CHAPTER VII

OBJECT DETECTION WITH ARTIFICIAL INTELLIGENCE: YOLO APPLICATION

Volkan KAYA[1] & İsmail AKGÜL[2]

[1](Asst. Prof. Dr.), Erzincan Binali Yıldırım University,
Faculty of Engineering and Architecture,
Departmant of Computer Engineering, Erzincan, Turkey
E-mail: vkaya@erzincan.edu.tr
ORCID: 0000-0001-6940-3260

[2](Asst. Prof. Dr.), Erzincan Binali Yıldırım University,
Faculty of Engineering and Architecture,
Departmant of Computer Engineering, Erzincan, Turkey
E-mail: iakgul@erzincan.edu.tr
ORCID: 0000-0003-2689-8675

1. Introduction

In parallel with the developments in the field of technology, the increase in the available information in many fields increases the computational complexity of a particular task and also causes an increase in the human workforce (Došilović et al., 2018). In recent years, the development of intelligent software that can think and make decisions like humans have been an important research topic in many areas. The importance of artificial intelligence technology, which is included in these smart software, is increasing day by day (Li et al., 2018; Samek & Müller, 2019; Buiten, 2019; Liu et al., 2021). Artificial intelligence is a computational science that makes programming based on human intelligence to achieve a certain goal (McCarthy, 2007). The concept of artificial intelligence was first introduced in 1956 by scientists who predicted that machines could think like humans, and then studies were carried out for the development of artificial intelligence (Zhang & Lu, 2021). Artificial intelligence is a set of systems that

adopt human intelligence to perform specific tasks and perform tasks related to mental processes (Wang et al., 2020). Therefore, artificial intelligence (AI) is smart software that allows any workload that a human can do to be done faster and helps people's decision-making processes in many areas.

Today, with the development of technology, the increase in the amount of data and the growth in computational power accordingly increase the importance of artificial intelligence. Data is the most important element for artificial intelligence to provide more efficient and successful results than humans (Haefner et al., 2021). The amount of data is important for the training and testing of artificial intelligence. Therefore, the large amount of data leads to the successful implementation of artificial intelligence solutions. For this reason, artificial intelligence-supported applications reduce human labor costs and have become a necessity in many intelligent application areas such as text processing, voice processing, image processing, data processing, big data analytics, cyber security, and autonomous driving systems (Sekaran et al., 2020; Bardou et al., 2018; Zhang & Dahu, 2019; Zhang et al., 2018; Bag et al., 2021; Zeadally et al., 2020; Namatherdhala et al., 2022; Zhang & Lu, 2021).

In recent years, different architectures based on artificial intelligence have been developed and studies have been carried out in many different areas by using these architectures. You only look once (Yolo) algorithm, which is included in these architectures and has shown great success in many areas in recent years, is preferred in many application areas by offering a new approach for object detection. Yolo is a regression-based algorithm that has shown outstanding success in the fields of object detection based on a convolutional neural network (CNN). Thanks to this algorithm, any image is passed through the neural network at once and the coordinates and classes of all objects in the image are determined. This algorithm determines the location of objects by dividing an input image given to the network into grids with a certain S×S size and also includes the region where the objects are located in a bounding box. The bounding box is determined depending on the threshold value of the grids (Redmon et al., 2016).

However, the most important difference that distinguishes the Yolo algorithm from other architectures is that it has superior success in real-time applications as well. For this reason, in object detection applications, both faster and more accurate results are obtained compared to other architectures. Figure 1 shows the basic working principle of the network structure of the

Yolo algorithm, which detects the coordinates and classes of the objects in an image.

