

Integrated Real Time Image Processing In Robotic Automation Line

¹Ismail Atalay, ¹Oğuz Alper İsen, ¹Emin Cantez, ¹Serkan Aydın, ¹Onur Akyel
¹Coşkunöz Kalıp Makina Sanayi ve Ticaret A.Ş., Bursa, Turkey

Abstract

Robotic automation systems includes higher production rates and increased productivity, more efficient use of materials, better product quality, improved safety, and reduced factory lead times. Higher output and increased productivity have been two of the biggest reasons in justifying the use of automation. Despite the claims of high quality from good workmanship by humans, automated systems typically perform the manufacturing process with less variability than human workers, resulting in greater control and consistency of product quality. Also, increased process control makes more efficient use of materials, resulting in less scrap.

Despite all these advantages, the final product control is still carried out by workers. In this study, it is planned to prevent quality problems with a video processing that integrates with the robotic automation line in order to solve the final quality problem. This system, which controls the movements of the worker in the predetermined motion routes with the signals coming from the line. It aims to ensure that the products going to the customer are completely error free.

Key words: Robotic Automation, Quality, Reduced Cost, Image Processing, Less Scrap

1. Introduction

In developing automation technologies, it is observed that production times are much shorter nowadays. With these automation technologies, we can access information such as the number of products produced, their type and production time. However, the number of fully automatic production lines is not preferred due to price / performance data. Instead, a common cell is created with the production worker and the automation line system, and thus production continues at very high levels of accuracy.

At this stage, by using an artificial intelligence algorithm that controls the automation lines, it can record potential errors that will occur during production by observing the movements of the operation worker of the system.

This artificial intelligence algorithm, in which there are many different versions of tracking applications, monitors what the operation worker needs to do for correct production by collecting recording data with the help of a video camera and examines this event when it sees an activity outside the cycle that needs to be done, and reports this if it is out of malfunction or predetermined errors and production reduces scrap cost.



Figure 1: Automation line

Artificial intelligence, enabling machines to solve any problem like humans is a science. Usually this takes the characteristic of human intelligence and applies an algorithm to machines. How the machines will react according to the requirements can make flexible or effective approaches. Artificial intelligence is related to many branches of science, Although it seems to be directly related to computer science, mathematics, psychology, philosophy It is closely related with different disciplines such as [2].

When considering computers, mechanical It is very convenient in calculation. These machines are suitable for people to do It allows to do simple non-existent problems accurately, efficiently and faster. When the problem gets more complex, things get very difficult for machines. Machines like humans they cannot perceive different situations or adapt to new situations. The main task of artificial intelligence is to improve the behavior of machines in complex problems, to solve problems quickly [3].

There are many different approaches to artificial intelligence. Some of these approaches seems very useful and alternative methods are constantly produced. End user who have been doing research on artificial intelligence in years often think about specific problems. They have done studies and worked for the solution of these problems. These problems has produced a large number of solutions with high accuracy and effectively have had [2].

2. Materials and Method

Deep learning helps machines understand the world and solve complex problems. It is the most popular approach to developing artificial intelligence. Deep learning method, multiple layers of nonlinear processing units for feature extraction and conversion Uses. Successive layers take the output of the previous layer as input and process does [3]. The data used in deep learning has a large number of properties. Some of these While some are high-level features, some are low-level features. Lower level features higher level they derive the features. The basis of deep learning is exactly the data itself. There is learning. When looking at deep learning methods, it is efficient algorithms for the extraction of hierarchical features that best represent the data used in [3].

In short, while old algorithms were dependent on people, deep learning algorithms distinguishing features are that it can learn by itself. Deep learning algorithms difference

from machine learning algorithms; will be able to process very large data and this data. It requires hardware with high performance. High performance equipment has an efficient structure for deep learning development and is the main reason for the abbreviation [3].

Also image processing methods and Python [9] libraries are functional suites that aid in deep learning. It is planned to use object identification functions with deep learning methods, after the image to be brought to the deep learning stage is taken raw and a cleaner image is obtained by using OpenCV[10].

Artificial Neural Networks (ANN):

ANN, inspired by the human brain, imitating the way the brain works, to perform functions such as learning, generalizing and producing new knowledge are machine systems developed for ANNs are different from traditional algorithms. They use calculation method. ANNs have fault tolerance and makes use of experiences, but also depends on hardware.

