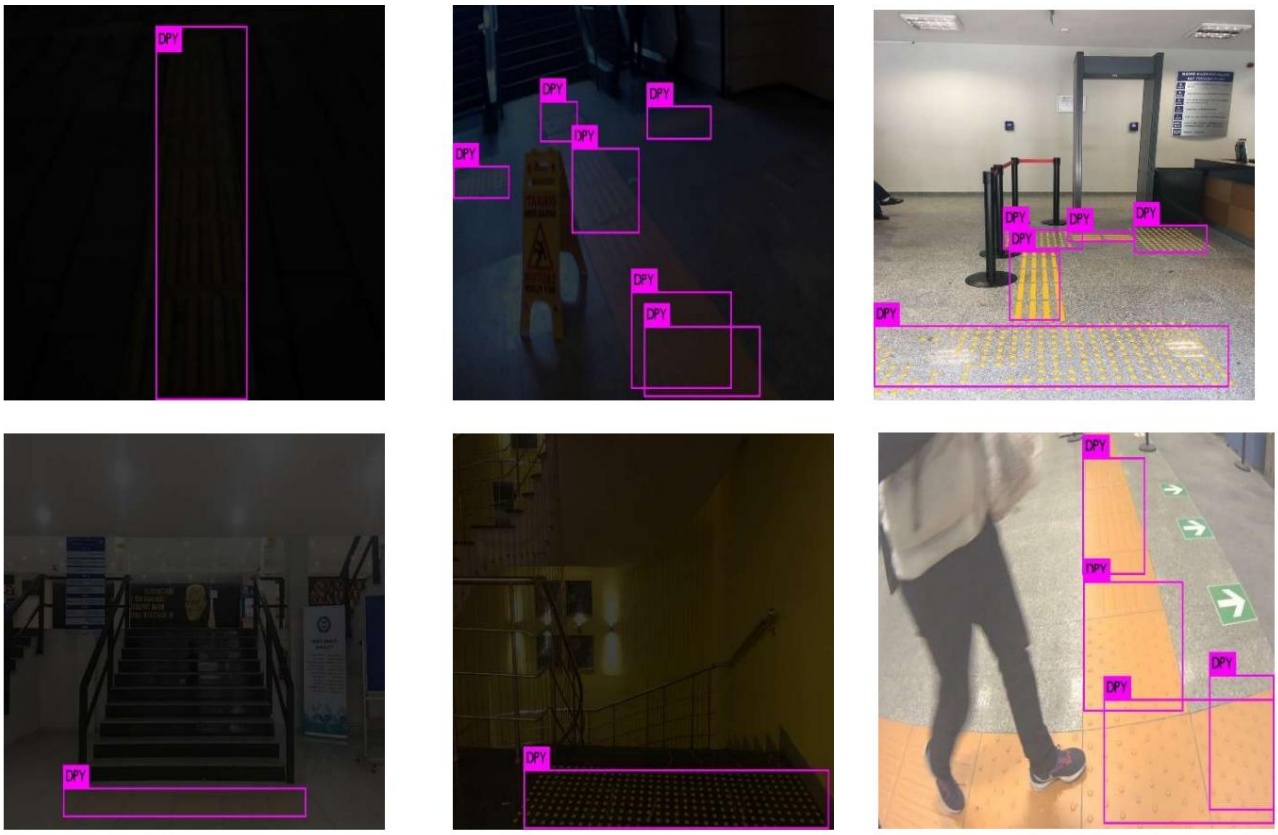


Figure 3.5. Sample images of the validation dataset.

Table 3.5. Comparison table with other methods.

