Technological Applied Sciences Status : Research Article
ISSN: 1308-7223 Received: 13.06.2022
Article ID: 2A0189 Accepted: 25.10.2022

İsmail Akgül
Erzincan Binali Yıldırım University, iakgul@erzincan.edu.tr,
Erzincan-Türkiye
Yıldız Aydın
Erzincan Binali Yıldırım University, yciltas@erzincan.edu.tr,
Erzincan-Türkiye

DOI	http://dx.doi.org/10.12739/NWSA.2022.17.4.2A0189	
ORCID ID	0000-0003-2689-8675	0000-0002-3877-6782
Corresponding Author	İsmail Akgül	

OBJECT RECOGNITION WITH DEEP LEARNING AND MACHINE LEARNING METHODS

ABSTRACT

The field of computer vision, which is widely studied area, is basically the imitation of the human vision system with digital devices. Computer systems perform operations through digital images or video images and decide according to the result. In this context, object recognition must be performed at the first stage in order to extract meaningful information from the image. In this study, the application of object recognition was developed using the deep learning method, which is especially popular in recent years. It has also been compared with classical machine learning methods often used in recognition applications. The proposed method developed with Convolutional Neural Network (CNN) has been compared by using the Histogram of Oriented Gradient (HOG) features and classifying them with Support Vector Machines (SVM) and K-Nearest Neighbor (KNN) methods. Experimental results show that the proposed CNN method is more successful than HOG+SVM, and HOG+KNN methods.

Keywords: CNN, HOG, SVM, KNN, Object Recognition

1. INTRODUCTION

Digital image use has grown as a result of the decrease in the cost of image capturing equipment along with increasingly developing technology. This rise has made computer vision a very popular field studied. The computer vision (CV) technology produces automatic and real time results that make it simpler in many fields of medicine, education etc. for similar activities to function. Computer vision approach is developed by parsing business tasks that can include object recognition, surveillance and description. The first step in any project for computer vision is to collect useful information from a scene, train a device for the best detection and separation of visible objects. A unique object, a certain individual, a particular car and model may be used, etc. The utilization of technical products that have been created for the purposes of identifying objects minimizes human resources. An instance of an object when it is detected (e.g. a face), it is possible to obtain more information, including the following: (i) to recognize a specific situation (e.g., to identify the subject's face), (ii) for object tracking over an image sequence (for example, to track the face in a video), and (III) to obtain more information about the object (e.g., to determine the gender of the subject), and at the same time (a) the presence or location of other objects in the scene (e.g., may be close to the face and a hand on a similar scale) and (B) more information about the

How to Cite:
Akgül, İ. ve Aydın, Y., (2022). Object Recognition with Deep Learning and Machine Learning Methods. Technological Applied Sciences, 17(4):54-61.
DOI:10.12739/NWSA.2022.17.4.2A0189.

scene (e.g., the type of scene, indoor and outdoor etc.), among other contextual information.

In this study, the CNN method for object recognition was compared with the Histogram of Oriented Gradient-HOG [1] attribute, a global feature used in classical classifiers (Support Vector Machines-SVM [2] and K-Nearest Neighbor-KNN [3]). Related works are given in the second part (II) of the study. The third (III) section describes the method of object recognition. Experimental results were given in the fourth (IV) section, and finally the results are discussed in the fifth (V) section.

2. RESEARCH SIGNIFICANCE

In this study, the application of object recognition was developed using the deep learning method, which is especially popular in recent years. It has also been compared with classical machine learning methods often used in recognition applications.

Highlights:

- Object recognition using artificial intelligence methods
- Development of a CNN model
- Comparison of deep learning and machine learning methods