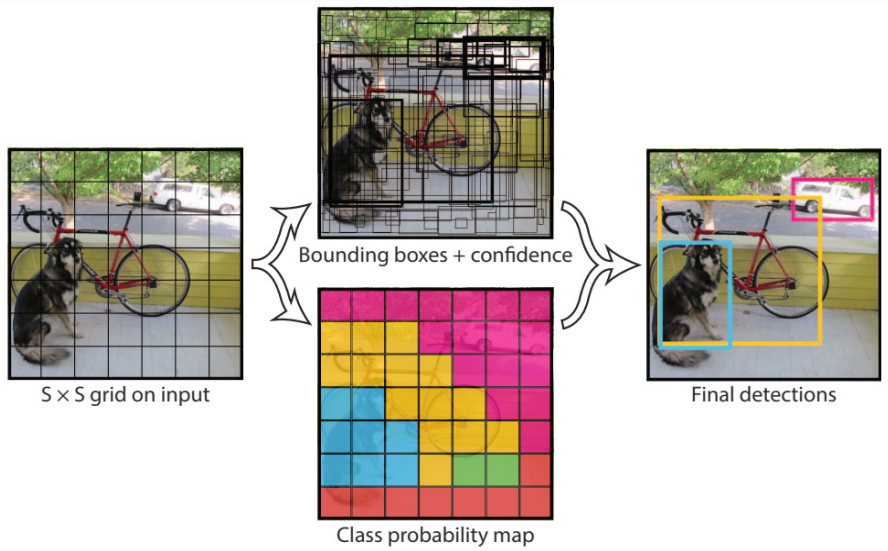


Figure 1: The working principle of the network structure of the Yolo algorithm (Redmon et al., 2016)

In this network structure, output predictions are produced by passing an image given as input through a convolutional neural network that determines the bounding box and the class probability map. Each bounding box consists of a grid cell with x and y coordinates, width (w) and height (h) values of the entire image, and the confidence estimate value by Intersection over Union (IoU) between the predicted box and any baseline truth box. Therefore, the Yolo algorithm detects the objects in the image together with their coordinates at the same time (Redmon et al., 2016; Sang et al., 2018).

Therefore, in this study, Yolo-v2 (Redmon & Farhadi, 2017), Yolo-v3 (Redmon & Farhadi, 2018), and Yolo-v4 (Bochkovskiy, et al., 2020) algorithms, which are different versions of the Yolo algorithm, which have shown successful results in object detection applications in the field of artificial intelligence in recent years, were used. With these algorithms, comparative experimental analyzes were performed using a dataset (Kaggle, 2022) containing traffic sign images. In the study, the materials and methods used in Chapter 2, the experimental results in Chapter 3, and the results and recommendations in Chapter 4 are explained in detail.

2. Material and Methods

In this section, the system structures and datasets used for the training and testing of artificial intelligence-based Yolo algorithms, which are preferred to detect traffic signs on the highway, are mentioned.

2.1. System Configuration

In this study, the training and testing of Yolo algorithms were carried out using Python programming language in a Google Colaboratory (Colab, 2022) environment with an NVIDIA Tesla K80 graphics processor together with a computer with an Intel Core i7-9750H 2.60 GHz processor, 16 GB RAM, 512 GB SSD, and Nvidia GeForce GTX 1650 graphics card.

2.2. Traffic Signs Dataset and Preprocessing

In this study, a public access dataset containing traffic signs on the highway was used to test the success of Yolo algorithms developed for application areas based on artificial intelligence (Kaggle, 2022). In this dataset, there are 4706 images in total, including trespass sign, no entry, straight or left only, straight or right only, left only ahead, 20-speed limit end, 30-speed limit, 20-speed limit, right only ahead, no right turn, no left turn, stop, no parking, park, bus stop, red light, yellow light, and green light images belonging to 18 different classes, and label files of 4674 of these images. It has been observed that the number of images and the number of label files are not equal. In addition, the number of labels of the classes in the dataset was examined and it was determined that the number of labels belonging to the 20-speed limit end (0 labels) and 20-speed limit (7 labels) classes was not sufficient. For these reasons, images of the 20-speed limit end and 20-speed limit classes with insufficient data, as well as the images that do not have a label file and have no data in the label file, were removed from the dataset by preprocessing with a developed algorithm. As a result of these pre-processes, a total of 47 images were extracted from the dataset, and a dataset consisting of 3508 training images and 1151 test images, and a total of 4659 images with 16 classes was obtained. The number of labels belonging to the classes included in this dataset is given in Table 1 in detail. In addition, the images of the traffic signs in the dataset are shown in Figure 2.

