SCIENTIFIC PUBLICATIONS OF THE STATE UNIVERSITY OF NOVI PAZAR
SER. A: APPL. MATH. INFORM. AND MECH. vol. 13, 2 (2021), 105–115.

Recognition and Detection with Deep Learning Methods

Ömer Can Eskicioğlu, Edin Dolićanin, Ali Hakan Işık, Rifai Kuçi

Abstract: The method of recognizing traffic signs through image processing has increased in popularity along with advanced driver assistance systems. Drivers may have difficulty reading and detecting traffic signs due to fatigue, weather conditions and speed while driving. In our study, traffic signs rectangular, square, circle and so on. Regardless of the type of different plates seen in the country, even if the correct detection is aimed. By sending the model as a parameter while training, the only thing that needs to be done within the scope of adding a new plate is to retrain our model. Before starting learning, the image was enhanced to improve the performance of the algorithm by using the Contrast Restricted Adaptive Histogram Equation (CLAHE) method in data processing. In our study, results were obtained with 2 deep learning models unlike classical CNN architecture. VGG-16 and Xception deep learning models were compared with each other. SGD and Adam optimization methods were tried for both models and the optimum method was found for our study. Our study has reached an accuracy value of up to 98.38%. The speed performance of our method is sufficient to enable a real-time system implementation in the future. In order to understand the results of our experimental tests in the system to be used, it has been turned into a return parameter and the driver can be integrated with the vehicle regardless of the screen and used with voice assistant or small structures to be added independently of the vehicle.

Keywords: image preprocessing, VGG-16, Xception, CLAHE method, traffic sign detection

1 Introduction

The main purpose of traffic signs is to prevent accidents by ensuring the safety of traffic and to ensure that traffic is handled by drivers in the most correct way. Traffic signs are designed to be easily noticed and understood by drivers in every environment. Since the psychological and physical conditions of people will change most of the time, someone who ispsychologically overwhelmed may not read traffic signs and signs correctly. One of the most important problems of cities is transportation problem, and there are heavy traffic

Manuscript received October 30, 2021. ; accepted November 29,2021.

Ömer Can Eskicioğlu is with Burdur Mehmet Akif Ersoy University, Engineering and Architecture Faculty, Burdur, Turkey; Edin Dolićanin is with State University of Novi Pazar, Novi Pazar, Serbia; Ali Hakan Işık is with Burdur Mehmet Akif Ersoy University, Engineering and Architecture Faculty, Computer Engineering Department, Burdur, Turkey; Rifai Kuçi is with Burdur Mehmet Akif Ersoy University, Engineering and Architecture Faculty, Burdur, Turkey.

congestion especially in the central points of crowded cities. [10] Therefore, it is predicted that this will cause many accidents. In such cases, the system that informs the driver about the traffic signs while driving is predicted to greatly reduce the accidents that will occur [13].

Fig. 1: Examples of Traffic Signs Smooth and Noisy

The rapidly increasing population has also increased the number of vehicles. With the increase in the number of vehicles, there has been a significant increase in the number of roads and traffic signs in order to ensure traffic safety. Along with advancing information and technology, the automotive industry made many progress. As one of these developments, Advanced Driver Assistance Systems come to the fore today. Nowadays, the use of large-sized touch indicators and screens, which are generally seen in vehicles in the upper segment, has increased in vehicles. The increasing trend of electric vehicles and autonomous driving experience leads to very different driving area designs [8]. Blind spot warning systems, identification methods of traffic signs and signs can be given as examples of these systems.

Considering the researches, the majority of traffic accidents are caused by driver-induced errors. The vast majority of these errors are caused by not obeying the legal rules such as traffic rules and speed.

Among the studies on the recognition of traffic sign signs in the studies in the literature are the following;Identifying traffic signs in real time using color and shape information by Hüseyin Yalçın and Hasan Irmak [20], tried to classify traffic signs with support vector machines by Emrah Onat and Ömer Özdil [13]. In the hierarchical convolutional neural networks and traffic sign identification studies carried out by [6] Emin Alper Sürücü and Hatice Doğan, the traffic signs can be recognized with high accuracy, but in the data set they use for traffic signs, a study specific to Germany traffic signs has been carried out. If Germany encounters a sign that is not on the traffic signs, the wrong result will be returned [17]. In the study conducted by Ersin Demir, a classification study was made only for the signs with circles [2].