3. RELATED WORKS

Alaeddine and Jihene have established a new deep network model the network in their work on the CIFAR-10 standard dataset [4]. Due to their CIFAR-10 data set tests, which provided a deeper network model than the deep network model, they showed that increased depth leads to better results of recognition. Abouelnaga et al. using principal component analysis (PCA) on CIFAR-10 data, combined it with the Convolutional Neural Network (CNN) to reduce K-Nearest Neighbor (KNN) over conformity and improve its accuracy and found that the KNN-CNN group increased the accuracy of the model [5]. Thakkar et al. added batch normalization layers between the convolution and activation layers of the DenseNet, VGG, Residual network, and Inception (v3) network models. In this way, they compared both ReLU and EluP activation functions by training on the CIFAR-10 data set and achieved very significant achievements in these models [6]. In another study, Coates et al. have applied ready-to-use feature learning algorithms such as sparse auto-encoders, sparse RBM’s, K-mean clustering, and Gaussian mixtures to CIFAR-10, NORB, and STL data sets using single-layer networks. They conducted experiments on data sets using these multiple unsupervised feature learning algorithms to identify the effect of various parameters on classification performance. They have shown that the K-means clustering algorithm, an extremely simple learning algorithm that does not have hyper parameters to set, can achieve state-of-the-art performance when used in conjunction with network parameters as a result of their work. They achieved higher accuracy in the CIFAR-10 and NORB datasets than previously published results. In addition, they determined that while complex algorithms may have more representation power, simple but faster algorithms may be more competitive [7].

Xu et al. convolutional neural networks in ReLU (rectified linear units), Leaky ReLU (rectified linear units leak), PReLU (parametric rectified linear units) and PReLU (random leaks rectified linear units) have studied the performance of different types of activation functions such as. An analysis of four types of rectified activation functions on the CIFAR-10 and CIFAR-100 data sets showed that the common belief that ReLU performs best is not true, and that Leaky Relu performs better than the original ReLU [8]. Giuste and

Vizcarra have successfully classified the Histogram of Oriented Gradients (HOG) and pixel intensities to classify images in the CIFAR-10 dataset, using combinations of different image property sources from both manual and deep learning approaches, but they said it is still need to be improved. Furthermore, the optimized model CIFAR-10 of the VGG16 ImageNet (CIFAR-VGG) improved the classification of images further. They demonstrate that multiple ways to remove significant features from images will increase classification by integrating them [9]. Romanuke's research discussed the method of dropping out of the convolutional neural networks in order to avoid overcoming. CIFAR-10 has a mean gain of between 10%, and 50 percent from its experiments with this technology on EEACL26 and NORB results, by using two generic network architectures with 4 or 5 convolutional layers [10]. In his research, Tang applied the replacement of the softmax layer used in the output labels in classification operations with a linear support vector machine, as opposed to feeding the result of supervised or unsupervised learning in deep networks as input to SVM. The replacement of softmax with linear SVMs showed major gains to the 2013 data sets of MNIST, CIFAR-10 and ICML [11].

4. EXPERIMENTAL METHOD

4.1. Dataset

Krizhevsky and Hinton researched, data set sizes 32x32 and airplane, automobile, bird, cat, deer, dog, frog, horse, ship, and truck consisting of 60000 color images belonging to 10 different classes colored in the study CIFAR-10 (Canadian Institute for Advanced Research) [12]. This dataset is generally used for training algorithms for machine learning and vision [4]. Figure 1 shows several examples of CIFAR-10 data.

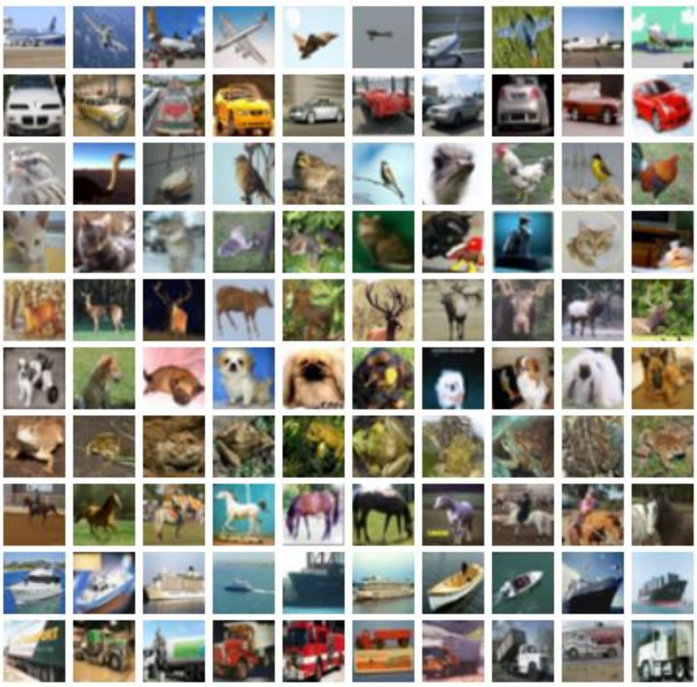

Figure 1. Example images of 10 different classes in the CIFAR-10 data

The training and evaluation datasets include 50.000 and 10.000 pictures respectively, with the same number of photographs from each class.

