Uluslararası İleri Doğa Bilimleri ve Mühendislik Araştırmaları Dergisi
Sayı 7, S. 239-243, 6, 2023

Araştırma Makalesi

IJANSER

https://as-proceeding.com/index.php/ijanser
ISSN: 2980-0811

International Journal of Advanced Natural Sciences and Engineering Researches
Volume 7, pp. 239-243, 6, 2023

Research Article

Symbol Detection with Deep Learning

Fatih Aslan[1,2,*] and Yaşar Becerikli[1,3,**]

[1] *Computer Engineering Department, Kocaeli University, Kocaeli/Turkiye*
[2] *Computer Engineering Department, Yalova University, Yalova/Turkiye*
[3] *Digital Forensics Specialization Department, Forensic Medicine Institution, İstanbul/Turkiye*

(fatihaslan82@gmail.com) Email of the corresponding author
**(ybecerikli@kocaeli.edu.tr)*

(Received: 07 July 2023, Accepted: 24 July 2023)

(5th International Conference on Applied Engineering and Natural Sciences ICAENS 2023, July 10 - 12, 2023)

ATIF/REFERENCE: Aslan, F. & Becerikli, Y. (2023). Symbol Detection with Deep Learning. *International Journal of Advanced Natural Sciences and Engineering Researches*, 7(6), 239-243.

Abstract – This study shows the feasibility of detecting some symbols from a distance. Therefore, it will be possible to use these symbols as identification of things, e.g. people, vehicles, objects, information labels etc. There is such a necessity especially for construction sites to know the rough location of workers. In order to achieve this goal, firstly some two-dimensional symbols that are distinguishable from each other by both human eye and computer are generated. Afterwards, they are put on a three dimensional surface and annotated. Lastly, the labeled data are trained by the deep learning method YOLO version 5. Results indicate that it is highly efficient to use the chosen symbols for recognizing from a distance. For now, up to 10 meter images are taken into account. Real time tests are taken and the accuracy is about 85%. As the dataset will be combined with real-life images, the accuracy results will be even higher.

Keywords – Symbol Detection, Symbol Recognition, Deep Learning, Identification

I. INTRODUCTION

Computer Vision has been highly developed and many techniques are generated in order to gain information about objects and automatically classifying objects on images has been intensely studied. Therefore, it is very important to draw meaningful conclusions from images such as human recognition, specific object recognition, label reading etc.

Recognizing labels and symbols via camera is crucial for many industrial and individual requirements such as autonomous systems and blind people assistance. Below, some studies are mentioned about symbol detection and recognition.

QR (quick response) code reading distance is studied in [1]. It is shown that a QR code with 10cmx10cm size can be read up to 300 cm. In the study, Maximum scanning distance changes linearly with respect to QR code size. If the angle of image containing QR code is below 45 degrees, then QR code scanning is not practically usable. Also, reading distance of a QR code is close to ten times of QR code size according to the article in [2].

A challenging automatic road traffic sign is studied in [3]. The signs were determined by using Convolutional Neural Network (ConNN) and Viola-Jones structure for the detection and recognition of traffic signs, even if they did not comply with the standard signs. Viola-Jones method was used to identify possible regions of traffic signs from the whole image, then these regions were transferred to the deep learning network. In the study, a 90% success rate was achieved.

In a deep learning based study using YOLO version 3 and Spatial Pyramid Pooling (SPP), four different traffic signals were studied [4].

In [5], a study was conducted which is sensitive to different light conditions and scale changes based on Histogram of Gradient and Support Vector Machines. The proposed approach was tested in three variants, each aimed at detecting one of the three categories of mandatory, prohibitory and danger traffic signs, according to the experimental setup of the recent German Traffic Sign Detection Benchmark competition.

In another study on traffic signs with contour detection, the image was segmented using the contour descriptor method and HSV (Hue Saturation Value) color space. With this identifier, each shape is determined as triangle, square, round or hexagon and appropriate shape identification is made from the related group [6].

In a study on symbol recognition, various shapes that are composed of 9x9 matrices were tried to be recognized. The images of these shapes in a square box were first converted into gray level and reduced to binary values with a certain threshold value. Then the features were extracted and matched with the template information in order to recognize images [7].

Also automatic vehicle plate recognition systems have been under investigation so far. In [8], an SVM based algorithm was used and accurate results were taken up to 5 meters.

According to the papers mentioned above, barcode and QR code like readings are not feasible from a distance in order to gain meaningful information about labeled objects. Therefore some symbols can be used accordingly in order to recognize objects.

II. METHOD

In order to achieve such a goal, some symbols are created and these symbols are placed on a spherical surface. Many images of those symbols are taken from different angles and labeled in order to build a dataset. Afterwards, this dataset is trained by YOLO v5 deep learning method. The method and results are given below.

A. *Proposed Symbols*

In the scope of labeling objects, first different shapes are considered and 14 of them are chosen as in Fig-1. They are treated carefully and selectively in order to distinguish from a distance.

Fig.1 Symbols

As seen, “x” used and “+” is not used for the sake of least confusion. Same situation applies to the symbol “arrow-up”.