In this article, it has been tried to classify traffic signs and signs with Convolutional

Neural Networks (CNN). Accuracy Score 0.98 was found to be an acceptably high value. Unlike some studies, in our study, traffic signs can be classified correctly regardless of shape such as circle, rectangular circle, square, hexagon. Our training allows them to save the models and use them in real-time. The results of the data we tested in our study are shown on random pictures we selected from the test set.

2 Materials and methods

GTSRB - German Traffic Sign Recognition Benchmark dataset consisting of approximately 52 thousand images including traffic signs was used. The dataset is reserved for 39 thousand training and 12 thousand tests. It goes through various image preprocessing steps before images are given to the CNN architecture. In the structure used, the size reduction process is done primarily in the pictures. As a result of this process, the image sizes have been reduced to 32x32. These images are filtered to gray color scale. It is aimed to increase the performance of the algorithm by using the Contrast Limited Adaptive Histogram Equalization (CLAHE) method.

2.1 Convolutional Neural Networks

Convolutional neural networks are often used to analyze visual information. It is a subbranch of deep learning. Convolutional neural networks are deep artificial neural networks that are frequently used in image, object classification and clustering using similarities of attributes. [22] ANNs determine the relationship between input and target variables of previous samples by weight assignment method. [12] CNN architecture can be used in various fields such as disease diagnosis, animal, vehicle classification, fire detection. [4] CNN consists of 7 layers.

1. Input Layer
2. Convulation Layer
3. Rectifed Linear Units Layer (ReLu)
4. Pooling Layer
5. Fully-Connected Layer
6. Dropout Layer
7. Classification Layer

The first stages in CNN architecture consist of Convolution and Pooling layers. [14] In the final stages there is the Fully Connected layer and then the Classification layer. In order to prevent the memorization of the network and to obtain alternative and high accuracy results, the screening process is performed in the Dropout Layer. [11] In CNN, it aims to find the whole image by separating the qualities of small images.

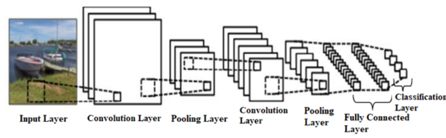

Fig. 2: General Architecture of the Convolutional Neural Network [24]

Convolutional Neural Network Layers

2.1.1 Input Layer

This layer is one of the first and basic layers of CNN. It is the layer that enables the desired input picture, video or audio file to be converted to the appropriate format and given to the network raw.

2.1.2 Convulation Layer

It is mentioned as the transformation layer in CNN. In this layer, the filters are based on the basic process of hovering and panning the image taken in the input layer. Filters can be of different sizes. It can be in sizes you deem appropriate for the solution of the problem. Most used filters; It is 2x2, 3x3, 5x5, 7x7. Filters are represented by matrices of specified sizes. [5]

2.1.3 Rectified Linear Units Layer(ReLu)

It comes after the convolution layer and is also known as the activation layer. Usually they do not have a linear structure. These are the functions that enable the outputs to be produced against the inputs to the units to be determined. There are several types of activation function such as Sigmoid, Tanh, and Relu. Relu is the most common activation function used for fast back propagation and feed forward in CNN. [15] This layer is used to transform the deep web we use into a nonlinear structure. [14]

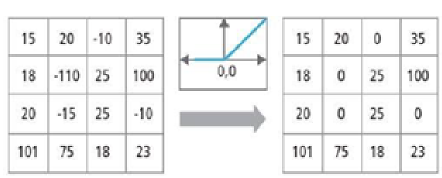

Fig. 3: Transform Operations of Relu Activation Function on Matrix [7]

2.1.4 Pooling layer

Pooling layer is basically the layer where the reduction and reduction of the size of the data we have is done. [9] Depending on the model, the width and height values of the input dimensions are reduced before entering the convolutional layer that follows it. No matter how much information is lost in the pooling layer, the missing values mean less computational load for the next layers. No feature loss occurs in the size reduction process.