Study	Method Used	Truth	fps	Advantages	Disadvantages
[4]	Canny, Laplace, Sobel Edge Specification Algorithms and Decision trees	88.48%	16.27	Traditional image processing belonging to algorithms examples are available.	Low fps value real time in systems Suitable It is not.
[49]	Gray level co-occurrence matrices and Support Vector Machines	-	-	Supervised learning examples of methods available.	System outputs any related to a piece of information not given.
[50]	Multicolor Image segmentation And Kirsch Edge Detection Algorithm.	-	2	Traditional image processing algorithms and one of the hardware parts pertaining to occasional use examples are available.	to low fps has. Daily use in life quite costly and It is a difficult system.
[24]	YOLO-V2	71%(mAP)	70	To other YOLO models a shallower network compared to fps because of its structure It has a high value.	Accuracy and mAP values are low.
[22]	YOLO-V3	80%(mAP)	60	multilayer network examples of structures available.	on the image small objects in detecting truth value is low.
	YOLOV3-Dense	92%(mAP)	60	Thanks to DenseNet all in the image attributes found on the network repeat on all layers reused and small on image detection of objects provided.	on the image all attributes using wax found parameter increased its number.

As seen in Table 3.5, the YOLOV3-Dense model uses traditional image processing It gives better results than the systems made using these methods. traditional methods It seems that the biggest disadvantage of the developed systems is the low fps value. This Due to this disadvantage, the systems cannot be used in real time.

In the YOLOV3-Dense model, not only the accuracy value is high but also the fps The value is quite higher compared to traditional methods. In this way, real-time It can be used in running systems.

4. RESULTS

In this study, YOLO-V3 model used in object detection field is DenseNet model. By combining with , tactile parquet surfaces were determined. With this model, tactile Detecting parquet surfaces and determining their position in the image can be done. Attribute found in YOLOV3-Dense model and YOLO-V3 model It is aimed to improve low resolution images in the layers. This For the purpose, the features in the image are used equally in all layers. The distribution was achieved with the DenseNet method and the network performance was increased. Experimental As a result of the studies, the YOLOV3-Dense model had an F1-score of 89% and an average sensitivity of 92%. and with 81% IoU values, it gives much better results than other YOLO models. It is understood.

Marmara Tactile Parquet Surface created by us in future studies It is planned to obtain better results by expanding the (MDPY) data set. In this study detection by optimizing the data augmentation methods and detection algorithms used. Increasing the accuracy of detection and the convolutional neural network obtained as a result of the study mobile devices that can be used in daily life by using their weights on different platforms. Applications are planned to be developed. Tactile detection detected in real time accessible to disabled individuals by vocalizing the location of parquet surfaces by the system. It can be used as an effective assistive technology by providing guidance. In future studies, many data such as people, vehicles, trees, signs will be included in the MDPY dataset. By labeling the object, visually impaired individuals are informed about the obstacles and objects around them. information can be provided. Thanks to the planned applications, the user will benefit from extra You can use the system only with your smartphone, without needing any hardware. The locations of the planned applications and the incorrectly installed DPYs are GPS determined with the help of the relevant municipalities, institutions and authorized persons. grievance can be resolved. Vehicles, people and objects that prevent the use of DPY roads It can be detected autonomously and authorized persons can be informed and Criminal proceedings may be taken against individuals.

RESOURCES

[1] World Health Organization: https://www.who.int/en/news-room/fact-sheets/detail/blindness-and-visual-impairment Access Date:

12.01.2019 [2] Lu, J., Siu, KWM, & Xu, P. (2008). A comparative study of tactile paving design standards in different countries. Paper presented at the Computer-Aided Industrial Design and Conceptual Design, 2008. CAID/CD 2008. 9th International Conference on, Kunming, China.