B. *Labeling*

First, the chosen symbols are put on a spherical surface using open source Blender animation and design program instead of physically drawing and sticking on a surface. This speeded up labeling and training processes.

A surface is created with a shape like a hat or a helmet. This is chosen to be able to label people wearing hardhat (helmet) in construction sites.

Then, the symbols are put on the surface one by one as seen in Fig-2 and taken a picture of them from 50 different angles with a static light source. This process is done semi-automatically thanks to the program.

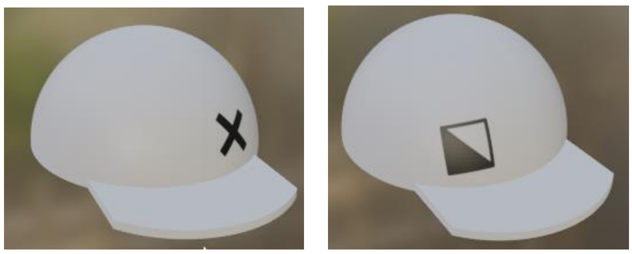

Fig. 2 Symbol Placing On A Surface

After labeling symbols, the helmets are combined with random real life backgrounds including construction site scenes as in Fig-3 in order to generate realistic situations and increase accuracy.

Backgrounds are taken from a famous helmet dataset [9].

Fig. 3 Annotating Symbols

C. *Training with YOLO*

Unlike traditional machine learning methods, artificial neural networks based deep learning methods have been a great success and gained popularity in case of accuracy and speed for object detection and recognition. Deep learning methods are grouped into two main categories: One-Stage (Proposal-Free) and Two-Stage (Proposal). You Only Look Once (YOLO) is a One-Stage deep learning method which takes an input image once in order to detect coordinates of searched objects if they exist.

As a One-Stage method, YOLO has several advantages over Two-Stage methods. The first is that it is fast. YOLO, which sees object detection as a regression problem, normally works at 45 fps (frame per second), while fast versions go up to 150 fps. Unlike the sliding window and region estimation techniques, the YOLO method is based on the whole image in the training and testing phase, so background errors are reduced by half compared to two-step methods. Finally, YOLO learns the generalizable properties of objects and thus gives less errors even in unpredictable images.

In YOLO [10], the input image is divided into SxS grids. If the center of an object coincides with a grid cell, that cell is made responsible for detecting the object. Each cell estimates the bounding boxes represented by B and the confidence score. These confidence values are the value that determines whether the relevant bounding box contains the object. This confidence value is formally defined as:

$$Pr(Object) * IoU^{truth}_{pred}$$

Here, IoU (Intersection over Union) is the ratio between the predicted box and the ground truth.

Each bounding box believed to contain the object estimates five parameters: x,y,w,h and c. The (x,y) coordinates determine the center of the bounding box, w and h determine the width and height of the bounding box, and the confidence value c determines the IoU ratio between the labeled bounding box and the predicted box.

Each grid cell also predicts C conditional class probabilities of Pr($Class_i$ | Object). Without looking at number of boxes B, only one set of class probabilities are predicted for each grid cell.

For testing, the conditional class probabilities and the individual box confidence predictions are multiplied.

The following formula gives confidence scores of classes for each box. These scores defines quality of predictions.

$$\Pr(\text{Class}_i|\text{Object}) * \Pr(\text{Object}) * \text{IOU}^{\text{truth}}_{\text{pred}} = \Pr(\text{Class}_i) * \text{IOU}^{\text{truth}}_{\text{pred}}$$

Since 2016, YOLO has evolved and improved many times. YOLO NAS is the current latest version.

In this study, YOLO v5 which was released in 2020 is chosen. The other versions will be tested later on.

The created dataset which consists of 1400 images is trained with YOLO v5. The batch size is 32 and the number of epochs is 50. The images are taken apart 20% and 80% for testing and training respectively.

III. Tests And Results

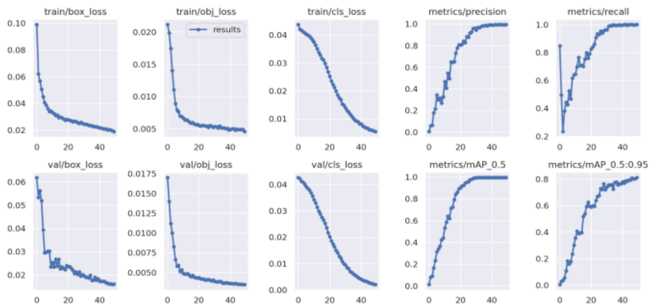

Fig.4 Results of Training Symbols

The training results are given in Fig-4. Here, precision is the ratio True Positives divided by True Positives and False Positives. Also, recall is the division of True Positives by True Positives and False Negatives. According to the results, the precision and recall values are quickly converged as the dataset is synthetic.

With the generated model, there are some simulation and real time tests conducted as seen in Fig-5 and Fig-6.

