**Journal of Artificial Intelligence and Data Science (JAIDA)**
**e-ISSN: 2791-8335**
**https://dergipark.org.tr/pub/jaida**

**Vol. 2, No. 2, pp. 63-68, 2022**
**Research Article**

# Accuracy Comparison of CNN Networks on GTSRB Dataset

Gökberk AY [1,*], Akif DURDU [2], Barış Samim NESİMİOĞLU [1]

[1] KTO Karatay University, Faculty of Engineering and Natural Sciences, Mechatronics Engineering, Karatay/KONYA, Turkey.

[2] Konya Technical University, Faculty of Engineering and Natural Sciences, Electrical and Electronics Engineering, Selçuklu/KONYA, Turkey.

## Abstract

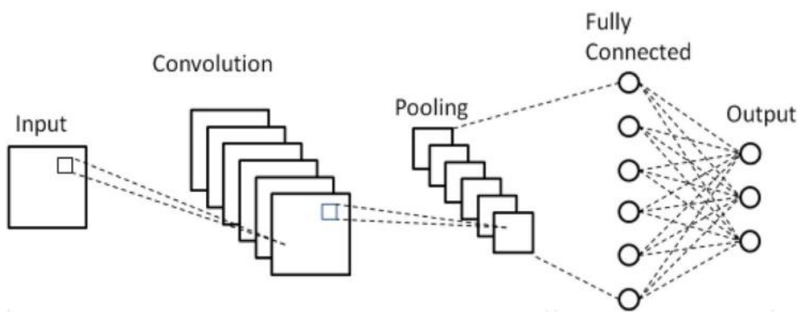
In this era, interpreting and processing the data of traffic signs has crucial importance for improving autonomous car technology. In this respect, the relationship between the recognition of traffic signs and industrial applications is highly relevant. Although real-world systems have reached that related market and several academic studies on this topic have been published, regular objective comparisons of different algorithmic approaches are missing due to the lack of freely available benchmark datasets. From this point of view, we compare the AlexNET, DarkNET-53, and EfficientNET-b0 convolutional neural network (CNN) algorithms according to validation performance on the German Traffic Signs Recognition Benchmark (GTSRB) dataset. Considering the equal training and test conditions 70% of data as training, 15% of data as training validation, and 15% of data were chosen as test data. Experimental results show us that EfficientNET-b0 architecture has 98.64%, AlexNET architecture has 97.45% and DarkNet-53 architecture has 94.69% accuracy performance.

***Keywords: AlexNET; CNN; DarkNET-53; EfficientNET-b0; GTSRB; Traffic Sign Classification.***

## 1. Introduction

The Convolutional Neural Network (CNN) is a kind of deep learning way that heap various convolutional layers in sequence. Nowadays, CNN has become a revolutionary technique in the field of artificial intelligence for classifying objects. Owing to this point of view, the classification of traffic signs has crucial importance in both constructing healthy traffic for humans and avoiding accidents. Although the classification of objects is easy-task for humans in common-used items used in daily routine, the classification of traffic signs is not so under complex scenarios [1]. The generalized basic block diagram of the CNN is given in figure 1 below.

***Figure 1.*** *Generalized block diagram of CNN [2].*




The main causes of these signs being misunderstood are the quality of sign material, blurry images, background color intersection, and weak light or weather conditions. An ideal traffic sign recognition system that has vital importance for automated drive systems comprises detection and classification implementations. The real-time performance of classification has a crucial effect on decisions instantly. The German Traffic Signs Recognition Benchmark (GTSRB) [3] dataset is created for challenging this performance comparison. Dataset has 43 classes with over 50000 real-taken pictures. The random example representatives of 43 classes in GTSRB dataset is given in figure 2 below.

*Corresponding author e-mail address: gokberk.ay@karatay.edu.tr

Received: 17/07/2022; Accepted: 22/11/2022

***Figure 2.*** *Random example representatives of 43 classes in GTSRB Dataset* [4].

Based on the serial development of computer vision systems, derived new methods enable accurately classifying complex objects. Classification of traffic signs seems basic because of their shapes or colors but it has a complex challenge concerning environmental items [5]. The main problems in that matter like bad conditions, low image quality, image blurring, and the same image color between the environment and traffic signs, etc.

