AUSUD
BANDIRMA ONYEDİ EYLÜL ÜNİVERSİTESİ
JITSA

Review Article

Different application areas of object detection with deep learning

Sevcan Turan[1,*], Bahar Milani[,2], Feyzullah Temurtaş[3]
[1] Çan Vocational School, Çanakkale Onsekiz Mart University, Çan, Turkey
[2] Department of Computer Engineering, Faculty of Engineering and Natural Sciences, Bandırma Onyedi Eylül University, Bandırma, Turkey
[3]Department of Electrical and Electronic Engineering, Faculty of Engineering and Natural Sciences, Bandırma Onyedi Eylül University, Bandırma, Turkey
*Correspondence: sevcanturan@comu.edu.tr
DOİ: 10.51513/jitsa.957371

Abstract: Automation is spread in all daily life and business activities to facilitate human life and working conditions. Robots, automated cars, unmanned vehicles, robot arms, automated factories etc. are getting place in our lives. For these automated actors, one important task is recognizing objects and obstacles in the target environment. Object detection, determining the objects and their location in the environment, is one of the most important solution for this task. With deep learning techniques like Convolutional Neural Network and GPU processing, object detection has become more accurate and faster, and getting attention of researchers. In recent years, many articles about object detection algorithms and usage of object detection have been published. There are surveys about the object detection algorithms, but they have introduced algorithms and focused on common application areas. With this survey, we aim to show that object detection algorithms have very large and different application area. In this study, we have given a brief introduction to deep learning. We have then focused on standard object detection algorithms based on deep learning and their applications in different research areas in recent years to give an idea for future works. Also, the datasets and evaluation metrics used in the research are listed.

Key Words: Object detection, deep learning, CNN, LSTM

Derin öğrenme tabanlı nesne algılama işlemlerinin farklı uygulama alanları

Özet: Otomasyon, insan yaşamını ve çalışma koşullarını kolaylaştırmak için günlük yaşamda ve iş hayatında yaygınlaşmaktadır. Robotlar, sürücüsüz arabalar, insansız araçlar, robot kollar, otomatik fabrikalar vs. hayatımıza hızla girmektedir. Bu otomatikleştirilmiş aktörler için önemli görevlerden biri, çalışılacak ortamdaki nesneleri ve engelleri tanımaktır. Nesne algılama -nesnelerin cinsinin ve ortamdaki konumlarının belirlenmesi- bu görev için en önemli çözümlerden biridir. Evrişimli Sinir Ağı ve GPU işleme gibi derin öğrenme teknikleri ile nesne algılama işlemleri daha doğru ve hızlı sonuç üretmeye başlamış ve araştırmacıların dikkatini çekmiştir. Son yıllarda nesne algılama algoritmaları ve nesne algılamanın kullanımı ile ilgili birçok makale yayınlanmıştır. Nesne algılama algoritmaları hakkında inceleme makaleleri de bulunmaktadır, ancak genel itibariyle algoritmaları tanıtmış ve çok yaygın olarak bilinen uygulama alanlarına odaklanmışlardır. Diğer inceleme makalelerinden farklı olarak, bu çalışmada nesne algılama algoritmalarının çok geniş ve farklı uygulama alanına sahip olduğu gösterilmek istenmektedir. Çalışmada, derin öğrenmeye kısa bir giriş yapıldıktan sonra derin öğrenmeye dayalı standart nesne algılama algoritmaları ve bunların son yıllarda farklı araştırma alanlarındaki uygulamalarına yer verilerek gelecekteki çalışmalar için rehber olmak amaçlanmaktadır. Ayrıca makalelerde kullanılan veri setleri ve değerlendirme ölçütleri de listelenmiştir.

Anahtar Kelimeler: Nesne algılama, derin öğrenme, CNN, LSTM

* Corresponding author. Tel.: 0 286 416 77 05
ORCID: 0000-0003-4278-7406, 0000-0002-5295-4215, 0000-0002-3158-4032
Received 27 June 2021; accepted 31 August 2021
Peer review under responsibility of Bandirma Onyedi Eylul University.

1. Introduction

Object detection is fundamental to defining an object's type and location in an image or video frame. Studies about object detection can be categorized into two domains, detection of known objects and detection an object that belongs to a category. For a few years, the object detection accuracy was limited due to restriction of computation cost, machine learning algorithms, and sample datasets size (Krizhevsky et al., 2017). With innovations in deep learning, GPU processing, large datasets the object detection in a category has become the most attractive topic (L. Liu et al., 2020).

In 2012, at ImageNet Large Scale Visual Recognition Challenge (ILSVRC), there has been a giant leap in object category detection with deep learning (Russakovsky et al., 2015). In this challenge, Krizhevsky et al. (2017) presented AlexNet, a Deep Convolutional Neural Network (DCNN). They have changed the direction of studies, so deep learning has become an essential tool for object detection. The innovation in object detection is used in health, robotics, transportation, security, manufacturing, scientific research about sky and earth, etc. (L. Liu et al., 2020).

In some of the surveys structure, metrics, performance etc. of object detection algorithms have been examined and compared, and then some common application areas have been given as examples (Jiao et al., 2019; L. Liu et al., 2020; Padilla et al., 2020; Zou et al., 2019). Also, some of the surveys have been focused on specific application area, for example salient objects (Borji et al., 2019; J. Han et al., 2018; W. Wang et al., 2021), traffic objects (Arnold et al., 2019; Sanyal et al., 2020), remote sensing objects (Cheng & Han, 2016; K. Li et al., 2020), objects that recognized by unmanned aerial vehicle (Cazzato et al., 2020; Mittal et al., 2020), etc. There is not a survey focused on showing how wide the usage area of the object detection. This study's contribution is presenting different application area of object detection based on deep learning with dataset and evaluation metrics they have used. The studies are selected from the separate topic for projecting the usage area of object detection to motivate the next studies.

In this study, in the second part most used deep learning architecture is introduced. Then, in the third part, common object detection algorithms and object detection research are examined. In the fourth part the discussions and in the last part the conclusions are presented.

2. Deep Learning Architecture

Deep learning, a branch of machine learning, is a technique for predicting the system's expected results from the various number of sample data without human interaction (LeCun et al., 2015). Its structure is based on an artificial neural network with a multi-hidden layer. As seen in Fig. 1, there are three main layers: input, hidden, and output. The input layer is used for collecting the input data, and the hidden layer is used for extracting features from the data and, the output layer is used to predict the results. Each layer consists of neurons, and in neurons, the inputs x_i are multiplied with the weights w_i, and the results are summed. The produced results are passed as input to the next layer (Öztemel, 2012).

Deep learning consists of two primary processes, as shown in Fig.1: a) Forward computing for predicting results from the input and b) Backpropagation for updating the weights to reduce the prediction error.

There are three main types of learning strategy for training deep neural network (*Ufldl Tutorial*, n.d.):

- Supervised Learning: Data and its labels determined by an expert of the domain are used to train the network. The predicted result is compared to the tags. And due to the difference between them, the weight of the neurons is updated.
- Semi-Supervised Learning: Data with and without labels used for training. There are more unlabelled data and a few labelled data available.
- Unsupervised Learning: Data without labels are used for training, and the system clusters the input according to similarities to each other.