[3] Asami, T., & Ohnishi, K. (2014). Crosswalk location, direction and pedestrian signal state extraction system for assisting the expedition of person with impaired vision. Paper presented at the Mecatronics (MECATRONICS), 2014 10th France-Japan/8th Europe-Asia Congress on, Tokyo

[4] Ghilardi, MC, Macedo, RCO, & Manssour, IH (2016). A New Approach for Automatic Detection of Tactile Paving Surfaces in Sidewalks. Procedia Computer Science, 80, 662-672. doi: https://doi.org/10.1016/j.procs.2016.05.356

[5] Mancini, A., Frontoni, E., & Zingaretti, P. (2018). Mechatronic System to Help Visually Impaired Users During Walking and Running. IEEE Transactions on Intelligent Transportation Systems, 19(2), 649-660. doi: 10.1109/TITS.2017.2780621

[6] Shoval, S., Borenstein, J., & Koren, Y. (1998). The NavBelt-a computerized travel aid for the blind based on mobile robotics technology. IEEE Transactions on Biomedical Engineering, 45(11), 1376-1386. doi: 10.1109/10.725334

[7] Yang, K., Cheng, R., Wang, K., & Zhao, X. (2015). A ground and obstacle detection algorithm for the visually impaired. Paper presented at the 2015 IET International Conference on Biomedical Image and Signal Processing (ICBISP 2015), Beijing, China.

[8] Tsai, C.-W., Lai, C.-F., Chao, H.-C., & Vasilakos, A. (2015). Big data analytics: A survey. Journal of Big Data, 2. doi: 10.1186/s40537-015-0030-3

[9] Garcia-Garcia, A., Orts, S., Oprea, S., Villena Martinez, V., Martinez-Gonzalez, P., & Rodríguez, J. (2018). A Survey on Deep Learning Techniques for Image and Video Semantic Segmentation. Applied Soft Computing, 70, 41-65. doi: 10.1016/j.asoc.2018.05.018

[10] Krizhevsky, A., Sutskever, I., & E. Hinton, G. (2012). ImageNet Classification with Deep Convolutional Neural Networks. Advances in Neural Information Processing Systems, 25. doi: 10.1145/3065386

[11] Deng, J., Dong, W., Socher, R., Li, LJ, Li, K., Li, FF, & Ieee (2009, Jun 20-25). ImageNet: A Large-Scale Hierarchical Image Database. Paper presented at the IEEE-Computer-Society Conference on Computer Vision and Pattern Recognition Workshops, Miami Beach, FL.

[12] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., . . Li, F.F. (2014). ImageNet Large Scale Visual Recognition Challenge. International Journal of Computer Vision, 115. doi: 10.1007/s11263-015-0816-y

[13] He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. Paper presented at the Proceedings of the IEEE international conference on computer vision.

[14] Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. arXiv preprint arXiv:1905.11946.

[15] McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A proposal for the dartmouth summer research project on artificial intelligence, august 31, 1955. AI magazine, 27(4),

12. [16] Özkaynak, E. Artificial Neural Networks.

[17] LeCun, Y., Bengio, Y., & Hinton, G. (2015). deep learning Nature, 521, 436. doi: 10.1038/nature14539

[18] LeCun, Y., Bengio, Y., & Hinton, G. (2015). deep learning Nature, 521(7553), 436-444. doi: 10.1038/nature14539

[19] http://cs231n.github.io/neural-networks-1/ Access Date: 12.01.2019

[20] Özkaynak, E. (2003). Artificial neural networks. Istanbul: Papatya Publishing.

[21] Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. [Review]. Proceedings of the IEEE, 86(11), 2278-2324. doi: 10.1109/5.726791

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

[23] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.

[24] Redmon, J., & Farhadi, A. (2017, 21-26 July 2017). YOLO9000: Better, Faster, Stronger. Paper presented at the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

[25] Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.

[26] Mizuno, T., Nishidate, A., Tokuda, K., & Arai, K. (2008). INSTALLATION ERRORS AND CORRECTIONS IN TACTILE GROUND SURFACE INDICATORS IN EUROPE, AMERICA, OCEANIA AND ASIA. IATSS Research, 32(2), 68-80. doi: https://doi.org/10.1016/S0386-1112(14)60210-7

[27] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., & Berg, A.C. (2016). Ssd: Single shot multibox detector. Paper presented at the European conference on computer vision.