Nowadays, machine learning [6] contains a hot topic called deep learning which is used for classification [7], [8], detection [9], [10], and recognition [11], [12] with achieved beyond the success of traditional methods like support vector machine (SVM) [13] or extreme learning machine (ELM) [14], [15]. When the traffic sign data has some trouble qualifications like shaded or occluded image, the accuracy rate may be decreased. Especially, the big training data enhance the generalizability of deep learning while the large model structure and its nonlinearity. According to the official results of the publication of the GTSRB dataset, various top five methods and research results are shown in Table 1.

**Table 1.** *Top five methods and performances of GTSRB dataset.*

| No | Method | Performance |
|---|---|---|
| 1 | CNN with 3 Spatial Transformers [16] | 99.71 % |
| 2 | Committee of CNNs [17] | 99.46 % |
| 3 | Color-blob-based COSFIRE filters for object recognition [18] | 98.97 % |
| 4 | Human Performance [4] | 98.84 % |
| 5 | Multi-Scale CNNs [19] | 98.31 % |

This article compares three of CNN algorithm architectures accuracy performances: AlexNET, DarkNET-53, and EfficientNET-b0, respectively. Details of the respective algorithms are described in Chapter II, and the details of simulation results and discussions are presented in Chapter III. Chapter IV interprets the performances and future-works of the architectures and draws conclusions.

## 2. Materials and Method

### 2.1 AlexNET Architecture

The Large-Scale Visual Recognition Challenge (LSVRC) is a competition in which research groups participate to evaluate learning algorithms on a large-label image dataset (ImageNet). The competition aimed to achieve higher accuracy in various visual recognition tasks. AlexNET is a type of convolutional neural network architecture developed by Alex Krizhevsky which won the LSVRC competition in 2012 [20]. A tabular representation of the architecture is shown in Table 2.

**Table 2.** *AlexNET architecture* [20]

| Layer | Filter Size | Size | Activation Function |
|---|---|---|---|
| Input | - | 227x227x3 | - |
| Convolution 1 | 11x11 | 55x55x96 | ReLu |
| Max Pooling 1 | 3x3 | 27x27x96 | - |
| Convolution 2 | 5x5 | 27x27x256 | ReLu |
| Max Pooling 2 | 3x3 | 13x13x256 | - |
| Convolution 3 | 3x3 | 13x13x384 | ReLu |
| Convolution 4 | 3x3 | 13x13x384 | ReLu |
| Convolution 5 | 3x3 | 13x13x256 | ReLu |
| Max Pooling 3 | 3x3 | 6x6x256 | - |
| Dropout 1 | - | 6x6x256 | - |
| Fully Connected 1 | - | 4096 | ReLu |
| Dropout 2 | - | 4096 | - |
| Fully Connected 2 | - | 4096 | ReLu |
| Fully Connected 3 | - | 1000 | Softmax |

## 2.2 DarkNET-53 Architecture

Image classification is a kind of application area of convolutional neural networks. YOLO (You Look Only Once) is an efficient real-time object recognition algorithm first described [21].

Darknet-53 is a convolutional feature extraction network mainly consisting of a series of 1x1 and 3x3 convolutional layers, with a total of 53 layers including the last fully connected layer but excluding the residual layer. Each convolutional layer is followed by a batch normalization (BN) layer [22] and a LeakyReLu [23] layer. A tabular representation of the architecture is shown in table 3.

**Table 3.** *DarkNET-53 architecture* [24].

| Layer | Filter Size | Repeat | Output Size |
|---|---|---|---|
| Input | - | - | 416x416 |
| Convolution | 3x3 | 1 | 416x416 |
| Convolution | 3x3 | 1 | 208x208 |
| Convolution | 1x1 | 1 | 208x208 |
| Convolution | 3x3 | | 208x208 |
| Residual | | | 208x208 |
| Convolution | 3x3 | 1 | 104x104 |
| Convolution | 1x1 | 2 | 104x104 |
| Convolution | 3x3 | | 104x104 |
| Residual | | | 104x104 |
| Convolution | 3x3 | 1 | 52x52 |
| Convolution | 1x1 | 8 | 52x52 |
| Convolution | 3x3 | | 52x52 |
| Residual | | | 52x52 |
| Convolution | 3x3 | 1 | 26x26 |
| Convolution | 1x1 | 8 | 26x26 |
| Convolution | 3x3 | | 26x26 |
| Residual | | | 26x26 |
| Convolution | 3x3 | 1 | 26x26 |
| Convolution | 1x1 | 4 | 13x13 |
| Convolution | 3x3 | | 13x13 |
| Residual | | | 13x13 |