2.1. Convolutional neural network (CNN)

CNN is the most used architecture for the object detection process introduced by LeCun et al. (1989). CNN is used for image classification (Cevikalp et al., 2020), image segmentation (Turan & Bilgin, 2019), speech recognition (Pavan et al., 2020), object detection (R

Girshick et al., 2016; Ross Girshick, 2015; Krizhevsky et al., 2017; Law & Deng, 2018; Redmon et al., 2016; Ren et al., 2017), object tracking (Yuan et al., 2020), network management (Vinayakumar et al., 2017), etc. In Fig. 2, the architecture of CNN is shown.

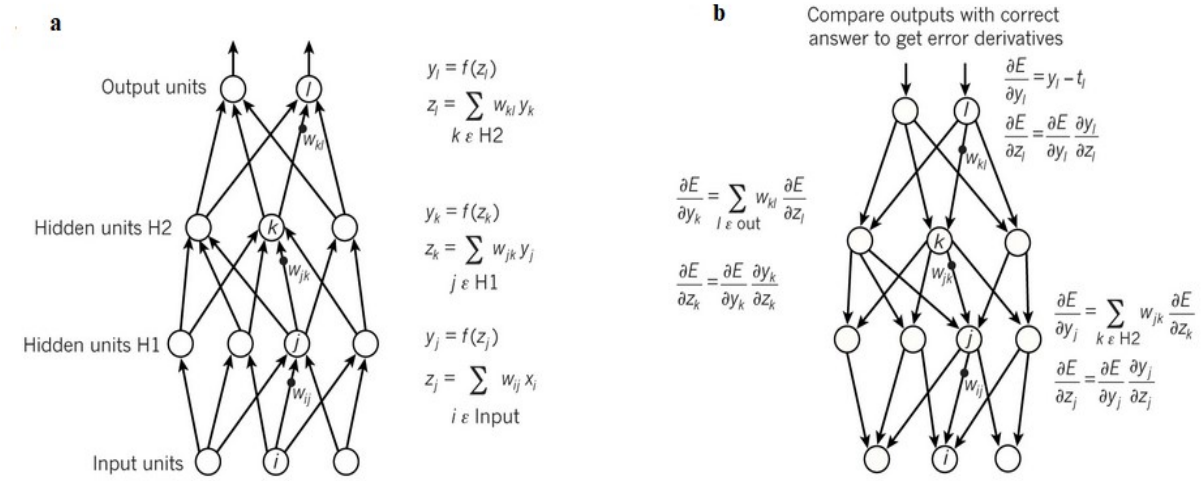

Figure 1. *Deep learning architecture. a) Computation the forward pass. b) Backpropagation for learning (LeCun et al., 2015)*

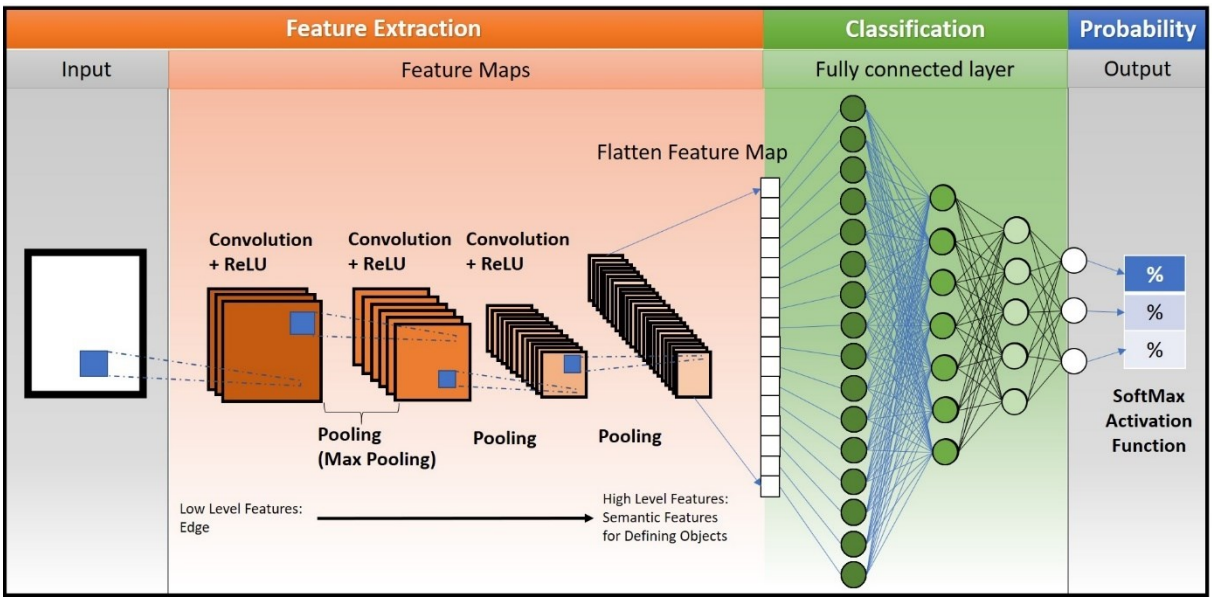

Figure 2. *Architecture of CNN*

The main part of the architecture is as follows (Indolia et al., 2018):

a. Convolution Layer
In the convolution layer, the input is convolved with filters, and this is done several times. In the early convolution steps, the basic features are obtained as edge, and in the last stages, the features represent the object is obtained.

b. Pooling
In pooling, the feature map size is reduced to get the most relevant feature representing the target objects. For this process, a 2×2 grid is taken from feature map, and in general with maximum pooling (MaxPooling) algorithm the new feature value for the next step is calculated. In the MaxPooling algorithm, the maximum value in grid is taken.

c. Activation Function
For providing nonlinearity, an activation function is used. In a neuron, the input goes to the activation function. Then the result is sent to the

next layer of the network. The most used function is Rectified Linear Unit function (ReLu). In ReLu, negative values are converted to 0, and the positive values stay the same.

d. Fully Connected Layer and Classifier
The fully connected layer is the classic neural network layer whose nodes are fully connected to the previous layer used to calculate the object class score. In this layer, first, the feature map is flattened, and a feature vector is obtained for using as the input of the classic neural network system. The result of the network is classified with function like SoftMax.

2.2. Recurrent neural network (RNN)

RNN is the network that can be used for learning from sequence data introduced by Elman (1990). As shown in Fig. 3 at time t, x_t and h_{t-1} which is previous hidden state value are the inputs, and o_t is the output, and h_t is the new hidden state of the network.