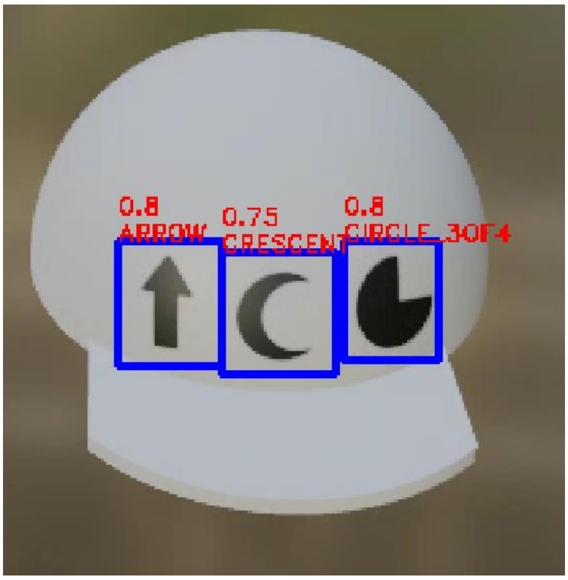

Fig. 5 Simulation Tests

The symbols are used in triple combination for this study. As 14 different symbols are used, it means 14x14x14 = 2744 different symbol sets can be generated. Therefore, most construction sites can be handled that way as the number of combinations is quite high.

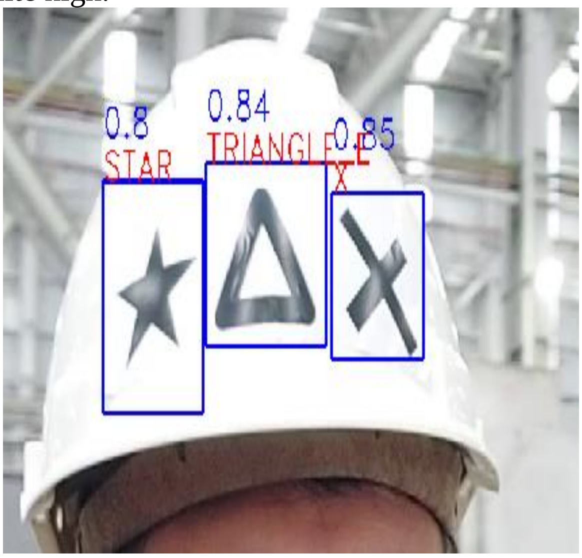

Fig. 6 Real Time Tests

Real time and simulation images are taken from up to 10 meters in this study.

IV. Conclusion

It is clear that symbols can be used as identification tags to recognize objects, vehicles, people, animals etc. With the help of deep learning method YOLO, 14 different symbols are used and 1400 synthetic images are labeled. They are trained and tested. Around 85% accuracy is gained with real time performance.

V. Future Work

Maximum reading distance is the most important aim of this study. To achieve this goal, more studies and tests will be focused on distance. Perhaps, some symbols will be excluded and some new symbols will be added.

Real time images will be added into the dataset and accuracy will be improved. Therefore, recognition distance will be improved as well as accuracy.

Also, new symbols will be tested to implement two element combinations which are easier to handle or to increase the number of labeling.

A generic tool will be implemented for companies in the aim of determining label size and maximum reading distance.

References

[1] ANDREEV, P., Aprahamian, B., & Marinov, M. (2019, June). QR code's maximum scanning distance investigation. In *2019 16th Conference on Electrical Machines, Drives and Power Systems (ELMA)* (pp. 1-4). IEEE.

[2] *Garg, G., "QR Code Scanning Distance: An In-Depth Guide", https://scanova.io/blog/qr-code-scanning-distance, (accessed June. 20, 2023)*

[3] Jose, A., Thodupunoori, H., & Nair, B. B. (2019). A novel traffic sign recognition system combining viola–jones framework and deep learning. In *Soft Computing and Signal Processing: Proceedings of ICSCSP 2018, Volume 1* (pp. 507-517). Springer Singapore.

[4] Tai, S. K., Dewi, C., Chen, R. C., Liu, Y. T., Jiang, X., & Yu, H. (2020). Deep learning for traffic sign recognition based on spatial pyramid pooling with scale analysis. *Applied Sciences*, *10*(19), 6997.

[5] Salti, S., Petrelli, A., Tombari, F., Fioraio, N., & Di Stefano, L. (2015). Traffic sign detection via interest region extraction. *Pattern Recognition*, *48*(4), 1039-1049.

[6] Keser, T., Kramar, G., & Nožica, D. (2016, October). Traffic signs shape recognition based on contour descriptor analysis. In *2016 International Conference on Smart Systems and Technologies (SST)* (pp. 199-204). IEEE.

[7] Cao, N. (2020). Research and development of icon recognition system based on machine vision. In *IOP Conference Series: Materials Science and Engineering* (Vol. 740, No. 1, p. 012183). IOP Publishing.

[8] Wibirama, S., & Nugroho, H. A. (2017, July). Long distance Automatic Number Plate Recognition under perspective distortion using zonal density and Support

Vector Machine. In *2017 3rd International Conference on Science and Technology-Computer (ICST)* (pp. 159-164). IEEE.

[9] "Hard Hat Workers Dataset", Available: https://public.roboflow.com/object-detection/hard-hat-workers"

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