2.1.5 Fully Connected Layer

This layer comes after the overlapping convolution and pooling layer due to the CNN architecture. It connects all the neurons in the previous layers to the neurons in its structure. Generally, there are 1 or 2 full connection layers in CNN networks. There are attributes and features before this layer. This layer also determines the classes. There may be differences in the number of layers according to different architectures.

2.1.6 Dropout Layer

When training is done on big data in CNN, our CNN network is memorizing. [14] This layer is used to prevent the network from memorizing. [16] The basic rationale at the layer is to remove various nodes on the network. Removed nodes differ during each update.

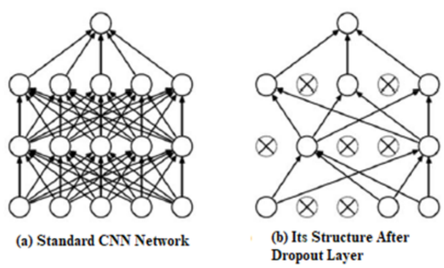

Fig. 4: Applying DropOut Layer to CNN Network [14]

2.1.7 Classification Layer

This layer is the last layer in the model. The full link comes after the layer and does the classification process. The output value in the layer is equal to the number of objects to be classified.

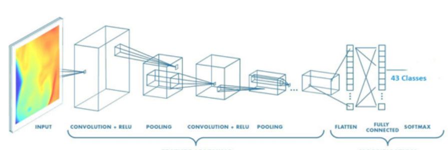

Fig. 5: Example CNN Architecture

2.2 Contrast Limited Adaptive Histogram Equalization

In the CLAHE method, the quality of the image can be improved by expanding the dynamic range with image histograms. [1] In histogram equalization, it is tried to obtain a more regular and smooth result image by normalizing the density distribution in the images. However, in histogram equalization studies, since the distribution in the density of the image is used, it may cause a faded effect if the intensity level is brought to the middle in the images. It may cause the formation of some noise pixels in images with high density in narrow regions. [1] In order to solve such problems, local histogram equalization studies are constantly being developed [21]. The general purpose image is divided into small rectangles in a rectangular shape. Standard histogram equalization is applied in the divided regions. The widths, heights and numbers of the optimum zones may vary according to the image. [1] After histogram equalization in quadrilaterals in the lower regions, an improved image is obtained by bi-linear interpolation method. [18] However, there is a noise problem in the adaptive histogram equalization method. In order to prevent this situation, it is necessary to limit the improvement in contrast equalization in regions with common input. To overcome this problem, a method called contrast-limited adaptive histogram equalization has been developed.

2.3 Visual Geometry Group (VGG16)

The vgg network proposed by the Oxford Visual Geometry Group (VGG) is a homogeneous architecture used to achieve better results in the ILSVRC-2014 competition. The VGG network is a model consisting of 13 convolution 3 fully connected layers. There are 41 layers in total, including Relu layer, Fully-connected layer, Maxpool, Softmax layer, and Dropout layer layers. VGG model is as shown in Figure 6. The size of the images in the VGG network input layer is generally 224x224x3. [3, 19] In our study, 32x32x3 and 64x64x3 were used in different ways.

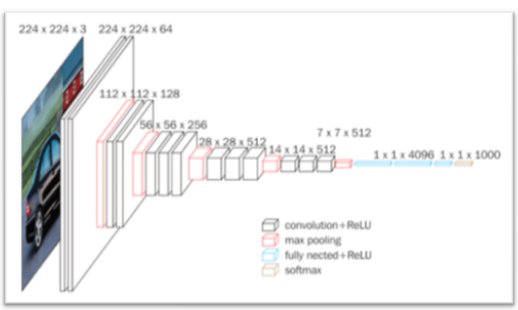

Fig. 6: Structure of Example VGG16 Architecture

2.4 Xception

Xception model was brought to the literature by Francois Chollet. Xception is an extension of the Inception architecture, replacing Inception modules with deeply separable convolutions. [23]