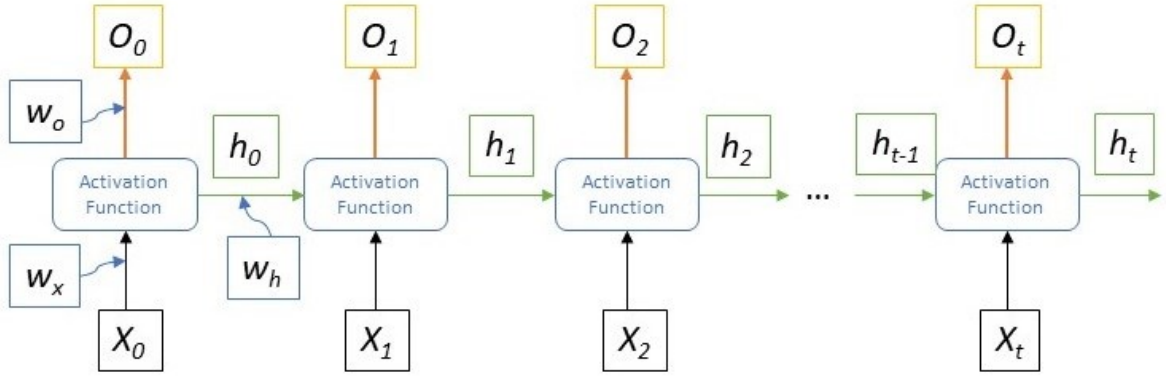

Figure 3. *Architecture of RNN (Many inputs, many outputs)*

For each time step, same w_x, w_y and w_h are used respectively as input, output, and hidden state weights. The hidden state of the network h_{i-1} where i=1,2,3…, t, is used as input for each next time step. In this way, RNN can use the historical data for prediction. But, according to vanishing gradient problem, RNN cannot be used for recalling long time history.

RNN is used for natural language process (Lawrence et al., 2000), speech recognition (Bahdanau et al., 2016), video processing (W. Zhang et al., 2016), etc.

2.3. Long short-term memory (LSTM)

LSTM, a type of RNN, is used for keeping information for a long duration which was developed by Hochreiter and Schmidhuber (1997). In LSTM, memory cells contain input gate, forget gate, cell state, output gate, input activation function, and output activation function (Gers et al., 1999), as shown in Fig. 4.

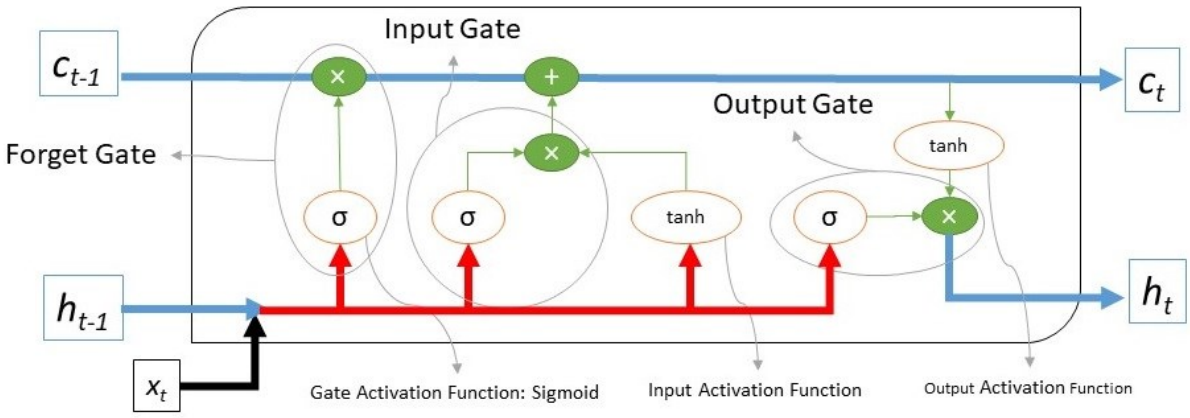

Figure 4. *Architecture of LSTM memory cell*

Cell state is used for transferring long-term historical information in the network. Forget gate is used to decide which values from the input and previous output of the network will affect the cell state. Input gate is used for deciding which input values will affect the cell state. Output gate is used for generating network result with cell state. At time t, input x_t, the previous output of network h_{t-1} and cell state c_{t-1} are given as input parameters to network, and output h_t and new cell state c_t are produced. At forget gate, h_{t-1} and x_t are feed to the activation function, and the result of the gate is bitwise multiplied with the c_{t-1} for deciding which values will be forgotten. At the input gate, h_{t-1} and x_t is fed to the activation function to determine which input values will affect the cell state, and the result of the gate is multiplied with the result of input activation function. And the resultant value is summed with cell state c_{t-1} and produced the new cell state c_t. At the output gate, h_{t-1} and x_t are feed to the gate activation function, and the c_t is fed to the output activation function, and the results are multiplied, and h_t is produced. h_t and c_t are passed to $t+1$ process.

LSTM is used for object detection in video (Sorano et al., 2020), object tracking (Tsai et al., 2020), speech recognition (J. Li et al., 2020), video processing (Xiao et al., 2020), natural language process (Jelodar et al., 2020), etc.

3. Works on object detection

In this section, the most common object detection algorithms and their applications in other research area are examined. The studies which use object detection have been selected from the recent years for limiting the paper content.

As seen from the studies in Table 1, most object detection algorithms based on CNN can be used to get the features of objects. CNN-based algorithms handle the inputs as independent from each other. It takes an image, detects the objects, passes to the new one, and the relationship between the scenes is not considered. RNN/LSTM is used if the input is handled as a sequence. Some researchers divide input into regions and consider regions as time series (W. Han et al., 2020; Kechaou et al., 2020).

AlexNet is the turning point of the studies. They used 15 million labelled RGB images with 256×256 resolution from the ImageNet dataset, five convolution layers, and three fully connected layers for training a CNN network for image classification. They used the ReLu activation function, 2×2 pooling grid size, and dropout layer. They used GPU for computation convolutions. As a result, they achieved a 15.3% error rate and gained significant success. This classification technique has been used as a backbone network for object detection, and it has been a guiding idea for researchers.

In object detection, there are two leading groups of CNN-based algorithms: One-stage and two-stage algorithms. One-stage algorithms are fast, but two-stage algorithms are more accurate. So, the researcher should decide whether the speed or accuracy is essential for this issue. Some of the common CNN based object detection algorithms are as follows:

- R-CNN, Fast R-CNN, Faster R-CNN: Region-Based Convolutional Neural Network (R-CNN) is a two-stage algorithm. At the first stage, 2000 proposal regions are found by a Selective Search Algorithm (SSA) from the image that are candidates for containing objects. Then these proposals are sent to the neural network, and a feature map is produced. Then object class and bounding box is getting from feature map. The objects are detected with 53.3% mean average precision (mAP) on Pascal VOC 2012 (*Pascalvoc*, n.d.) dataset (Girshick et al., 2014). In 2016 this team introduced a new version of R-CNN with 62.4% mAP (R Girshick et al., 2016). The R-CNN algorithm is so slow because of SSA and running network for 2000 times, therefore Fast R-CNN has been introduced to develop accuracy and computation performance (Ross Girshick, 2015). In the Fast R-CNN training process, images and object proposals are used as input and a feature map is produced. The features are selected for object detection in the Region of Interest Pooling algorithm by using object proposals and feature maps. After a fully connected layer, with Softmax classifier object categories, and with Bounding Box Regression bounding boxes are found. Objects are detected with 68.4% mAP

and 9.5-hour training time, 0.32 second testing time on Pascal VOC 2012+2007 dataset. R-CNN and Fast R-CNN are used SSA to find regions and are slow for real-time object detection. New algorithm Faster R-CNN has replaced this selection algorithm with Region Proposal Network (Ren et al., 2017). In Faster R-CNN, the image is feed to convolution layers, and a feature map is produced. Using feature maps, region proposals are learned with Region Proposal Network. Objects are detected, using Softmax classifier and bounding box regression, with 75.9% mAP on Pascal VOC 2007+2012 and MS COCO (Mscoco, n.d.) dataset combination.