4.2. Convolutional Neural Network (CNN) Model

Deep learning is one of the developing areas of computational science that involves large amounts of data to train a model [6]. On the other hand, an evolutionary neural network is one of the most common deep-learning approaches and is used effectively for applications related to image classification. In this direction, training and testing were carried out on the proposed CNN model shown in Figure 2.

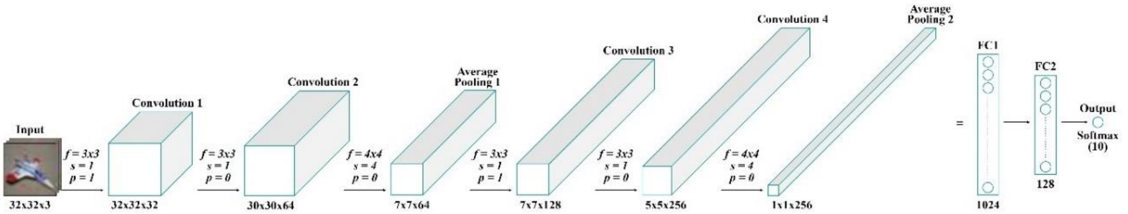

Figure 2. Recommended CNN model
(f:filter size, s:stride, p:padding, FC:fully connected)

In a model where two convolutions are established with an average common logic and this process is repeated twice, an input image is taken in 32x32 color format and a 10-Class output is obtained. The Model consists of 18 layers, 422.602 parameters, including 4 convolutions, 2 Joint, 3 forget, 5 ReLu, 1 flattening, 2 full connections and 1 output.

4.3. Histogram of Oriented Gradients (HOG)

The histogram of oriented gradient is one of the attributes used in different applications such as license plate recognition, vehicle recognition, and object recognition in the field of computer vision. Creates an attribute identification in localized areas using the tilt component of the picture. The HOG identifier describes objects by the distribution of density slopes or edge slopes. When extracting the HOG attribute, the image is divided into cells that are related to each other, and a histogram of the tilt orientations is calculated for the pixels in each cell. Combining these histograms creates a HOG identifier for the image. Finally, local histograms are contrast-normalized. The following process steps must be carried out to remove the HOG feature from an image:

- Calculation of image gradient
- Creating directional histograms for each cell
- Normalizing histograms in cell blocks

4.4. Support Vector Machines (SVM)

The basis is structural risk minimization [13] based on support vector machines find the hyperplane that separates the training dataset in the first step by the maximum margin. This hyperplane is used to categorize test samples in the second process. The sample area that cannot be linearly separated by a DVM is moved to a separate sample space, where it can be linearly separated.

This optimization problem is solved by the following equation if the training examples $(S = ((\vec{x_1}, y_1), \ldots, (\vec{x_\iota}, y_i)$, are expressed linearly as the hyperplane $(\vec{w}, b)$.

Minimization:

$$y_i[\vec{w}.\vec{x_\iota} + b] \geq 1, i = 1, \ldots, l \quad (1)$$

w here shows the weight vector, x input vector, b bias.

It realizes the maximum edge subplane.

4.5. K-Nearest Neighbor (KNN)

In the nearest neighbor algorithm, the test data to be classified is classified based on the classes of close neighbors. The KNN process, known as the instance-based clustering, comprises steps to decide the next neighbor and to identify these neighbors with class tags. Here K shows the number of examples (neighbors) to be referenced. The groups of its immediate neighbors are numbered accordingly. As this approach typically accounts for more than one neighbor, k is referred to as the closest neighbor classifier. In the training instance of an unknown I instance, the distance between x_i is calculated by the following formula.

$$d(I, x_i) = \sum_{f \in F} w_f \delta(I_f, x_{if}) \tag{2}$$

d here shows the training dataset, f attributes.

4.6. Media and Libraries Used

The methods applied in the study were made on a laptop with a 2GB video card with i7 processors. Application codes are written in Python programming language.