Artificial neural networks includes, input, weights, bias, aggregation function, activation consists of function and output. A simple artificial neural network example is given in Figure 3. Inputs are data coming to neurons. Entered a neuron as it could come from another neuron

It can also come from the outside world. The data that are arriving as inputs are recorded with a weight value. It is multiplied and summed by the bias value. In this way, the inputs will be produced It is known to adjust the effect on it. Addition function is the input by multiplying It is the function that gives the net input of that cell by adding its values. Depends on ANN Sometimes the value of the entries is important, sometimes the number of incoming entries is important[7].

The activation functions take the net inputs and specify an output against that input. Many It has an activation function. Linear activation function, sigmoid activation function, tanh activation function can be given as an example. Activation The output of the function is the output of the neuron. If this output value is desired, another It can be used as the input value of the neuron or as an output to the outside world [7].

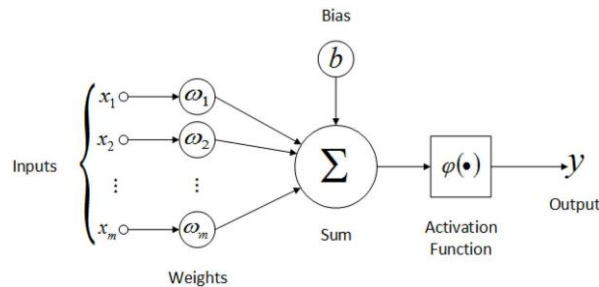


Figure 2: Artificial Neural Network

Recurrent Neural Networks (RNN):

Recurrent neural networks (RNNs) are neural networks with internal memory, so they are highly powerful neural networks. Previous inputs they received due to memory they can remember, and this increases their probability of accurately predicting the future. In traditional neural networks, all input data are assumed to be independent from each other.

The basic idea of recurrent neural networks is to use sequential information. While theoretically it is accepted that they can use the information in very long sequences, in practice only it is known to return to the information a few steps back.

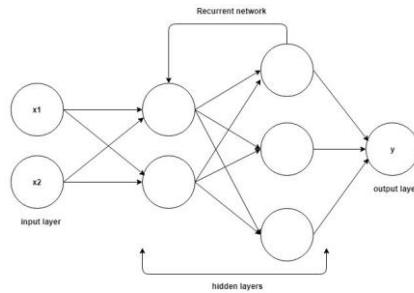


Figure 3: Recurrent Neural Network

Convolutional Neural Networks (CNN):

People look at an image and where in the image makes quick deductions of what they are. Modeling the human visual system CNNs created with today's computer vision field have achieved significant success.

It is. Such as object recognition, object classification, object tracking, sentence modeling It is used in many areas. CNNs consist of many layers. These are the introduction layer, convolution layer, ReLu, pooling layer, fully connected layer, dropout layer, classification layer and output layer [12]. Sample one The convolutional neural network is given in Figure [4].

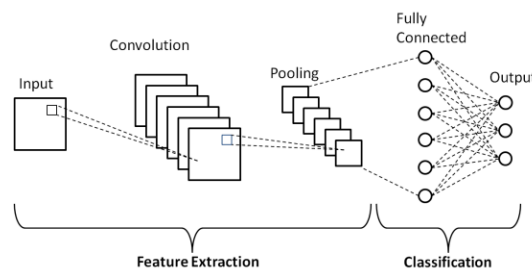


Figure 4: Convolutional Neural Network

Object Detection and Tracking Algorithms:

Object detection is used in almost every field today. Pedestrian detection, There are also many application areas such as people counting, autonomous vehicles and face detection. Object the general purpose of detection is to create a new image of a previously defined object class in the system. and identifying the positions of each object detected in this image.

Using a rectangle that encloses it. There are many accurate and fast algorithms for object detection and tracking. In deep learning The simplest deep learning approach for object detection and object tracking is convolution nerve network are. An output class is obtained for each input image. The image This system can also be used to detect various objects. The biggest problem with this approach is that the object in the image can be of different sizes. As a result of this problem, it will cause the calculation time to take considerably longer. Many for object detection neural network is available [13]. To these; R-CNN, Fast R-CNN, Faster R-CNN, SSD, YOLO as an example.