- YOLO (Redmon et al., 2016): YOLO is a one-stage algorithm. In this study, the image is divided into grids with dimensions 7×7. And each grid is responsible for detecting even if there is an object in it. If an object's centre is found in the grid, the grid is assigned to find the class category and two bounding boxes with the object's confidence score. Then the highest confidence score is selected for the category and bounding box. They have detected objects with 63.4% mAP on PASCAL VOC 2007+2012 dataset at real-time.
- SSD (W. Liu et al., 2016): In this study, VGG16 (Simonyan & Zisserman, 2015) has been used with some modification as the backbone to get a feature map. After VGG16, they have added six convolution layers, and classification and bounding box regression are done at different feature map scale, and for each class, 8732 predictions are found. Then the highest prediction rate is selected for determining objects. They have used six default boxes with different scales and orientations at different feature map sizes. If the boxes contain information about a class, then comparing ground truth, the offsets are determined, otherwise they are considered as background. Using different feature map sizes for prediction, they can detect the objects with varying sizes on the image. They have detected objects on 300×300 images with 74.3% mAP and 59 frames per second (fps), and 512×512 images with 76.9% mAP on the PASCAL VOC2007 dataset.
- RetinaNet (T. Lin et al., 2017): It is a one-stage detector that has achieved almost two-stage detectors. In this study, they have presented a new loss function named focal loss which eliminates the background values. They have used hard positive examples in training. In the algorithm of RetinaNet, Resnet (He et al., 2016) and Feature Pyramid Network (FPN) (T. Y. Lin et al., 2017) has been used as the backbone to get multi-scale feature maps, and an anchor box has been used for detecting object positions. For each FPN level, class and bounding box prediction networks are run separately. In this algorithm objects are detected with 39.1% average precision (AP) and 122 milliseconds on 600×600-pixel images from the MS COCO dataset, more accurately than the two-stage algorithms.
- CornerNet (Law & Deng, 2018): It is a one stage detector which eliminates the anchor boxes, and object's coordinate is represented with two corner point, left-top and right-bottom, and a bounding box is drawn around the object with these points. For getting feature maps from the image, the Hourglass network is used as a backbone network. From the feature map, with a new corner pooling algorithm, they have found candidates of the left-top and right-bottom corner points as heatmaps, embeddings, and offsets. The distance between the embeddings is used to find the left-top and right-bottom points that belong to the same object. Then, they draw the bounding boxes. With 42.1% AP, they have detected the objects on the MS CO-CO dataset.
- CenterNet (Duan et al., 2019): It is a one-stage detector developed based on CornerNet. They have cascaded the corner pooling algorithm and added the centre point pooling mechanism. If the centre point of the bounding box has the feature representation about a class, this bounding box is chosen. In this way,

they have reduced the incorrect bounding box detection. They have achieved the highest accuracy than the other one-stage detector with 47.0% AP on the MS COCO dataset.

According to research area and dataset properties, researchers can use ready-made algorithms like above as well as develop new algorithms themselves. Some examples of the studies in the recent years at different research areas that use object detection are shown in Table 1.

Table 1. *Samples of research used object detection in recent years.*

Ref. No	**Dataset**	**Evaluation Metric**	**Research Detail With Contributions and Limitations**
(Sardoğan et al., 2020)	Own dataset	*Diseased:* Accuracy: 0.84 Precision:0.92 Recall:0.80 F1-Score: 0.86 *Healthy:* Accuracy: 0.85 Precision:0.88 Recall:0.77 F1-Score: 0.82	They have used Faster R-CNN to detect the diseased and healthy leaves of apple tree from images which consists of many leaves. Their contribution is that the other studies generally have worked on images with one leaf, but their images contain many leaves. They have taken photos of trees leaves from different apple orchards in Yalova, Turkey for two years. They have annotated the images manually and showed the diseased and healthy leaves on the image. Their limitations are that the dataset needs to validation and accuracy must be improved.
(Chen et al., 2021)	Images collected by Taiwan Agricultural Research Institute, Council of Agriculture	F1-Score Mealybugs: 100% Coccidae: 89% Diaspididae: 97%	They have tried to detect three types of pests on the agricultural products before harming the product. They have focused on Mealybugs, Coccidae and Diaspididae which are the most seen pest in Taiwan. They have tested and compared the results of YOLOv4, SSD and Faster R-CNN, and found that YOLOv4 has provided the best results. They have also developed a mobile application based on cloud computation for farmers to detect pest on real time. Their contribution is that detection of different scale pets. Their limitation is that images in dataset contain only one type of pest, so the dataset is biased.
(Tassinari et al., 2021)	210-minute video captured in experimental farm at University of Bologna.	mAP: 0.64 - 0.66	They have detected dairy cows and collected information about them for sustainability of the highly utilized farming. They have used YOLOv3 to detect the target cows. They have selected four cows and they have gotten video contains these cows' daily activities. They have annotated the video for training process. As contribution, they can differentiate the individual cow using its color pattern, detect and track at video frame and store information about it. As limitation, they labeled each cow with two types of class separately according to left and right-side posture, but front and back postures were not considered. Each cow was represented with two class and when the number of cows increase the number of classes will increase with double.

(Hacıefendioğlu et al., 2021)	Images acquired by Google Earth Pro software after Palu earthquake in 2018.	Ground Failure area: 0.684 AP Buildings: 0.432 AP	They have used Faster R-CNN for detecting the ground failures and damaged buildings after the earthquake in Palu / Indonesia in 2018 from satellite images. They have mentioned that the ground failures are the most damaging reason, and they have tried to find these areas to assess the damages. Sometimes, it cannot reach the area according to collapsed buildings or the environment's structure, so object detection determines the ground failure area and damaged buildings quickly to intervene. As contribution, they detected target objects from satellite images with deep learning techniques unlike the others only compare the images before and after earthquake. As limitation, the accuracy must be improved, and they have detected failure area and damaged structure with two separate system.
(Fan et al., 2020)	COD10K (10000 images 78 object categories The researchers have generated data set), CHAMELEON (Skurowski et al., 2017) and CAMO (Le et al., 2019)	Train with CAMO, COD10K and extra, test with Chameleon, E-measure: 0.891 S-measure: 0.869 F-measure (FM): 0.74 MAE: 0.044	They have detected the objects which camouflaged in a natural environment with a new algorithm based on CNN. ResNet-50 has been used as a backbone network for getting feature maps, and with additional convolutions, down and upsampling procedures they have detected the camouflaged objects. As contribution, they have developed new algorithm and dataset, and they tested their algorithm with their own dataset and another public dataset. As limitation, accuracy can be improved and images containing more than one object can be used.
(Pi et al., 2020)	Volan2018 (65,580 annotated frame from YouTube) (Pi et al., 2020)	80.69% mAP videos from helicopter footage 74.48% mAP for videos from drone.	They have used aerial imagery for detecting damaged areas after a disaster. They have detected six classes: the flooded area, building roofs (damaged/undamaged), cars, debris, vegetation. They have used YOLOv2 as an object detection algorithm. First, they have pre-trained YOLO on COCO and VOC, and then with transfer learning, they have trained the system with the Volan2018 dataset. As contribution, they have detected targets from aerial image of different hurricanes and geographical areas getting with drone and helicopter and they have created new dataset. As limitation when the altitude changes the accuracy decreases.
(Deng et al., 2020)	KITTI (Geiger et al., 2012) and WayMo Open Dataset (Sun et al., 2020)	81.62% AP on KITTI 75.59% mAP on Waymo	They have developed a 3D object detection algorithm from point cloud represented as voxel for autonomous cars. They have used a 3D backbone network to get feature maps and converted them to Birds Eye View. Then with the Volvex RoI Pooling algorithm, they have found the region proposals. With a 2D backbone network, they have detected the