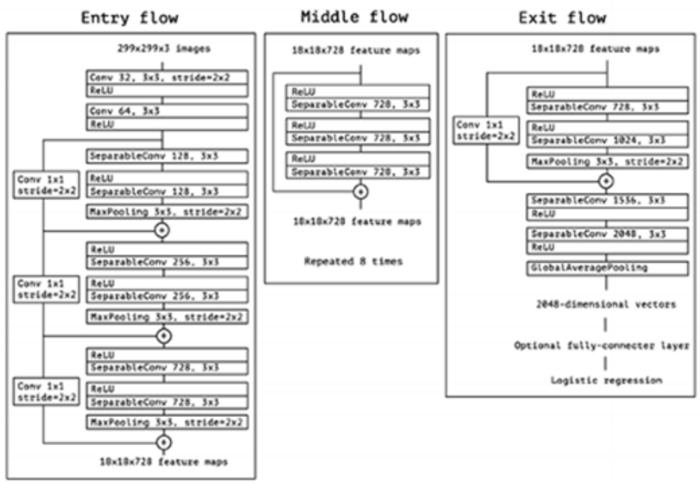

Fig. 7: Sample Xception Architecture Structure [23]

3 Results and Discussion

In this section, we will examine the success of the identification and detection of traffic signs, which is the purpose of our study, on the deep learning architectures we have created.The optimum performance was determined by giving different values to the hyper parameters in our study. Performance has been tested on different deep learning algorithms and optimization methods.. The data related to the classification metrics in the study were analyzed and evaluated. In Table 1, metric results can be seen by using 2 layer CNN architecture.The average precision is around 95% and the average recall is around 95%. They are acceptable values. Table 1 shows the accuracy, weighted and macro average values of our classes. These are the values obtained as a result of 12630 test images.

Table 1: Classification Report

	Precision	Recall	F1 - Score	Support
Accuracy			0.95	12630
Micro avg	0.92	0.93	0.92	12630
Weighted avg	0.95	0.95	0.95	12630

Figure 8 shows the loss and accuracy graph during the training phase. There is an inverse ratio between loss value and accuracy value. When the epoch number of turns is increased, it can be determined that the loss value gradually decreases and the accuracy value increases. It is the result of 30 rounds of training on approximately 39 thousand data in education.

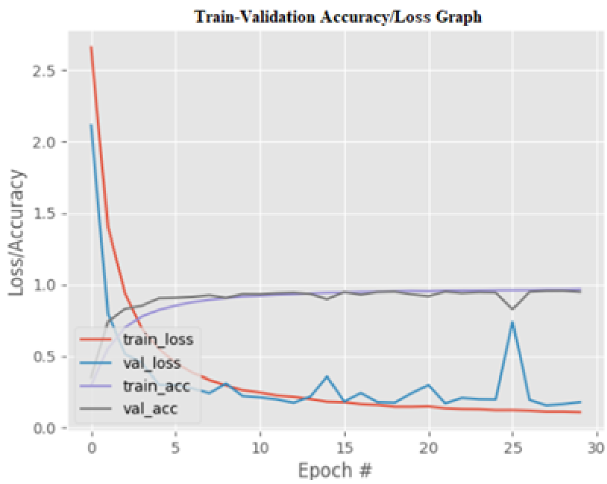

Fig. 8: Train-Validation Accuracy/Loss Graph

The study was also used in 2 different deep learning models. On these models, the results were compared using 2 different optimization methods. Looking at the results, it makes more sense to use the VGG-16 deep learning architecture for this problem. Highly accepted results were obtained with the Adam optimization method in VGG-16 deep learning model.

Table 2: Results of Architectures for 32x32 Video Images

Input : 32 x 32		Precision	Recall	F1 - Score	Total Accuracy
VGG - 16	SGD	0.10	0.14	0.10	0.1444
	Adam	0.97	0.97	0.97	0.9693
Xception	SGD	0.59	0.62	0.59	0.6247
	Adam	0.61	0.64	0.61	0.6361

Table 2 above shows different deep learning models and optimization methods for 32x32 input view. The dataset used in our study is operated with the models and optimization methods found above. As a result of the study, VGG-16 deep learning architecture and Adam optimization method take the highest values among other possible possibilities for 32 x 32 inputs. The high value received is deemed appropriate in terms of usability.