Each region that can be an object in an image extracted in the R-CNN is passed through CNN one by one. In Fast R-CNN, for regions of interest that can be objects in an image, CNN is run once and a feature map is created. Even if it does a faster object search than R-CNN, it uses selective search, which is a slow and time consuming process is not fast enough for [13]. Selective search is used for region recommendations in R-CNN and Fast R-CNN algorithms.

Selective search is a slow and time consuming process, which affects the performance of the network. Faster R-CNN, on the other hand, uses a different network to estimate region suggestions instead of using the selective search algorithm for region suggestions, and this different network Faster R-CNN considerably accelerates it . The comparison of R-CNN, Fast R-CNN and Faster R-CNN algorithms is given in Table 1.

	R-CNN	FAST R-CNN	FASTER R-CNN
Test times per image	50s	2s	0.2s
Accelaration	1x	25x	250x
Average Prediction (%)	66	66.9	66.9

The YoloV3 algorithm is very fast because it passes the image through the neural network in one go and predicts the class and coordinates of all objects. The most important feature that distinguishes YoloV3 from other object detection algorithms is that it is ahead of its competitors in general average precision (mAP) values in real-time object detection and tracking.

Figure 7 shows the comparison graph of the algorithms used in object detection and tracking with 0.5IoU (mAP-50) in the COCO data set. In this study, YoloV3 algorithm, which is very good in terms of both speed and accuracy, was chosen.The YoloV3 algorithm is illustrated in Section 4.1.

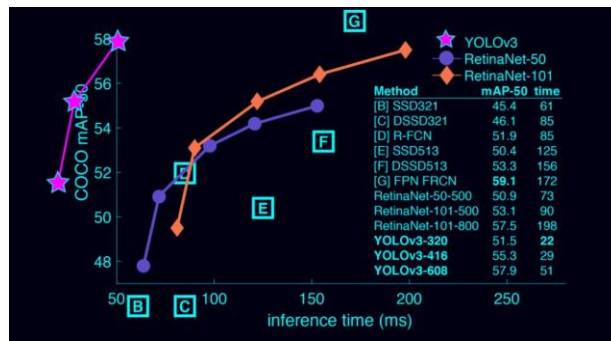


Figure 5: Comparison between object detection algorithm

3. Results

3.1 Selecting the fast and accurate method:

In addition to accuracy, fast detection is also very important during the monitoring of the operator in the production area. For this reason, the YoloV3 algorithm, which works at both high speed and high accuracy, was chosen for this problem and Darknet was used as a framework. The YoloV3 algorithm divides the input picture into NxN pieces. Each piece is responsible for finding out if there is an object in it. Each of these parts constitutes as many predictors as the number of anchor boxes. Using the Anchor boxes method The main reason is that if there is more than one object in any part, the algorithm detects all objects in this part. Within the prediction vector, the confidence score includes the x and y coordinates of the object's midpoint, the probability of the connected class. The confidence score shows how sure the part is whether there is an object in it. The probability of bound class is the predicted value as much as the number of classes in the model . Since there is 1 class specific to this problem, there will be 1 connected class possibility. After object detection, the main purpose is to draw a rectangle around the detected object

While the algorithm is detecting an object, it can also create more than one unnecessary box for an object or boxes in areas where there are no objects. In order to throw these boxes, boxes below the previously determined threshold level are solved based on the confidence score .

3.2 Data set creation and labeling process:

Video recordings started with the start signal received from the system with the camera placed on the roof during production. Then, a data set consisting of video files was created by making approximately 20 correct and 2 incorrect productions. Each of these videos was 2250 x 1250 pixels in size.

Since a data set of this size will increase the duration of the training process considerably, the dimensions of the photos can be changed to 720x480 pixels. All photos in the data set that have

changed dimensions are labeled. The purpose of the labeling process is to record the coordinates of the object in a text file. After the photo labeling, the data set was broken down as training (train), test, validation. Of the 5000 photographs, 4000 were taken as training data, 750 as test data, and 2550 as verification data.

3.3 Setting neural network parameters

It is very important to adjust the parameters of the neural network used according to the characteristics of the data set. These parameters can differ between neural networks. In addition, a setting made in the same neural networks may not affect two different data sets in the same way. In this study, Yolo parameters are summarized.