### 2.3 EfficientNET-b0 Architecture

EfficientNET can be defined as both convolutional neural network architecture and a scaling method that uniformly scaled the dimensions of network depth width and resolution, respectively in a compound coefficient called fi constant. The intuition justifies the compound scaling method that if the input image is bigger, the network needs more layers to increase the receptive field and more channels to capture more fine-grained patterns on the bigger image [25].

Contrary to other classical CNN architectures, EfficientNET uses an inverted residual block called MBConv which was initially proposed for the architecture of MobileNETV2 [26] for improving efficiency causes. EfficientNET has various versions from b0 to b7 inside itself, in this paper the b0 version is selected. The total amount of parameters in EfficientNet-b0 architecture is 5.3 million. The tabular representation of the EfficientNET-b0 baseline network is given in Table 4.

***Table 4.*** *EfficientNET-b0 architecture* [25].

| Layer | Filter Size | Resolution |
|---|---|---|
| Input | - | 224x224 |
| MBConvolution | 3x3 | 112x112 |
| MBConvolution | 3x3 | 112x112 |
| MBConvolution | 5x5 | 56x56 |
| MBConvolution | 3x3 | 28x28 |
| MBConvolution | 5x5 | 14x14 |
| MBConvolution | 5x5 | 14x14 |
| MBConvolution | 3x3 | 7x7 |
| Convolution | 1x1 | 7x7 |
| Pooling & Fully Connected | - | 7x7 |

## 3. Results and Discussion

In our work, to provide an efficient training, 70% of the GTRSB data set were utilized for training, whereas the remaining 30% was employed for training validation and test data as 15%-15% respectively. In that regard, we observe that classification provided by the EfficientNet-b0 of CNN algorithms allows us to classify the traffic signs dataset better. To this end, according to simulation results that were obtained from MATLAB, all the algorithms performed quite well, i.e., with an accuracy of 94.69%, 97.45% and 98.64% for DarkNet-53, AlexNET and EfficientNET-b0, respectively. As might be shown from the obtained results, the EfficientNET-b0 algorithm takes the best result with 98.64 % accuracy (performance) value on GTSRB dataset. Obviously, the performance of EfficientNET-b0 can be considered as one of the outperformers according to the results summarized in Table 1.

The performance comparison of the related CNN algorithms is tabularized in Table 5.

***Table 5.*** *Comparison results of related CNN algorithms.*

| Layer | Performance (%) |
|---|---|
| EfficientNET-b0 | 98.64 % |
| AlexNET | 97.45 % |
| DarkNET-53 | 94.69 % |

## 4. Conclusion

In this paper, trained AlexNET, DarkNET-53, and EfficientNET-b0 models for classifying the German Traffic Signs Recognition Benchmark dataset were successfully implemented in the simulation base. Here, the classification methods based on related CNN architectures were implemented on MATLAB and analyzed with evaluation metrics. The simulation outcomes show the EfficientNET-b0 classifying method based on the CNN Deep Neural Network classifier model meets the objective efficiently and produces a better accuracy performance outcome for classifying traffic signs data compared with other related methods which AlexNET and DarkNET-53, respectively. As a potential future work, for increasing knowledge about this classification issue, the amount of data on related algorithms may be changed to observe result variety better.

**Declaration of Interest**

The authors declare that there is no conflict of interest.

**References**

[1] J. Zhang, W. Wang, C. Lu, J. Wang, and A. K. Sangaiah, "Lightweight deep network for traffic sign classification," *Annales des Telecommunications/Annals of Telecommunications*, vol. 75, no. 7–8, 2020.

[2] V. H. Phung and E. J. Rhee, "A deep learning approach for classification of cloud image patches on small datasets," *Journal of Information and Communication Convergence Engineering*, vol. 16, no. 3, 2018.