Table 3: Results of Architectures for 64x64 Video Images

Input : 64 x 64		Precision	Recall	F1 - Score	Total Accuracy
VGG - 16	SGD	0.28	0.33	0.29	0.3331
	Adam	0.98	0.98	0.98	0.9838
Xception	SGD	0.73	0.75	0.72	0.7468
	Adam	0.91	0.83	0.83	0.8312

As seen in Table 3, VGG-16 and Xception models are compared for 64x64 inputs. Models were run sequentially using SGD and Adam optimization methods. VGG-16 and Xception models achieve high rates with the Adam optimizer. For different image inputs such as 32x32 and 64x64 for identification and detection of traffic signs, using the VGG-16 deep learning model with the Adam optimized method should be preferred.

4 Conclusion

In our study, traffic signs recognition process was carried out with different deep learning models. The histogram adaptive contrast equalization technique and noise reduction methods have facilitated the perception of the CNN architecture. The accuracy value has been increased with hyper parameter optimization. The highest 98% accuracy has been achieved on our studied models. It is the result of 30 rounds of training on approximately 39 thousand data in education. It has been determined that by increasing the epoch number of turns, the accuracy value has increased continuously, but this time the train_loss and val_loss values have decreased. In order for the training to have a high success, the accuracy value should be high and the values of train_loss and val_loss should be close to 0.

Regardless of the types of traffic signs (square, hexagonal, circle, etc.), traffic sign detection was achieved with image processing and diagnosis. Our work has a structure that can be operated with different models and is open to development. As a suggestion to friends who want to improve, our model will try to determine the value of the plate as it sees the plates and it has been provided to be rotated in a way that the drivers can understand what the plate is detected as a parameter. In this way, by communicating or integrating with advanced driver assistance systems, it is a predictable result that the drivers can drive their cars safely while listening to what the signs are. The rotating parameter has been prepared as a format that can be used by developers. In this way, small models can be made and a great cost can be achieved.

References

[1] K.U.R.T. BURÇIN, V.V. NABIYEV, *Dijital Mamografi Görüntülerinin Kontrast Sınırlı Adaptif Histogram Eşitleme ile İyileştirilmesi*, Proceedings of the VII. Ulusal Tıp Bilişimi Kongresi, Gazimağusa, KKTC, 14–17, 2010.

[2] E. DEMIR, *Trafik Levhası Belirleme ve Tanımlanması*, http://doczz.biz.tr/doc/277231/i-trafik-levhası-belirleme-ve-tanımlanması-ersin, 2011.

[3] F. DOĞAN, İ. TÜRKOĞLU, *Derin Öğrenme Algoritmalarının Yaprak Sınıflandırma Başarımlarının Karşılaştırılması*, Sakarya University Journal of Computer and Information Sciences, 1(1), 10–21, 2018.

[4] A. ETYEMEZ, K.U.R.T. MUSTAFA, *Yapay Sinir Ağları Yöntemi ile Optimum Takım Seçimi*, El-Cezeri Journal of Science and Engineering, 6(2), 323–332.

[5] A. GÜLCÜ, Z. KUŞ, *Konvolüsyonel Sinir Ağlarında Hiper-Parametre Optimizasyonu Yöntemlerinin İncelenmesi*, Gazi Üniversitesi Fen Bilimleri Dergisi Part C: Tasarım ve Teknoloji, 7(2), 503–522. 2019.

[6] H. GÜNDÜZ, S. KAPLAN, S. GÜNAL, C. AKINLAR, *Circular traffic sign recognition empowered by circle detection algorithm*, In 2013 21st Signal Processing and Communications Applications Conference (SIU) (pp. 1–4). IEEE. 2013.

[7] S.L. HIJAZI, R. KUMAR, C. ROWEN, *Using Convolutional Neural Networks for Image Recognition*, Erişim Tarihi: 15.06.2019. https://www.semantic scholar.org/paper/Using-Convolutional-Neural-Netwo rks-for-Image-By-Hija- zi-Kumar/bbf7b5bdc39f9b8849c639c11f4726e3, 2015.