Table 1: Number of labels of traffic signs in the dataset

	Number of labels in the dataset		
Dataset classes	**Train**	**Test**	**Total**
trespass sign	532	171	703
no entry	114	46	160
straight or left only	84	55	139
straight or right only	191	62	253
left only ahead	171	73	244
30-speed limit	51	34	85
right only ahead	380	121	501
no right turn	384	102	486
no left turn	895	272	1167
stop	53	26	79
no parking	939	232	1171
park	896	237	1133
bus stop	363	126	489
red light	1286	277	1563
yellow light	183	38	221
green light	1070	266	1336
Total	**7592**	**2138**	**9730**

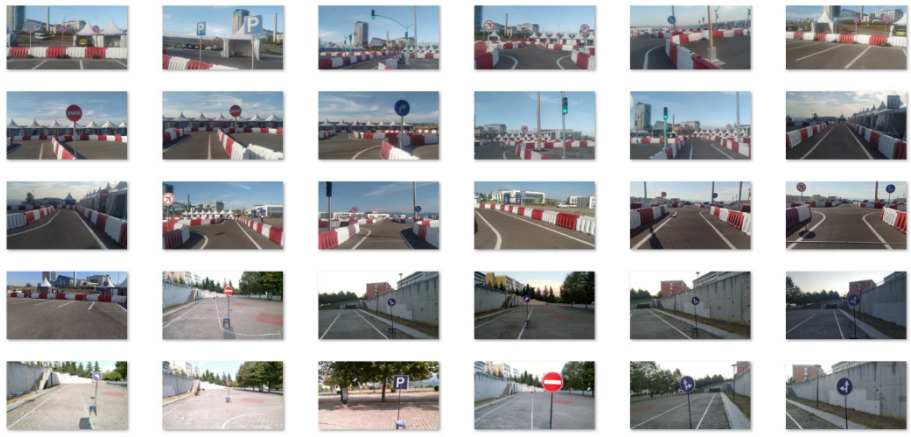


Figure 2: Images of traffic signs used in training Yolo algorithms

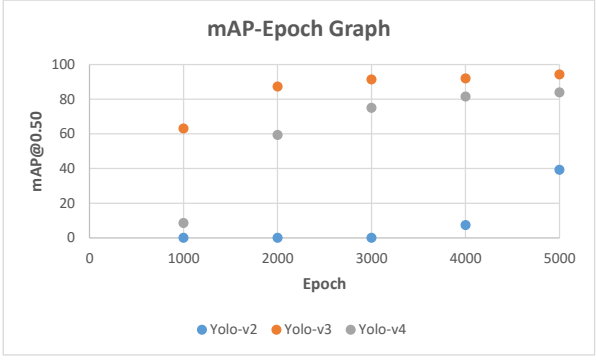
2.3. Training and Testing of Yolo Algorithms

With the Yolo-v2, Yolo-v3, and Yolo-v4 algorithms discussed in the study, a total of 4659 images in the traffic sign dataset were used to support the driver/

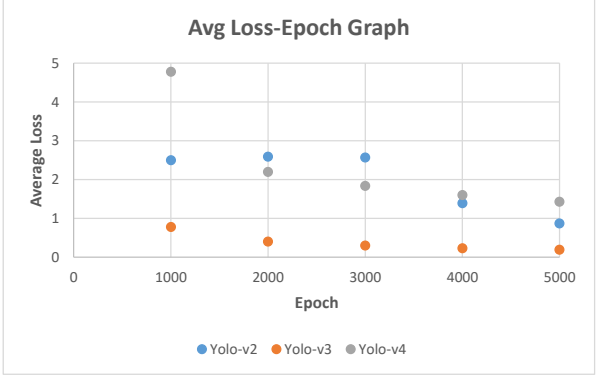
driverless systems used on the highway. 3508 images in the dataset were used for training Yolo algorithms and 1151 images were used for testing Yolo algorithms. Each of the Yolo algorithms has been trained and tested in 1000, 2000, 3000, 4000, and 5000 epochs.