[28] Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R., & LeCun, Y. (2013). Overfeat: Integrated recognition, localization and detection using convolutional networks. arXiv preprint arXiv:1312.6229.

[29] Xiao, J., Liu, Z., Zhao, P. F., Li, Y., & Huo, J. W. (2018). Deep Learning Image Reconstruction Simulation for Electromagnetic Tomography. [Article]. Ieee Sensors Journal, 18(8), 3290-3298. doi: 10.1109/jsen.2018.2809485

[30] Pasquet-Itam, J., & Pasquet, J. (2018). Deep learning approach for classifying, detecting and predicting photometric redshifts of quasars in the Sloan Digital Sky Survey stripe 82. [Article]. Astronomy & Astrophysics, 611, 11. doi: 10.1051/0004-6361/201731106

[31] Hagita, K., Higuchi, T., & Jinnai, H. (2018). Super-resolution for asymmetric resolution of FIB-SEM 3D imaging using AI with deep learning. [Article]. Scientific Reports, 8, 8. doi: 10.1038/s41598-018-24330-1

[32] Fang, R.G., Pouyanfar, S., Yang, Y.M., Chen, S.C., & Iyengar, SS (2016). Computational Health Informatics in the Big Data Age: A Survey. [Article]. Acm Computing Surveys, 49(1), 36. doi: 10.1145/2932707

[33] Noor, A. M., Holmberg, L., Gillett, C., & Grigoriadis, A. (2015). Big Data: the challenge for small research groups in the era of cancer genomics. [Review]. British Journal of Cancer, 113(10), 1405-1412. doi: 10.1038/bjc.2015.341

[34] Andreu-Perez, J., Poon, CCY, Merrifield, RD, Wong, STC, & Yang, GZ (2015). Big Data for Health. [Article]. Ieee Journal of Biomedical and Health Informatics, 19(4), 1193-1208. doi: 10.1109/jbhi.2015.2450362

[35] Pang, SC, Orgun, MA, & Yu, ZZ (2018). A novel biomedical image indexing and retrieval system via deep preference learning. [Article]. Computer Methods and Programs in Biomedicine, 158, 53-69. doi: 10.1016/j.cmpb.2018.02.003

[36] Sodemann, AA, Ross, MP, & Borghetti, B.J. (2012). A Review of Anomaly Detection in Automated Surveillance. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 42(6), 1257-1272. doi: 10.1109/TSMCC.2012.2215319

[37] Bertini, M., Del Bimbo, A., & Seidenari, L. (2012). Multi-scale and real-time non-parametric approach for anomaly detection and localization. Computer Vision and Image Understanding, 116(3), 320-329. doi: https://doi.org/10.1016/j.cviu.2011.09.009

[38] Xu, D., Ricci, E., Yan, Y., Song, J., & Sebe, N. (2015). LearningDeep Representations of Appearance and Motion for Anomalous Event Detection.

[39] Ribeiro, M., Lazzaretti, A. E., & Lopes, H. S. (2018). A study of depth convolutional auto-encoders for anomaly detection in videos. [Article]. Pattern Recognition Letters, 105, 13-22. doi: 10.1016/j.patrec.2017.07.016

[40] Krizhevsky, A., Sutskever, I., & E. Hinton, G. (2012). ImageNet Classification with Deep Convolutional Neural Networks (Vol. 25).

[41] Szegedy, C., Wei, L., Yangqing, J., Sermanet, P., Reed, S., Anguelov, D., . . Rabinovich, A. (2015, 7-12 June 2015). Going deeper with convolutions. Paper presented at the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

[42] Perlin, HA, & Lopes, H.S. (2015). Extracting human attributes using a convolutional neural network approach. Pattern Recognition Letters, 68, 250-259. doi: https://doi.org/10.1016/j.patrec.2015.07.012

[43] Hasan, M., Choi, J., Neumann, J., Roy-Chowdhury, AK, & Davis, LS (2016, 27-30 June 2016). Learning Temporal Regularity in Video Sequences. Paper presented at the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

[44] Mahadevan, V., Li, W., Bhalodia, V., & Vasconcelos, N. (2010, 13-18 June 2010). Anomaly detection in crowded scenes. Paper presented at the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition.