			cars. As contribution, they have developed new algorithm based on point cloud and test it on public dataset. As limitation, accuracy can be improved.
(Liu et al., 2020)	ECSSD (Yan et al., 2013) DUT-OMRON (Yang et al., 2013) PASCAL-S (Y. Li et al., 2014) HKU-IS (G. Li & Yu, 2016) DUTS (Wang et al., 2017)	*ECSSD:* FM: 0.934 MAE:0.044 *DUT-OMRON:* FM:0.809 MAE:0.055 *PASCAL-S:* FM:0.880 MAE:0.071 *HKU-IS:* FM:0.927 MAE:0.035 *DUTS:* FM:0.870 MAE:0.043	They have introduced an algorithm which consists of four main process part to detect a salient object. In the first part, there is a VGG16 backbone network. It is used for extracting shallow and deep features from the image at a different level. And the second part, each level of features is upsampled and utilized for getting enhanced feature maps. Results are sent to the third part for getting distinguishing features with upsampling operations. The last part of the algorithm generates the saliency map by using feature maps from deep to shallow, and finally determines the salient object. As contribution, they have developed new algorithm and tested it on public dataset and gotten higher accuracy than the compared studies. As limitation, on different dataset the accuracy and error rate are not stable, training process may be changed.
(Lee, 2020)	Drone pictures and videos from search engine, and labeled manually by them	The highest value of AP on third strategy 0.736 with one pass evaluation.	They have presented object detection and tracking algorithms on video with three types of combination strategies on detection and tracking. They have used Tiny YOLO v3 as an object detector and SiamRPN as a visual tracker. In the first strategy, they have focused on tracking using the object's previous position, and if the tracker cannot find the object, the object detector is assigned to find it and sent results to the tracker. In the second strategy, the object detector and visual tracker are switched, and both produce the results to use at the next step. In the third strategy, the object detector is the base process. If the detector fails, a request will be sent to the tracker. They have tested system for drone detection and tracking on videos. As contribution, they have developed new architecture to detect and track drones on video with network which has been trained a few static images. As limitation, they have said the speed must be improved for real time application and accuracy must be improved too.
(Ovodov, 2020)	DSBI (R. Li et al., 2018), Angelina Braille Images Datase (Ovodov, 2020)	*DSBI train and test:* FM on Point Base: 0.9994 FM on Character level: 0.9976	The author has developed an algorithm with some modification on RetinaNet architecture to recognize the Braille alphabet letters from the image. The letters are detected as point base and letter base. The point base detection has a higher f-measure value. As contribution, new dataset has been created and the images getting with a mobile phone can be used to

		Train with DSBI+ Angelina, and test with Angelina: FM on Point Base: 0.9991 FM on Character level:0.9981	recognize the text and results have been produced more quickly and accurately than the other studies. As limitation, the resultant class score can be given in AP metric to compare the general object detection algorithms, and at character level detection the FM values decrease.
(Hung et al., 2020)	BBBC038 (Caicedo et al., 2019), BBBC022 and BBC041 (Ljosa et al., 2012)	Nuclei detection: 82% mAP Cells with malarie: 78% mAP	They have used Keras R-CNN, which is based on Faster R-CNN for object detection in biological images. They have tried to detect two types of objects. First, they have detected cells from the image taken after staining the cell to highlight the DNA. And secondly, they have detected cells in blood smears that were taken from a patient with malarie parasite. As contribution, they have detected cell classes with different state of malaria parasite very quickly than traditional techniques. As limitation, the accuracy must be much more improved because the application area is human health.
(Wu et al., 2020)	Own dataset	57.6% mAP	They have used Mask R-CNN to detect mobile phones in a room for a wireless charging system. The study needs to find locations, numbers, and types of receivers in the Resonant Beam Charging system's coverage to start the charging process. After detection, these devices will be charged automatically by the system. They have focused on mobile phones. They have used different Intersection Over Union values (IoU), and different distances between the receiver and the charging system to measure accuracy. As contribution, their proposed system has decreased the detection time of mobile phones by one third. As limitation, accuracy must be improved and admitted dataset is needed.
(F. Han et al., 2020)	Own dataset	87.42% mAP	They have used CNN to detect objects underwater. According to some underwater imaging constraints as illumination and less image quality, object detection algorithms do not give satisfactory results. As contribution, they have compared their algorithms results with common algorithms and gotten high accuracy. They have said that their algorithm has enough accuracy to detect object for their underwater robot. As limitation, according to underwater imaginary challenging's the preprocessing procedures must be eliminated.
(J. Zhang et al., 2020)	GTSDB (Houben et al., 2013)	*GTSDB:*	They have developed an algorithm to detect the traffic signs based on Faster R-CNN with some additions. They have gotten features

	CCTSDB (J. Zhang et al., 2017) Lisa (Møgelmose et al., 2012)	Recall 90.5% Precision 98.7% *CCTSDB:* Recall 83.62% Precision 99.7% *Lisa:* Recall 85.6% Precision 98.9%	with ResNet50 as a backbone network and run an attention mechanism on features to find essential features of signs. Then they have fed the features in different scales to Region Proposal Network. After detecting regions, they have predicted the traffic sign. As contribution, they have detected traffic signs which has different dimensions and different environmental conditions, and they have tested their system on different public dataset. As limitation, average accuracy is needed to be improved.
(Zhu et al., 2020)	Own dataset	88.1% mAP	They have developed a CNN-based algorithm with Feature Pyramid Network to detect objects on power transmission line. They have used multi-scale feature maps to get high accuracy. For the pre-train, MS COCO dataset has been used. As contribution, they have developed bounding box regression algorithm with orientation angle because the orientation of the objects is also important for the transmission lines. As limitation, accuracy must be improved.
(Kechaou et al., 2020)	PASCAL VOC	52.0% mAP on test without post-processing	They have used an encoder-decoder architecture for generic object detection. In the encoder, they have used Resnet for getting a feature map, and in the decoder, they have used CNN and LSTM to predict object class and bounding box. For each step in the decoding layer, they have found objects in the attended region. It repeats this process until each object is detected with one class and one bounding box in the image. As contribution, they have eliminated post-processing procedures on images. As limitation, the AP values of small objects are low so the accuracy must be improved.
(Sorano et al., 2020)	Soccer match video from Italian league and their annotated from Wyscout.	No result for object detection	They have developed system with ResNet for feature extraction and YOLOv3 for object detection to analyze the soccer videos for annotating the pass movements. Their project will be used as a guide for training footballers and examine the team's performance. In the study, YOLOv3 has been re-trained to detect the ball and the footballer and bidirectional LSTM has been used for the classification of the movement whether it is a pass or not. Their contribution is that they have trained and tested the system with matches under the same conditions (as the stadium, lights) and different conditions and they have detected pass movement. As limitation on object detection, accuracy must be improved to detect the ball and footballer.