3. Results

In the study, Yolo-v2, Yolo-v3, and Yolo-v4 algorithms were trained and tested using a dataset containing traffic signs. Algorithms were run for a total of 5000 epochs, with results for every 1000 epochs. Therefore, many experimental analyzes have been made for the training and testing of each algorithm, and the success results of the algorithms have been obtained. Comparative success graphs obtained as a result of training and testing Yolo algorithms are given in Figure 3.



(a)



(b)

Figure 3: Comparative success graphs obtained as a result of training and testing Yolo algorithms (a) Mean Average Precision (mAP) (b) Average Loss

When Figure 3a is examined, it is seen that the best learning occurs with the Yolo-v3 algorithm, and the worst learning occurs with the Yolo-v2 algorithm after every 1000 epochs. When Figure 3b is examined, it is seen that as a result of 5000 epochs, the Yolo-v3 algorithm has reached the lowest average loss value, while the Yolo-v4 algorithm has reached the highest average loss value. In addition, all numerical results obtained as a result of training and testing Yolo algorithms are given in Table 2 comparatively.

Table 2: Numerical results obtained as a result of training and testing Yolo algorithms

Yolo Algorithms	mAP	Avg Loss	Precision	Recall	f1-score	Support
Yolo-v2	0.39	0.87	0.53	0.46	0.49	2138
Yolo-v3	0.94	0.19	0.91	0.94	0.93	2138
Yolo-v4	0.84	1.43	0.75	0.87	0.81	2138

In addition, the confusion matrix results obtained as a result of training and testing the Yolo-v2, Yolo-v3, and Yolo-v4 algorithms are given in Table 3, Table 4, and Table 5, respectively.

Table 3: Confusion matrix results obtained as a result of training and testing the Yolo-v2 algorithm

Dataset classes	True Positive	False Positive	AP
trespass sign	98	87	57.54%
no entry	2	14	8.35%
straight or left only	7	6	31.25%
straight or right only	20	24	38.55%
left only ahead	34	45	49.51%
30-speed limit	1	0	59.35%
right only ahead	82	43	70.62%
no right turn	56	88	42.05%
no left turn	188	61	68.68%
stop	0	0	11.33%
no parking	173	90	68.47%
park	177	109	75.09%
bus stop	37	49	27.61%
red light	104	248	17.66%
yellow light	2	14	2.82%
green light	0	0	0.00%
	mAP@0.50		**39.30%**

Table 4: Confusion matrix results obtained as a result of training and testing the Yolo-v3 algorithm

Dataset classes	True Positive	False Positive	AP
trespass sign	161	12	95.68%
no entry	44	1	95.65%
straight or left only	51	7	97.18%
straight or right only	61	7	98.15%
left only ahead	70	12	94.26%
30-speed limit	33	0	98.62%
right only ahead	121	13	99.18%
no right turn	94	25	92.10%
no left turn	254	25	95.90%
stop	20	0	76.92%
no parking	224	21	95.17%
park	236	16	99.40%
bus stop	124	4	98.85%
red light	262	26	94.29%
yellow light	33	3	90.90%
green light	231	34	86.99%
	mAP@0.50		**94.33%**