In YoloV3, the dimensions of the input image are taken into the system as 608x608 by changing the size. If desired, it can be taken into the system as 418x418. But the larger this size, the better the detection of small objects. Besides These dimensions are known to be increased in multiples of 32 according to the characteristics of the video card to be trained.

The large and diverse data set is an important factor for learning. The larger the data set, the better the learning, but this increases the training time. In addition to this, the diversity of the data set is also very important, having data in all combinations that may be encountered increases the accuracy considerably. In cases where sufficient data cannot be increased, the system can be made more sensitive by data augmentation. Increasing data is to obtain new data in many different ways such as taking the symmetry of the image according to the axes, changing the color ratios.

3.3 Training and Test Process

After adjusting the neural network parameters by considering the data set, the training process is carried out. During the training process, batch and learning rate, which are among the neural network parameters, have a very important place. Batch value is the number of photos read in one iteration. During the training process, each iteration also renews the weight files.

It is recommended that the batch value data set is designed according to the characteristics. When the batch value increases, the time spent per iteration increases and the training time increases.

During the training, an average loss value per iteration is calculated. This value shows how close the estimated values are to the reality. The average loss value changes according to the value of the learning rate. One of the most appropriate methods for learning rate is to keep the learning rate high at the beginning and gradually decrease it. If the learning rate is kept high, the system is affected more by the data, and if the learning rate is kept low, the training takes longer. In addition, if the learning speed is chosen too low, it may cause failure to reach the global optimum value. The effect of various learning rates on convergence is given in Figure [x].

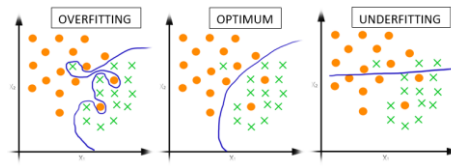


Figure 6: Learning rate effect

3.3 Experimental Result

The data set consists of 5000 photographs. After the data set was created, the labeling process was performed for all photos in the data set. Tagging for a photo took 25 seconds. The tagging process for all photos took approximately 15 hours. After the data set is labeled, the neural network parameters are set. Then, the training process was started and the average error graph was drawn. Learning rate has an important place for the error function. The average error function is not expected to decline neither too quickly nor too slowly.

Average error value is expected to decrease optimally. In this thesis study, the learning rate that decreases the average error value optimally was selected by training with different learning rates. In Figure 13, while the learning rate is 0.01, the average error graph is given. As can be seen from the figure, it will be seen that the average error graph decreases very rapidly. Although this learning rate has accelerated the training process, it is highly affected by the data set. For this reason, the learning rate should be reduced before training.

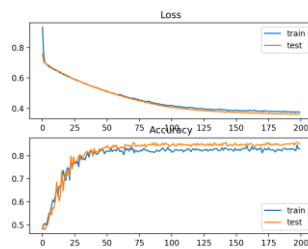


Figure 7: Learning rate = 0.01 result

In Figure [x], while the learning rate is 0.001, the average error graph is given. As can be understood from the figure, the average error graph decreases optimally. In addition, due to the decrease in the value of the learning rate, the duration of the training process has increased.

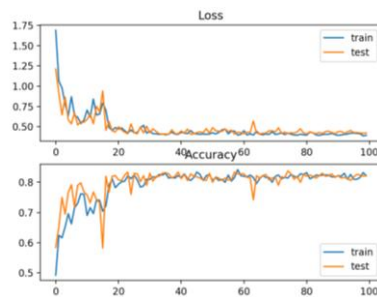


Figure 8: Learning rate = 0.001 result

The training process was stopped when the average error value reached about 0.009 levels and the change in average error values was very slow. The training process took 12 hours. Table 3 shows the time spent per iteration.

Iteration	Elapsed Time
100	10 min
1000	1 hours 20 min
5000	12 hours

Table 1: Elapsed time

Performance test was conducted after the training process. Performance test was performed using the PR curve, which is one of the performance metrics. Performance test is given in Figure 14.

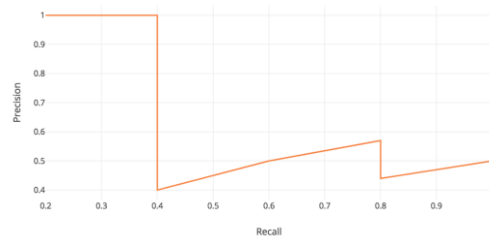


Figure 9: Performance Result

Besides the high accuracy of the test process, the detection is expected to make the process fast. The video was recorded at 25 fps. Yolov3 successfully performs the object identification process 25 times per second.