(W. Han et al., 2020)	Waymo Open Dataset (Sun et al., 2020)	Pedestrian: 60.1% mAP Vehicle: 51.0% mAP	They have detected objects from 360° rotation Lidar sensor's 3D point cloud data on traffic with low latency and high accuracy. They have divided the data into slices, and after spatial pooling, each slice has been sent to LSTM. Then feature pyramid of data has been extracted. From feature pyramid, objects have been detected with SSD and non-maximum suppression. As contribution, they have reduced the latency due to Lidar sensor and by this way, they have developed a system which can be used for autonomous car. As limitations, object representation in 3D point cloud must be learned for getting more accurate results.
(Zhao et al., 2020)	Visual Genome (Krishna et al., 2017)	For object detection 7.49 mAP	They have developed an algorithm to caption objects with attributes on image. They have used Inception-v3 (Szegedy et al., 2016) as backbone network. They have run an image attribute extractor based on Inception-v3 without bounding box regression for getting objects attributes. And, they have run backbone network for detection region proposals and object class. Finally, the attributes and region features have been sent to LSTM for captioning. As contribution, they have trained an end-to-end system which detects the objects and captions them. As limitation, the accuracy must be improved for better captioning.
(Kristo et al., 2020)	UNIRI-TID (Kristo et al., 2020)	97.93% AP	They have developed an object detection system based on YOLOv3 and new dataset to detect the people who move the border or forbidden area illegally or be terrorists. They have used thermal videos recorded in different weather conditions with different person pose state and movement characteristics. As contribution, they have adapted the RGB object detection algorithm to thermal images and create new dataset which includes different weather conditions like rainy, foggy. As limitation, if the heatmap of the objects are affected environmental conditions the accuracy decreases, and system can be tricked due to thermal imagery challenges.
(Loey et al., 2020)	Medical Masks Dataset (*Medical Mask*, n.d.) (in study Kaggle version) Face Masks Dataset (*Face Masks*, 2020)	81% AP	They have tried to detect faces with the medical mask, which has been a part of our life with Covid-19. They have aimed to produce guidance to governments for controlling compliance with the mask-wearing rule. They have gotten the best results using Resnet-50, YOLOv2, and ADAM optimizer(Kingma & Ba, 2014). As contribution, they have developed mean IoU method to increase the accuracy and they have used object detection

	to most important situation of our daily life. As limitation, the accuracy must be improved, and the number of the masked and unmasked face can be given.

4. Results and Discussions

With advances in deep learning and computer architecture, accuracy in object detection has increased, and processing time has been greatly reduced. Hereby as shown in part 3, object detection algorithms are applied in many research areas including health, country border security, epidemic disease, video/image captioning or annotation, transportation systems, autonomous vehicles, agriculture, damage assessment etc. Looking at the accuracy values, it can be said that the deep learning-based object detection algorithms provide solutions to many problems in different field.

In the line with the possibilities offered by deep learning, researchers need to decide whether they use common algorithms or develop new one. Because object properties such as different scales of object in same image, object density, the environment in which the object is placed etc. are challenging according to research area. Also, researchers need to decide the application will be offline or online. For online application, speed is important as accuracy. So, hardware must be also configured according to the computational needs.

As seen in Table 1, the accuracy must be improved for real field applications with eliminating limitations. Beside of this to develop the accuracy, admitted datasets are needed for many research areas. All the researching areas need to produce the results quickly so the researcher also focused on speeding up the computation and the speed of the algorithms must be reported.

5. Conclusion

In this study, common object detection algorithms, deep learning architectures and research in various fields using object detection have been chosen and briefly introduced. The methods, datasets and measurement metrics used in the reviews have been listed. This study aims to conduct literature research that will guide researchers who will work on object detection in their working area.

For future studies, researchers can focus on applying object detection to whole agriculture process because the quality of product is much more important according to global warming, shortage of resource and population crowd. For example, on hazelnut fields Erysiphe Pisi is causing Powdery Mildew disease which is the most damaging factor in recent times at North Region of Turkey and this disease should be detected and resolved immediately. For this task, an automatic disease detection and agricultural spraying system can be developed with unmanned aerial vehicles. Also, they can focus on reducing computational cost and speed up the process.

Researchers' Contribution Rate

All researchers have equal contribution rates.

Acknowledgement and/or disclaimers

There is no funding.

Conflict of Interest Statement

There is no conflict.

References

Bahdanau, D., Chorowski, J., Serdyuk, D., Brakel, P., & Bengio, Y. (2016). End-to-end attention-based large vocabulary speech recognition. *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 4945–4949. https://doi.org/10.1109/ICASSP.2016.7472618

Caicedo, J. C., Goodman, A., Karhohs, K. W., Cimini, B. A., Ackerman, J., Haghighi, M., Heng, C. K., Becker, T., Doan, M., McQuin, C., Rohban, M., Singh, S., & Carpenter, A. E. (2019). Nucleus segmentation across imaging experiments: the 2018 Data Science Bowl. *Nature Methods*. https://doi.org/10.1038/s41592-019-0612-7

Cevikalp, H., Benligiray, B., & Gerek, O. N. (2020). Semi-supervised robust deep neural networks for multi-label image classification. *Pattern Recognition*, *100*, 107164. https://doi.org/https://doi.org/10.1016/j.patcog.2019.107164

Chen, J.-W., Lin, W.-J., Cheng, H.-J., Hung, C.-L., Lin, C.-Y., & Chen, S.-P. (2021). A Smartphone-Based Application for Scale Pest Detection Using Multiple-Object Detection Methods. *Electronics*, *10*(4), 372. https://doi.org/10.3390/electronics10040372

Deng, J., Shi, S., Li, P., Zhou, W., Zhang, Y., & Li, H. (2020). Voxel R-CNN: Towards High Performance Voxel-based 3D Object Detection. *ArXiv:2012.15712*.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*(2), 179–211. https://doi.org/https://doi.org/10.1016/0364-0213(90)90002-E

Face Masks. (2020). https://www.kaggle.com/andrewmvd/face-mask-detection

Fan, D. P., Ji, G. P., Sun, G., Cheng, M. M., Shen, J., & Shao, L. (2020). Camouflaged object detection. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. https://doi.org/10.1109/CVPR42600.2020.00285

Geiger, A., Lenz, P., & Urtasun, R. (2012). Are we ready for autonomous driving? The KITTI vision benchmark suite. *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 3354–3361. https://doi.org/10.1109/CVPR.2012.6248074

Gers, F. A., Schmidhuber, J., & Cummins, F. (1999). Learning to forget: continual prediction with LSTM. *1999 Ninth International Conference on Artificial Neural Networks ICANN 99. (Conf. Publ. No. 470)*, *2*, 850–855 vol.2. https://doi.org/10.1049/cp:19991218

Girshick, R, Donahue, J., Darrell, T., & Malik, J. (2016). Region-Based Convolutional Networks for Accurate Object Detection and Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *38*(1), 142–158. https://doi.org/10.1109/TPAMI.2015.2437384

Girshick, Ross. (2015). Fast R-CNN. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*.