[8] M. KARADAĞ, U.C. VURAL, B. KARASU, *Otomotiv Sektöründe Cam*, El-Cezeri Journal of Science and Engineering, 6(2), 299–322.

[9] Y. KIM, M. TEASUP, *Human Detection and Activity Classification Based on Micro-Doppler Signatures Using Deep Convolutional Neural Networks*, IEEE Geoscience and Remote Sensing Letters, 13(1), 8–12, 2016.

[10] M.T. KORU, A.P. AKGÜNGÖR, E. KORKMAZ, *Trafik Tıkanıklığının Fiyatlandırılması ve Uygulamalarının İncelenmesi: Kızılay Ankara Örneği*, El-Cezeri Journal of Science and Engineering, 4(3), 497–508, 2017.

[11] H. LEE, R. GROSSE, R. RANGANATH, A.Y. NG, *Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations*, Proceedings of the 26th annual international conference on machine learning, ACM. 2009.

[12] A.K.S.U. NECATI, U.Ç.A.N. KENAN, *Zaman ve Konum Girdileri Kullanılarak Yapay Sinir Ağlarıyla Referans Evapotranspirasyonun Tahmin Edilmesi*, El-Cezeri Journal of Science and Engineering, 3(2), 2016.

[13] E. ONAT, Ü. ÜZDIL, *Traffic sign classification using hough transform and SVM*, In 2015 23nd Signal Processing and Communications Applications Conference (SIU), (pp. 2161–2165). IEEE. 2015.

[14] İ.N.İ.K. ÖZKAN, E. ÜLKER, *Derin Öğrenme ve Görüntü Analizinde Kullanılan Derin Öğrenme Modelleri*, Gaziosmanpaşa Bilimsel Araştırma Dergisi, 6(3), 85–104.

[15] Ş.I. SERENGIL, *A Gentle Introduction to Convolutional Neural Networks*, Erişim Tarihi: 10.06.2019. https://sefiks.com/2017/11/03/a-gentle-introduction-to-convolutional-neural-networks, 2017.

[16] N. SRIVASTAVA, G.E. HINTON, A. KRIZHEVSKY, I. SUTSKEVER, R. SALAKHUTDINOV, *Dropout: a simple way to prevent neural networks from overfitting*, Journal of machine learning research 15(1): 1929–1958, 2014.

[17] E.A. SÜRÜCÜ, H. DOĞAN, *Traffic sign recognition with hierarchical Convolutional Neural Network*, In 2018 26th Signal Processing and Communications Applications Conference (SIU) (pp. 1–4). IEEE. 2018.

[18] C.K. TEO, *Digital Enhancement of Night Vision and Thermal Images*, Thesis, Naval Postgraduate School, California, 2003.

[19] M. TÜRKOĞLU, D. HANBAY, *Apricot disease identification based on attributes obtained from deep learning algorithms*, In 2018 International Conference on Artificial Intelligence and Data Processing (IDAP) (pp. 1–4). IEEE. 2018.

[20] H. YALÇIN, H. IRMAK, M.M. BULUT, G.B. AKAR, *Real-time Traffic Sign detection and Recognition on FPGA*, In 2013 21st Signal Processing and Communications Applications Conference (SIU) (pp. 1–4). IEEE. 2013.

[21] H. YOON, Y. HAN, H. HAHN, *Image Contrast Enhancement based Sub-histogram Equalization Technique without Over-equalization Noise*, International Journal of Computer Science and Engineering 2009 : 3 (2).

[22] Convolutional Neural Network (Evrişimsel Sinir Ağları), https://medium.com/@rabiaokumus96/convolutional-neural-networks-evrişimsel-sinir-ağları-cceb887a2979.

[23] Image Recognition using Pre Trained Xception Model in 5 steps, https://medium.com/@gkadusumilli/image-recognition-using-pre-trained-xception-model-in-5-steps-96ac858f4206. 2019.

[24] WILDML. Understandıng Convolutional Neural Networks For NLP, http://www.wildml.com/2015/11/understanding-convolutional-neural-networks-for-nlp/. 2016.