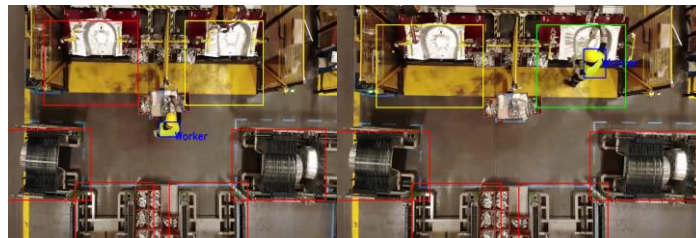


Figure 10: Bounding box and results

The network predicts 4 coordinates for each bounding box, t_x , t_y , t_w , t_h . If the cell is offset from the top left corner of the image by (c_x, c_y) and the bounding box prior has width and height p_w, p_h , then the predictions correspond to:

$$\begin{aligned}
 b_x &= \sigma(t_x) + c_x \\
 b_y &= \sigma(t_y) + c_y \\
 b_w &= p_w e^{t_w} \\
 b_h &= p_h e^{t_h}
 \end{aligned}$$

Figure 11: Rectangle formula

Each YOLO layer has 255 outputs: 85 outputs per anchor [4 box coordinates + 1 object confidence + 80 class confidences], times 3 anchors. If you use fewer classes, reduce filters to filters=[4 + 1 + n] * 3, where n is your class count. This modification should be made to the layer preceding each of the 3 YOLO layers.

N = 1 classes;

Filters = [4 + 1 + 1] * 3

Filters = 15

Conclusions

The results of real-time object detection studies in the production line, which are important issues for quality control processes in the automotive industry and smart production, are presented. Tracking methods were trained using YoloV3 algorithms. As long as it does not work outside the loop created, the object recognition process works correctly. However, when switching to different production styles (sample cross fixture work), considering the situation where the trained model is worked on a single fixture, errors may occur. This situation will be overcome as new data are collected. Since the video process requires high memory capacity and processing power, the capacity of the selected hardware may cause problems in the future. Therefore, future studies are planned to transfer the data to the cloud and work completely online.

Acknowledgements

This research is supported Coşkunöz Kalıp Makina Sanayi ve Ticaret A.Ş

REFERENCES

- [1] <https://medium.com/t%C3%BCrkiye/yapay-zekan%C4%B1n-tarih%C3%A7esi-vegeli%C5%9Fim-s%C3%BCreci-cb4c73deb01d>
- [2] <https://techy.works/genel/yapay-zeka-nedir/>
- [3] Kızrak, M. ve Bolat, B., (2018),"Derin Öğrenme ile Kalabalık Analizi Üzerine Detaylı Bir Araştırma", Bilişim Teknolojileri Dergisi, c.11, no.11, s. 263-286.
- [4] https://www.wipo.int/edocs/pubdocs/en/wipo_pub_1055/2019.pdf
- [5] Şeker, B. ve Diri, B. ve Balık, H., (2017),"Derin Öğrenme Yöntemleri ve Uygulamaları Hakkında Bir İnceleme", Gazi Mühendislik Bilimleri Dergisi, c.3, no.3, s 47-64.
- [6] Keskenler, M., Keskenler, E., (2017),"Geçmişten Günümüze Yapay Sinir Ağları ve Tarihçesi", Takvim-i Vekayi Dergisi, c.5, no.2, s.8-18.
- [7] <https://yapayzeka.ai/yapay-sinir-aglarinin-calisma-prensibi/2018/1/>
- [8] <https://pjreddie.com/darknet/yolo/>
- [9] <https://docs.python.org/3/>
- [10] <https://docs.opencv.org/4.4.0/>
- [11] Cassidy, A. P.; Deviney, F. A. Calculating Feature Importance in Data Streams with Concept Drift Using Online Random Forest. 2014 IEEE International Conference on Big Data (Big Data), 2014.
- [12] <https://stanford.edu/~shervine/l/tr/teaching/cs-230/cheatsheet-convolutional-neuralnetworks>
- [13] <https://towardsdatascience.com/r-cnn-fast-r-cnn-faster-r-cnn-yolo-object-detectionalgorithms-36d53571365e>