[3] J. Stallkamp, M. Schlipsing, J. Salmen, and C. Igel, "The German Traffic Sign Recognition Benchmark: A multi-class classification competition," 2011.

[4] J. Stallkamp, M. Schlipsing, J. Salmen, and C. Igel, "Man vs. computer: Benchmarking machine learning algorithms for traffic sign recognition," *Neural Networks*, vol. 32, 2012.

[5] G. Wang, G. Ren, Z. Wu, Y. Zhao, and L. Jiang, "A hierarchical method for traffic sign classification with support vector machines," 2013.

[6] Y. LeCun, G. Hinton, and Y. Bengio, "Deep learning (2015), Y. LeCun, Y. Bengio and G. Hinton," *Nature*, vol. 521, 2015.

[7] D. P. Kingma, D. J. Rezende, S. Mohamed, and M. Welling, "Semi-supervised learning with deep generative models," in *Advances in Neural Information Processing Systems*, 2014, vol. 4, no. January.

[8] J. Zhang, C. Lu, X. Li, H. J. Kim, and J. Wang, "A full convolutional network based on DenseNet for remote sensing scene classification," *Mathematical Biosciences and Engineering*, vol. 16, no. 5, 2019.

[9] G. Pang, C. Shen, L. Cao, and A. van den Hengel, "Deep Learning for Anomaly Detection: A Review," *ACM Computing Surveys*, vol. 54, no. 2. 2021.

[10] L. Liu *et al.*, "Deep Learning for Generic Object Detection: A Survey," *Int J Comput Vis*, vol. 128, no. 2, 2020.

[11] Z. Bi, L. Yu, H. Gao, P. Zhou, and H. Yao, "Improved VGG model-based efficient traffic sign recognition for safe driving in 5G scenarios," *International Journal of Machine Learning and Cybernetics*, vol. 12, no. 11, 2021.

[12] F. Zaklouta and B. Stanciulescu, "Real-time traffic sign recognition in three stages," *Rob Auton Syst*, vol. 62, no. 1, 2014.

[13] Y. Chen, W. Xu, J. Zuo, and K. Yang, "The fire recognition algorithm using dynamic feature fusion and IV-SVM classifier," *Cluster Comput*, vol. 22, 2019.

[14] Y. Chen, J. Xiong, W. Xu, and J. Zuo, "A novel online incremental and decremental learning algorithm based on variable support vector machine," *Cluster Comput*, vol. 22, 2019.

[15] G. Wang, G. Ren, Z. Wu, Y. Zhao, and L. Jiang, "A robust, coarse-to-fine traffic sign detection method," 2013.

[16] Á. Arcos-García, J. A. Álvarez-García, and L. M. Soria-Morillo, "Deep neural network for traffic sign recognition systems: An analysis of spatial transformers and stochastic optimisation methods," *Neural Networks*, vol. 99, 2018.

[17] D. Cireşan, U. Meier, J. Masci, and J. Schmidhuber, "Multi-column deep neural network for traffic sign classification," *Neural Networks*, vol. 32, 2012.

[18] B. Gecer, G. Azzopardi, and N. Petkov, "Color-blob-based COSFIRE filters for object recognition," *Image Vis Comput*, vol. 57, 2017.

[19] P. Sermanet and Y. Lecun, "Traffic sign recognition with multi-scale convolutional networks," 2011.

[20] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun ACM*, vol. 60, no. 6, 2017.

[21] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016, vol. 2016-Decem.

[22] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *32nd International Conference on Machine Learning, ICML 2015*, 2015, vol. 1.

[23] D. A. Clevert, T. Unterthiner, and S. Hochreiter, "Fast and accurate deep network learning by exponential linear units (ELUs)," 2016.

[24] H. Ma, Y. Liu, Y. Ren, and J. Yu, "Detection of collapsed buildings in post-earthquake remote sensing images based on the improved YOLOv3," *Remote Sens (Basel)*, vol. 12, no. 1, 2020.

[25] M. Tan and Q. v. Le, “EfficientNet: Rethinking model scaling for convolutional neural networks,” in *36th International Conference on Machine Learning, ICML 2019*, 2019, vol. 2019-June.

[26] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, “MobileNetV2: Inverted Residuals and Linear Bottlenecks,” 2018.