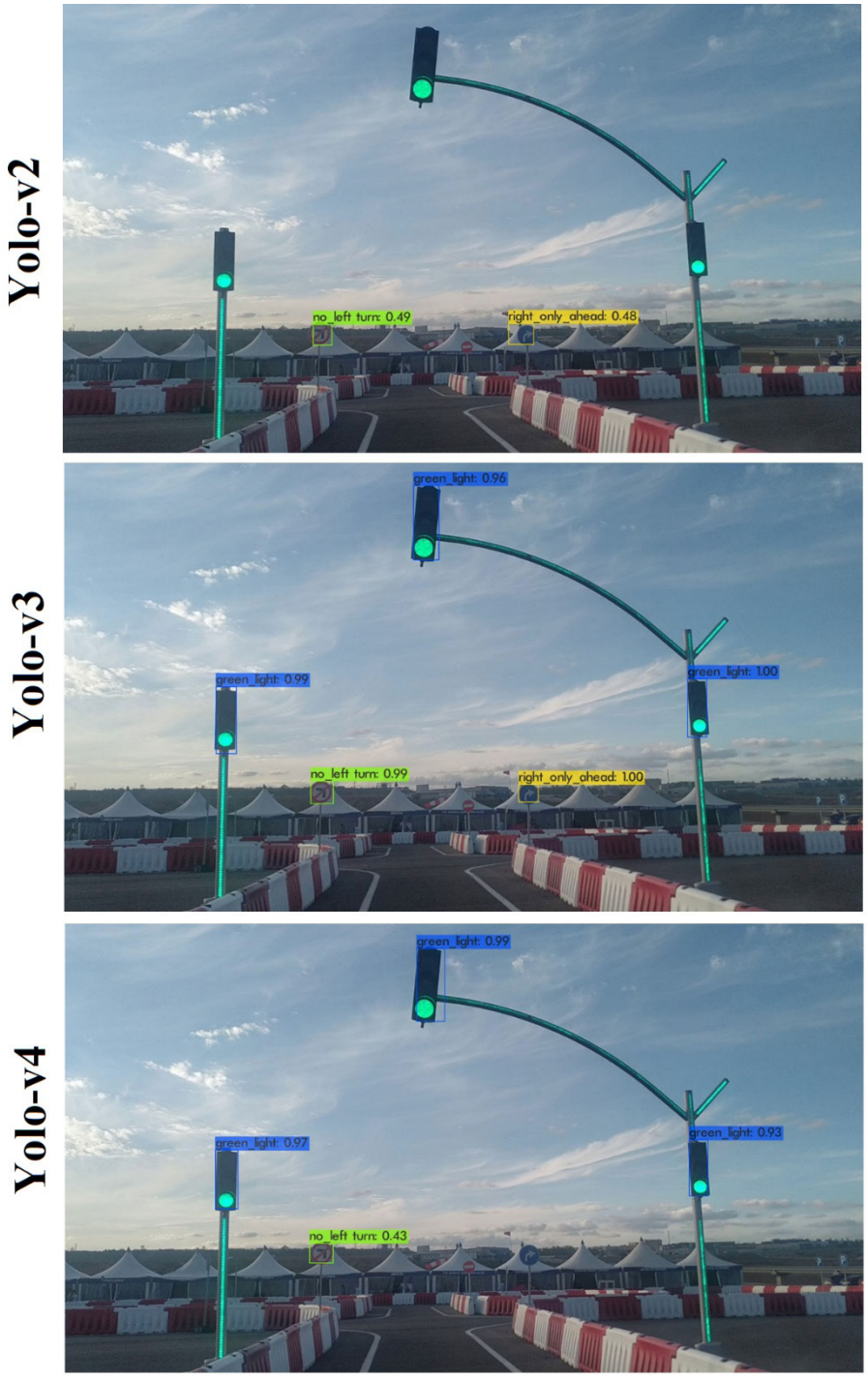
Table 5: Confusion matrix results obtained as a result of training and testing the Yolo-v4 algorithm

Dataset classes	True Positive	False Positive	AP
trespass sign	138	30	87.20%
no entry	40	18	87.51%
straight or left only	39	26	75.69%
straight or right only	56	42	85.47%
left only ahead	56	56	81.23%
30-speed limit	27	22	79.85%
right only ahead	114	76	88.77%
no right turn	75	79	72.72%
no left turn	253	92	90.79%
stop	16	4	64.14%
no parking	221	10	94.91%
park	233	28	99.31%
bus stop	103	12	91.28%
red light	242	43	89.48%
yellow light	26	10	72.49%
green light	222	59	82.27%
	mAP@0.50		**83.94%**

According to the results of the confusion matrix obtained as a result of the experimental analyzes for the training and testing of Yolo algorithms in Table 3, Table 4, and Table 5, it has been seen that the algorithm that most successfully detects traffic signs is Yolo-v3 at a rate of 94.33%. Then, it is seen that it is realized with Yolo-v4 at 83.94% and Yolo-v2 at 39.30%, respectively.