Hacıefendioğlu, K., Başağa, H. B., & Demir, G. (2021). Automatic detection of earthquake-induced ground failure effects through Faster R-CNN deep learning-based object detection using satellite images. *Natural Hazards*, *105*(1), 383–403. https://doi.org/10.1007/s11069-020-04315-y

Han, F., Yao, J., Zhu, H., & Wang, C. (2020). Underwater Image Processing and Object Detection Based on Deep CNN Method. *Journal of Sensors*, *2020*, 1–20. https://doi.org/10.1155/2020/6707328

Han, W., Zhang, Z., Caine, B., Yang, B., Sprunk, C., Alsharif, O., Ngiam, J., Vasudevan, V., Shlens, J., & Chen, Z. (2020). *Streaming Object Detection for 3-D Point Clouds*. http://arxiv.org/abs/2005.01864

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*. https://doi.org/10.1162/neco.1997.9.8.1735

Houben, S., Stallkamp, J., Salmen, J., Schlipsing, M., & Igel, C. (2013). Detection of traffic signs in real-world images: The German traffic sign detection benchmark. *The 2013 International Joint Conference on Neural Networks (IJCNN)*, 1–8. https://doi.org/10.1109/IJCNN.2013.6706807

Hung, J., Goodman, A., Ravel, D., Lopes, S. C. P., Rangel, G. W., Nery, O. A., Malleret, B., Nosten, F., Lacerda, M. V. G., Ferreira, M. U., Rénia, L., Duraisingh, M. T., Costa, F. T. M., Marti, M., & Carpenter, A. E. (2020). Keras R-CNN: library for cell detection in biological images using deep neural networks. *BMC Bioinformatics*, *21*(1), 300. https://doi.org/10.1186/s12859-020-03635-x

Indolia, S., Goswami, A. K., Mishra, S. P., & Asopa, P. (2018). Conceptual Understanding of Convolutional Neural Network- A Deep Learning Approach. *Procedia Computer Science*. https://doi.org/10.1016/j.procs.2018.05.069

Jelodar, H., Wang, Y., Orji, R., & Huang, H. (2020). Deep Sentiment Classification and Topic Discovery on Novel Coronavirus or COVID-19 Online Discussions: NLP Using LSTM Recurrent Neural Network Approach. *ArXiv:2004.11695*.

Kechaou, A., Martinez, M., Haurilet, M., & Stiefelhagen, R. (2020). Detective: An Attentive Recurrent Model for Sparse Object Detection. http://arxiv.org/abs/2004.12197

Kingma, D. P., & Ba, J. (2014). Adam: A

Method for Stochastic Optimization. http://arxiv.org/abs/1412.6980

Krishna, R., Zhu, Y., Groth, O., Johnson, J., Hata, K., Kravitz, J., Chen, S., Kalantidis, Y., Li, L. J., Shamma, D. A., Bernstein, M. S., & Fei-Fei, L. (2017). Visual Genome: Connecting Language and Vision Using Crowdsourced Dense Image Annotations. *International Journal of Computer Vision*. https://doi.org/10.1007/s11263-016-0981-7

Kristo, M., Ivasic-Kos, M., & Pobar, M. (2020). Thermal Object Detection in Difficult Weather Conditions Using YOLO. *IEEE Access*, *8*, 125459–125476. https://doi.org/10.1109/ACCESS.2020.3007481

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet Classification with Deep Convolutional Neural Networks. *Commun. ACM*, *60*(6), 84–90. https://doi.org/10.1145/3065386

Law, H., & Deng, J. (2018). CornerNet: Detecting Objects as Paired Keypoints. *Proceedings of the European Conference on Computer Vision (ECCV)*, 734–750.

Lawrence, S., Giles, C. L., & Fong, S. (2000). Natural language grammatical inference with recurrent neural networks. *IEEE Transactions on Knowledge and Data Engineering*, *12*(1), 126–140. https://doi.org/10.1109/69.842255

Le Cun, Y., Jackel, L. D., Boser, B., Denker, J. S., Graf, H. P., Guyon, I., Henderson, D., Howard, R. E., & Hubbard, W. (1989). Handwritten digit recognition: applications of neural network chips and automatic learning. *IEEE Communications Magazine*, *27*(11), 41–46. https://doi.org/10.1109/35.41400

Le, T.-N., Nguyen, T. V, Nie, Z., Tran, M.-T., & Sugimoto, A. (2019). Anabranch network for camouflaged object segmentation. *Computer Vision and Image Understanding*, *184*, 45–56. https://doi.org/https://doi.org/10.1016/j.cviu.2019.04.006

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Lee, D.-H. (2020). CNN-based single object detection and tracking in videos and its application to drone detection. *Multimedia Tools and Applications*. https://doi.org/10.1007/s11042-020-09924-0

Li, G., & Yu, Y. (2016). Visual Saliency Detection Based on Multiscale Deep CNN Features. *IEEE Transactions on Image Processing*, *25*(11), 5012–5024. https://doi.org/10.1109/TIP.2016.2602079

Li, J., Zhao, R., Sun, E., Wong, J. H. M., Das, A., Meng, Z., & Gong, Y. (2020). High-Accuracy and Low-Latency Speech Recognition with Two-Head Contextual Layer Trajectory LSTM Model. *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 7699–7703. https://doi.org/10.1109/ICASSP40776.2020.9054387

Li, R., Liu, H., Wang, X., & Qian, Y. (2018). DSBI: Double-Sided Braille Image Dataset and Algorithm Evaluation for Braille Dots Detection. *Proceedings of the 2018 the 2nd International Conference on Video and Image Processing*.

Li, Y., Hou, X., Koch, C., Rehg, J. M., & Yuille, A. L. (2014). The Secrets of Salient Object Segmentation. *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 280–287. https://doi.org/10.1109/CVPR.2014.43

Liu, Z., Li, Q., & Li, W. (2020). Deep layer guided network for salient object detection. *Neurocomputing*, *372*, 55–63. https://doi.org/https://doi.org/10.1016/j.neucom.2019.09.018

Ljosa, V., Sokolnicki, K. L., & Carpenter, A. E. (2012). Annotated high-throughput microscopy image sets for validation. In *Nature Methods*. https://doi.org/10.1038/nmeth.2083

Loey, M., Manogaran, G., Taha, M. H. N., & Khalifa, N. E. M. (2020). Fighting against COVID-19: A novel deep learning model based on YOLO-v2 with ResNet-50 for medical face mask detection. *Sustainable Cities and Society*. https://doi.org/10.1016/j.scs.2020.102600

***Medical Mask*.** (n.d.). Retrieved January 17, 2021, from https://humansintheloop.org/medical-mask-dataset/

Møgelmose, A., Trivedi, M. M., & Moeslund, T. B. (2012). Vision-based traffic sign detection

and analysis for intelligent driver assistance systems: Perspectives and survey. *IEEE Transactions on Intelligent Transportation Systems*. https://doi.org/10.1109/TITS.2012.2209421

Ovodov, I. G. (2020). Optical Braille Recognition Using Object Detection CNN. *ArXiv:2012.12412*.