5. FINDINGS AND DISCUSSIONS

In its original form, training was conducted according to the parameters shown in Table 1 on the CIFAR-10 data set, which was divided into 80% training and 20% testing.

Table 1. Training parameters of CIFAR-10 data

Parameters	Values
Epoch	50
Mini Batch Size	32
Dropout	0.5
Activation Function	ReLu
Optimization Algorithm	Adamax

Figure 3 shows the accuracy and loss graph obtained as a result of training of the CNN model. When the figure is examined, it is seen that the proposed model starts learning with an accuracy rate as low as 37% and goes to learning regularly and reaches 80% success accuracy as a result of 50 iterations. It is also understood that learning takes place at every stage of education. Similarly, the model which began training at a loss rate of 1.634, as a result of 50 iterations, achieved a loosing rate of 0.614.

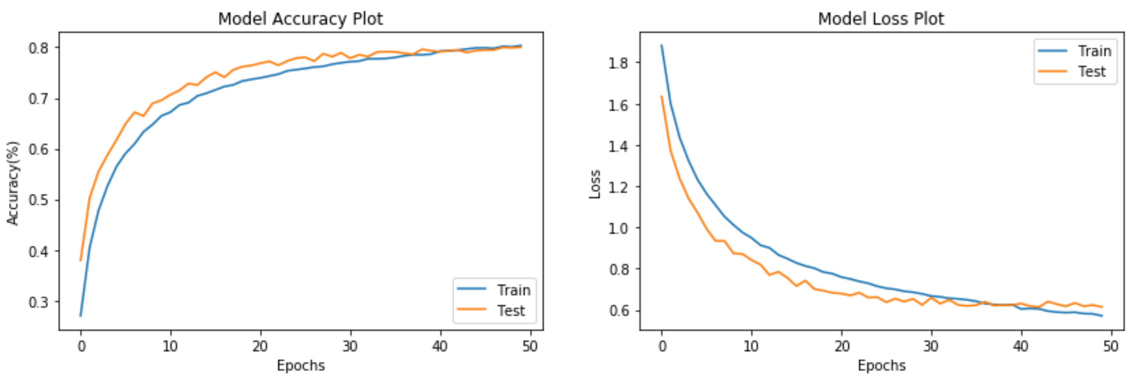

Figure 3. Accuracy and loss plot

The final results of accuracy and loss rates achieved as a result of the training are given in Table 2.

Table 2. Accuracy rate

Method	Accuracy (%)
HOG+SVM	51
HOG+KNN	49
CNN	80

The confusion matrices obtained by using CNN, SVM, and KNN classifiers from the test dataset with a total of 10000 images containing 1000 images from each class are given in Figure 4, Figure 5 and Figure 6, respectively. In Figure 4, where the proposed method is shown, it is seen that the highest performance is in the car class with 90.4%, and the lowest performance is in the cat class with 52.9%. It is worth noting that the cat class is most confused with the class where there are images of dogs.

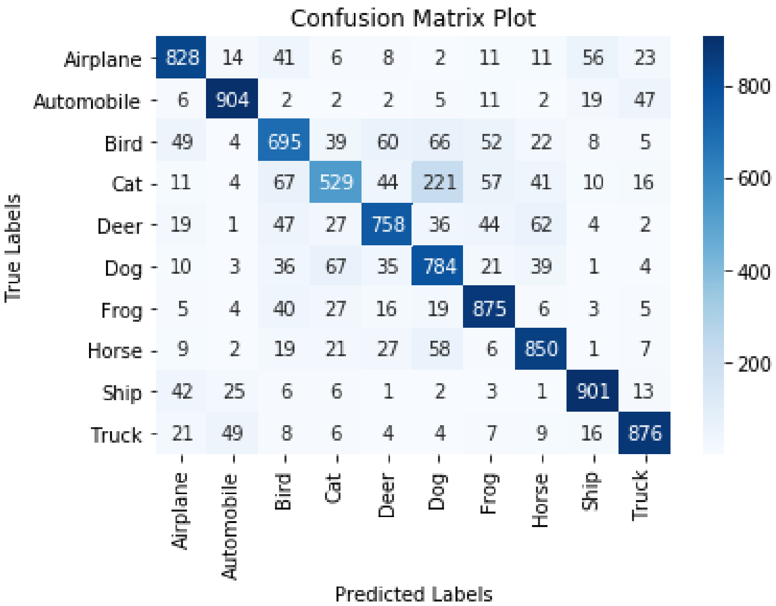

Figure 4. Confusion matrix from the application developed with the CNN classifier