3.1. Performance testing of Yolo algorithms

To determine the validity of the success rates obtained as a result of experimental analyzes with Yolo-v2, Yolo-v3, and Yolo-v4 algorithms, a performance test was performed for each algorithm separately and each algorithm was tested using real-life images. Examples of the success rates of each algorithm on real-life performance are shown in detail in Figure 4.


Yolo-v2
no_left turn: 0.49
right_only_ahead: 0.48
Yolo-v3
green_light: 0.96
green_light: 0.99
green_light: 1.00
no_left turn: 0.99
right_only_ahead: 1.00
Yolo-v4
green_light: 0.99
green_light: 0.97
green_light: 0.93
no_left turn: 0.43

Yolo-v2



Yolo-v3



Yolo-v4



Yolo-v2



Yolo-v3



Yolo-v4




Yolo-v2
no_left_turn: 0.66
no_right_turn: 0.77
Yolo-v3
no_left_turn: 0.99
no_right_turn: 0.98
Yolo-v4
no_left_turn: 0.98
no_right_turn: 0.96

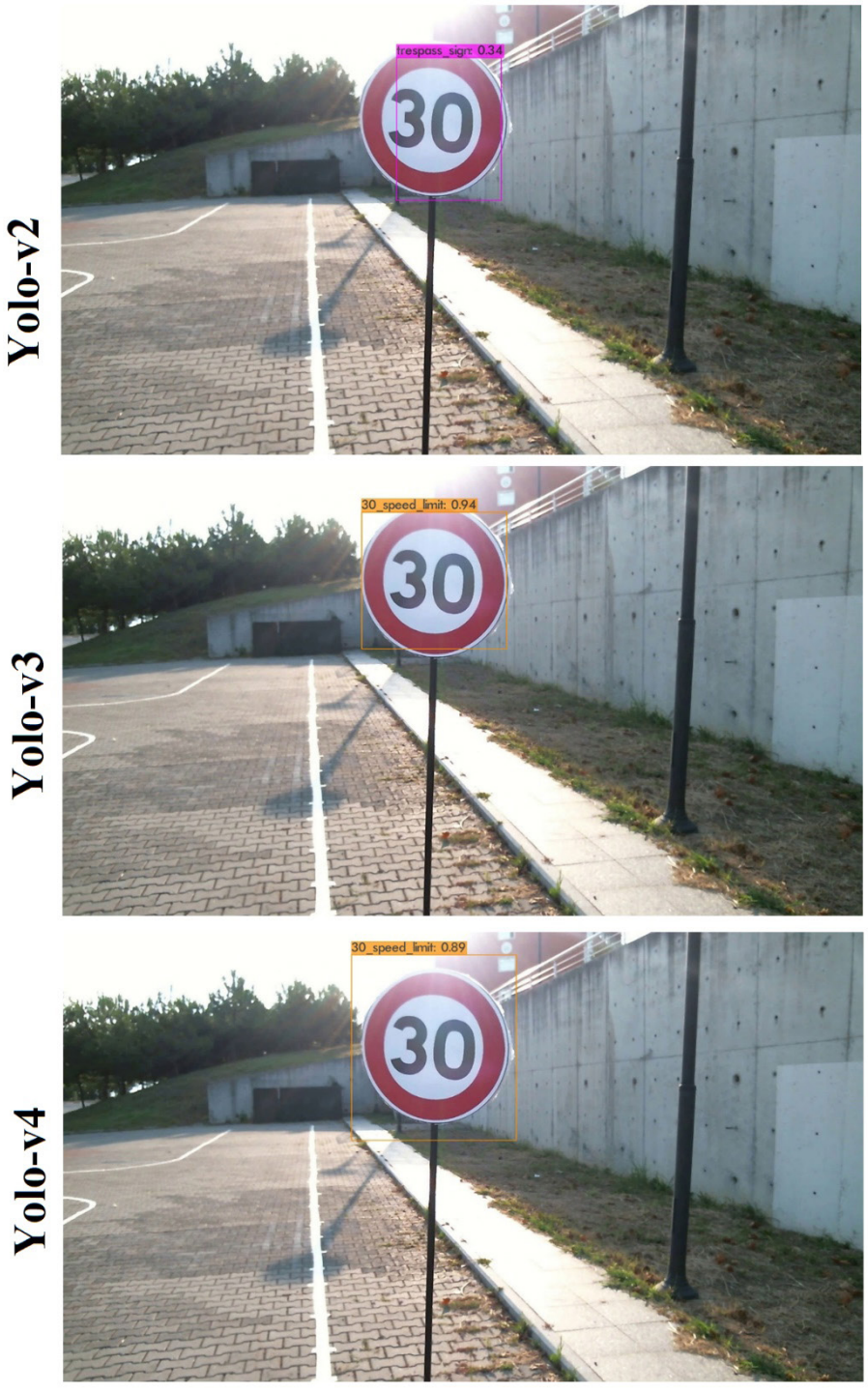

Yolo-v2
trespass_sign: 0.34
30
Yolo-v3
30_speed_limit: 0.94
30
Yolo-v4
30_speed_limit: 0.89
30



Figure 4: Images of performance testing of Yolo-v2, Yolo-v3, and Yolo-v4 algorithms

4. Conclusions

The increase in the amount of data in many areas due to the developments and innovations in the field of technology has increased the complexity of calculating a particular task and the cost of human workload. Therefore, the development of artificial intelligence software that can think and make decisions like humans has helped people's decision-making process in many areas and has achieved more efficient and successful results than humans.

In this study, Yolo algorithms (Yolo-v2, Yolo-v3, and Yolo-v4), which recognize and detect traffic signs on the highway, were used to support the driver/driverless systems used on the highway. For the training and testing of Yolo algorithms, a dataset containing a total of 4659 traffic sign images