Öztemel, E. (2012). *Yapay Sinir Ağları* (3rd ed.). Papatya Yayıncılık Eğitim A.Ş.

Pascalvoc. (n.d.). Retrieved January 11, 2020, from http://host.robots.ox.ac.uk/pascal/VOC/

Pavan, G. S., Kumar, N., Karthik N., K., & Manikandan, J. (2020). Design of a Real-Time Speech Recognition System using CNN for Consumer Electronics. *2020 Zooming Innovation in Consumer Technologies Conference (ZINC)*, 5–10. https://doi.org/10.1109/ZINC50678.2020.9161432

Pi, Y., Nath, N. D., & Behzadan, A. H. (2020). Convolutional neural networks for object detection in aerial imagery for disaster response and recovery. *Advanced Engineering Informatics*, *43*, 101009. https://doi.org/https://doi.org/10.1016/j.aei.2019.101009

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 779–788.

Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*(6), 1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031

Sardoğan, M., Özen, Y., & Tuncer, A. (2020). Faster R-CNN Kullanarak Elma Yaprağı Hastalıklarının Tespiti. *Düzce Üniversitesi Bilim ve Teknoloji Dergisi*, 1110–1117. https://doi.org/10.29130/dubited.648387

Skurowski, P., Abdulameer, H., Błaszczyk, J., Depta, T., Kornacki, A., & Kozieł, P. (2017). *CHAMELEON*. http://kgwisc.aei.polsl.pl/index.php/en/dataset/63-animal-camouflage-analysis

Sorano, D., Carrara, F., Cintia, P., Falchi, F., & Pappalardo, L. (2020). *Automatic Pass Annotation from Soccer VideoStreams Based on Object Detection and LSTM*. http://arxiv.org/abs/2007.06475

Sun, P., Kretzschmar, H., Dotiwalla, X., Chouard, A., Patnaik, V., Tsui, P., Guo, J., Zhou, Y., Chai, Y., Caine, B., Vasudevan, V., Han, W., Ngiam, J., Zhao, H., Timofeev, A., Ettinger, S., Krivokon, M., Gao, A., Joshi, A., … Anguelov, D. (2020). Scalability in Perception for Autonomous Driving: Waymo Open Dataset. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2443–2451. https://doi.org/10.1109/CVPR42600.2020.00252

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the Inception Architecture for Computer Vision. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. https://doi.org/10.1109/CVPR.2016.308

Tassinari, P., Bovo, M., Benni, S., Franzoni, S., Poggi, M., Mammi, L. M. E., Mattoccia, S., Di Stefano, L., Bonora, F., Barbaresi, A., Santolini, E., & Torreggiani, D. (2021). A computer vision approach based on deep learning for the detection of dairy cows in free stall barn. *Computers and Electronics in Agriculture*, *182*, 106030. https://doi.org/10.1016/j.compag.2021.106030

Tsai, W.-J., Huang, Z.-J., & Chung, C.-E. (2020). Joint Detection, Re-Identification, And Lstm In Multi-Object Tracking. *2020 IEEE International Conference on Multimedia and Expo (ICME)*, 1–6. https://doi.org/10.1109/ICME46284.2020.9102884

Turan, S., & Bilgin, G. (2019). Semantic nuclei segmentation with deep learning on breast pathology images. *2019 Scientific Meeting on Electrical-Electronics Biomedical Engineering and Computer Science (EBBT)*, 1–4. https://doi.org/10.1109/EBBT.2019.8741715

Ufldl Tutorial. (n.d.). UFLDL Tutorial. Retrieved December 24, 2020, from http://ufldl.stanford.edu/tutorial/

Vinayakumar, R., Soman, K. P., & Poornachandran, P. (2017). Applying

convolutional neural network for network intrusion detection. *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 1222–1228. https://doi.org/10.1109/ICACCI.2017.8126009

Wang, L., Lu, H., Wang, Y., Feng, M., Wang, D., Yin, B., & Ruan, X. (2017). Learning to Detect Salient Objects with Image-Level Supervision. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 3796–3805. https://doi.org/10.1109/CVPR.2017.404

Wu, A., Zhang, Q., Fang, W., Deng, H., Jiang, S., & Liu, Q. (2020). *Mask R-CNN Based Object Detection for Intelligent Wireless Power Transfer*. http://arxiv.org/abs/2004.10021

Xiao, H., Xu, J., & Shi, J. (2020). Exploring diverse and fine-grained caption for video by incorporating convolutional architecture into LSTM-based model. *Pattern Recognition Letters*, *129*, 173–180. https://doi.org/https://doi.org/10.1016/j.patrec.2019.11.003

Yan, Q., Xu, L., Shi, J., & Jia, J. (2013). Hierarchical Saliency Detection. *CVPR 2013*.

Yang, C., Zhang, L., Lu, H., Ruan, X., & Yang, M.-H. (2013). Saliency detection via graph-based manifold ranking. *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference On*, 3166–3173.

Yuan, D., Li, X., He, Z., Liu, Q., & Lu, S. (2020). Visual object tracking with adaptive structural convolutional network. *Knowledge-Based Systems*, *194*, 105554. https://doi.org/https://doi.org/10.1016/j.knosys.2020.105554

Zhang, J., Huang, M., Jin, X., & Li, X. (2017). A real-time Chinese traffic sign detection algorithm based on modified YOLOv2. *Algorithms*. https://doi.org/10.3390/a10040127

Zhang, J., Xie, Z., Sun, J., Zou, X., & Wang, J. (2020). A Cascaded R-CNN with Multiscale Attention and Imbalanced Samples for Traffic Sign Detection. *IEEE Access*. https://doi.org/10.1109/ACCESS.2020.2972338

Zhang, W., Xu, L., Li, Z., Lu, Q., & Liu, Y. (2016). A Deep-Intelligence Framework for Online Video Processing. *IEEE Software*, *33*(2), 44–51. https://doi.org/10.1109/MS.2016.31

Zhao, D., Chang, Z., & Guo, S. (2020). Cross-scale fusion detection with global attribute for dense captioning. *Neurocomputing*. https://doi.org/10.1016/j.neucom.2019.09.055

Zhu, J., Guo, Y., Yue, F., Yuan, H., Yang, A., Wang, X., & Rong, M. (2020). A deep learning method to detect foreign objects for inspecting power transmission lines. *IEEE Access*. https://doi.org/10.1109/ACCESS.2020.2995608