[45] Cong, Y., Yuan, J., & Liu, J. (2011, 20-25 June 2011). Sparse reconstruction cost for abnormal event detection. Paper presented at the CVPR 2011.

[46] Yu, H., & Winkler, S. (2013, 3-5 July 2013). Image complexity and spatial information. Paper presented at the 2013 Fifth International Workshop on Quality of Multimedia Experience (QoMEX).

[47] Shen, J., Liu, N., Sun, H., Tao, X., & Li, Q. (2019). Vehicle Detection in Aerial Images Based on Hyper Feature Map in Deep Convolutional Network. KSII Transactions on Internet & Information Systems, 13(4).

[48] Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. Paper presented at the Advances in neural information processing systems.

[49] Einloft, DC, Ghilardi, MC, & Manssour, IH (2016). Automatic Detection of Tactile Paving Surfaces in Indoor Environments. Paper presented at the Workshop of Undergraduate Works (WUW) in the 29th Conference on Graphics, Patterns and Images (SIBGRAPI'16).

[50] Jie, X., Xiaochi, W., & Zhigang, F. (2010). Research and implementation of blind sidewalk detection in portable eta system. Paper presented at the Information Technology and Applications (IFITA), 2010 International Forum on.

[51] Kassim, AM, Yasuno, T., Mohd Aras, MS, hj shukor, A., Jaafar, HI, Baharom, F., & Jafar, F. (2015). Vision based of tactile paving detection method in navigation system for blind person (Vol. 77).

[52] Lv, J.-J., Shao, X.-H., Huang, J.-S., Zhou, X.-D., & Zhou, X. (2017). Data augmentation for face recognition. Neurocomputing, 230, 184-196. doi: https://doi.org/10.1016/j.neucom.2016.12.025

[53] Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on Image Data Augmentation for Deep Learning. Journal of Big Data, 6(1), 60. doi: 10.1186/s40537-019-0197-0

[54] https://github.com/puzzledqs/BBox-Label-Tool Access Date: 12.01.2019

[55] Felzenszwalb, P.F., Girshick, R.B., McAllester, D., & Ramanan, D. (2009). Object detection with discriminatively trained part-based models. IEEE transactions on pattern analysis and machine intelligence, 32(9), 1627-1645.

[56] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.

[57] https://towardsdatascience.com/yolo-v3-object-detection-53fb7d3bfe6b Access Date: 12.01.2019

[58] Ju, Yun, L., Wang, J., Hui, & Chang (2019). The Application of Improved YOLO V3 in Multi-Scale Target Detection. Applied Sciences, 9, 3775. doi: 10.3390/app9183775

[59] Huang, G., Liu, Z., Maaten, L. et al, & Weinberger, K.Q. (2017, 21-26 July 2017). Densely Connected Convolutional Networks. Paper presented at the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

[60] Goutte, C., & Gaussier, E. (2005). A Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation. Paper presented at the Proceedings of the 27th European conference on Advances in Information Retrieval Research.

RESUME

Name and surname : Abdulsamet AKTAÿ

Place of birth : Erzurum

Foreign language : English

Email : abdulsamet.aktas@marmara.edu.tr

Education status

Degree	Department/Program	University/High School	Graduation Year
High school	Science	Erzurum Anatolian High School	2009
Licence	Computer engineering	Fÿrat University	2014

Work Experience

Year	Firm/Institution	Mission
2014	ACCRIS BILISIM SOFTWARE CONSULTANCY	software developer
2017	Marmara University	Research Assistant