with 16 classes downloaded from the internet was used. The success rates of Yolo algorithms were compared using this dataset. As a result of the comparison, it was concluded that the algorithm that successfully recognizes and detects traffic signs with a rate of 94.33% was the Yolo-v3 algorithm, and then it was determined by the Yolo-v4 algorithm with a rate of 83.94% and by the Yolo-v2 algorithm with a rate of 39.30%, respectively. In addition to this, a performance test has been carried out in order to recognize and detect traffic signs on the highway in real life and the success of Yolo algorithms has been tested.

As a result, based on experimental analysis, it has been seen that Yolo algorithms perform successfully in artificial intelligence-based applications. Therefore, the fact that Yolo algorithms can be applied to different artificial intelligence-based systems increases the importance of Yolo algorithms.

References

Bag, S., Pretorius, J. H. C., Gupta, S., & Dwivedi, Y. K. (2021). Role of institutional pressures and resources in the adoption of big data analytics powered artificial intelligence, sustainable manufacturing practices and circular economy capabilities. Technological Forecasting and Social Change, 163, 120420. https://doi.org/10.1016/j.techfore.2020.120420

Bardou, D., Zhang, K., & Ahmad, S. M. (2018). Lung sounds classification using convolutional neural networks. Artificial intelligence in medicine, 88, 58-69. https://doi.org/10.1016/j.artmed.2018.04.008

Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). Yolov4: Optimal speed and accuracy of object detection. arXiv preprint. https://doi.org/10.48550/arXiv.2004.10934

Buiten, M. C. (2019). Towards intelligent regulation of artificial intelligence. European Journal of Risk Regulation, 10(1), 41-59. https://doi.org/10.1017/err.2019.8

Colab (2022). "Google Colaboratory", Retrieved in October, 26, 2022 from https://colab.research.google.com

Došilović, F. K., Brčić, M., & Hlupić, N. (2018). Explainable artificial intelligence: A survey. In 2018 41st International convention on information and communication technology, electronics and microelectronics (MIPRO), IEEE, Opatija, Croatia, 0210-0215. https://doi.org/10.23919/MIPRO.2018.8400040

Haefner, N., Wincent, J., Parida, V., & Gassmann, O. (2021). Artificial intelligence and innovation management: A review, framework, and research agenda☆. Technological Forecasting and Social Change, 162, 120392. https://doi.org/10.1016/j.techfore.2020.120392

Kaggle, (2022). "Kaggle", Retrieved in October, 25, 2022 from https://www.kaggle.com/datasets/mehmetokuyar/traffic-sign-dataset

Li, L., Lin, Y. L., Zheng, N. N., Wang, F. Y., Liu, Y., Cao, D., Wang, K., & Huang, W. L. (2018). Artificial intelligence test: A case study of intelligent vehicles. Artificial Intelligence Review, 50(3), 441-465.

Liu, N., Shapira, P., & Yue, X. (2021). Tracking developments in artificial intelligence research: Constructing and applying a new search strategy. Scientometrics, 126(4), 3153-3192. https://doi.org/10.1007/s11192-021-03868-4

McCarthy, J. (2004). What is artificial intelligence. URL: http://www-formal.stanford.edu/jmc/whatisai.pdf

Namatherdhala, B., Mazher, N., & Sriram, G. K. (2022). Uses of Artificial Intelligence in Autonomous Driving and V2X communication. International Research Journal of Modernization in Engineering Technology and Science, 4(7). 1932-1936.

Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. arXiv preprint. https://doi.org/10.48550/arXiv.1804.02767

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016, 779–788.

Redmon, J., Farhadi, A. YOLO9000: Better, Faster, Stronger. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017, 6517–6525.

Samek, W., & Müller, K. R. (2019). Towards explainable artificial intelligence. In Explainable AI: interpreting, explaining and visualizing deep learning, Springer, Cham, 5-22. https://doi.org/10.1007/978-3-030-28954-6_1

Sang, J., Wu, Z., Guo, P., Hu, H., Xiang, H., Zhang, Q., & Cai, B. (2018). An improved YOLOv2 for vehicle detection. Sensors, 18(12), 4272. https://doi.org/10.3390/s18124272

Sekaran, K., Chandana, P., Jeny, J. R. V., Meqdad, M. N., & Kadry, S. (2020). Design of optimal search engine using text summarization through artificial intelligence techniques. TELKOMNIKA (Telecommunication Computing Electronics and Control), 18(3), 1268-1274. http://doi.org/10.12928/telkomnika.v18i3.14028

Wang, C. X., Di Renzo, M., Stanczak, S., Wang, S., & Larsson, E. G. (2020). Artificial intelligence enabled wireless networking for 5G and beyond: Recent advances and future challenges. IEEE Wireless Communications, 27(1), 16-23. https://doi.org/10.1109/MWC.001.1900292

Zeadally, S., Adi, E., Baig, Z., & Khan, I. A. (2020). Harnessing artificial intelligence capabilities to improve cybersecurity. IEEE Access, 8, 23817-23837. https://doi.org/10.1109/ACCESS.2020.2968045

Zhang, C., & Lu, Y. (2021). Study on artificial intelligence: The state of the art and future prospects. Journal of Industrial Information Integration, 23, 100224. https://doi.org/10.1016/j.jii.2021.100224

Zhang, R., Zheng, F., & Min, W. (2018). Sequential behavioral data processing using deep learning and the Markov transition field in online fraud detection. arXiv preprint. https://doi.org/10.48550/arXiv.1808.05329

Zhang, X., & Dahu, W. (2019). Application of artificial intelligence algorithms in image processing. Journal of Visual Communication and Image Representation, 61, 42-49. https://doi.org/10.1016/j.jvcir.2019.03.004