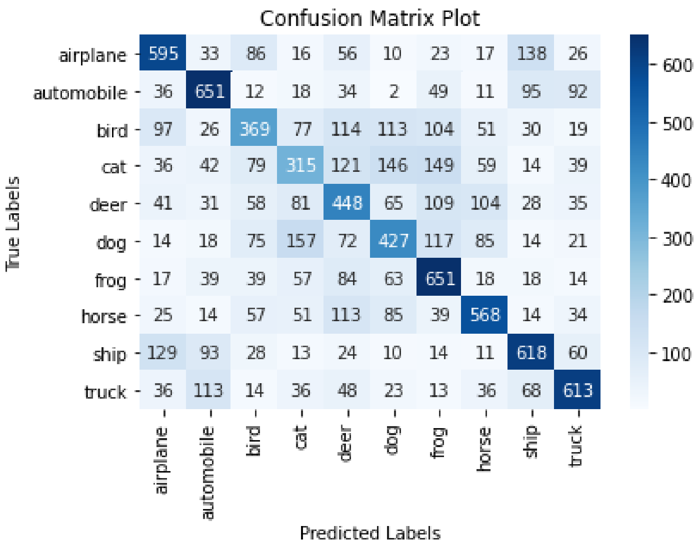

Figure 5. Confusion matrix from the application developed with the SVM classifier

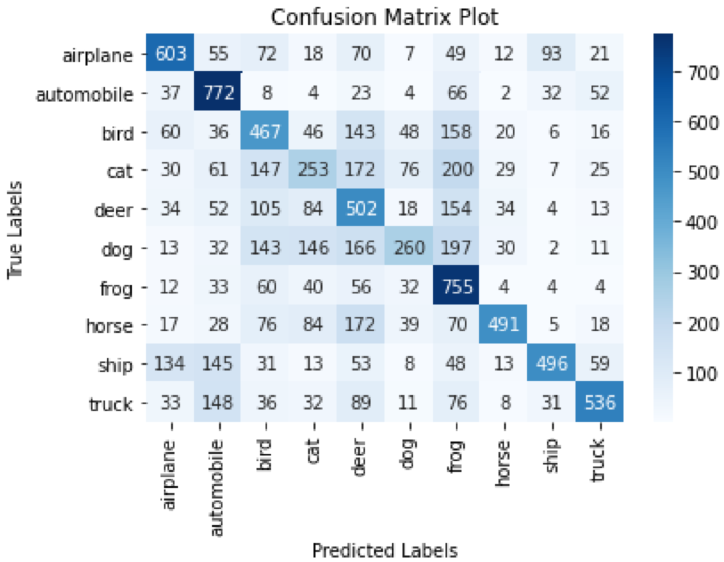

Figure 6. Confusion matrix from the application developed with the KNN classifier

6. CONCLUSION AND RECOMMENDATIONS

This research examined the success in object recognition of classical machine classifiers and methods of deep learning. HOG was used as an attribute; moreover, SVM, KNN, and CNN were used as classifiers. In the experimental results, the HOG+KNN application was the most unsuccessful with an accuracy rate of 49%, while the method developed with CNN was the most successful method with an accuracy rate of 80%. This shows that deep learning methods are more suitable for this dataset due to the large number of images. In conclusion, the experiments carried out within the scope of this study show the effectiveness of the CNN process. A hybrid application for object recognition applications is intended for future research with the integration of traditional master learning methods and deep learning methods.

CONFLICT OF INTEREST

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

FINANCIAL DISCLOSURE

The authors did not receive any financial support in conducting this study.

DECLARATION OF ETHICAL STANDARDS

The authors of this article declare that the materials and methods used in this study do not require ethical committee permission and/or legal-special permission.

REFERENCES

[1] Dalal, N. and Triggs, B., (2005). Histograms of oriented gradients for human detection. In 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05), San Diego, CA, USA, pp:886-893. https://doi.org/10.1109/cvpr.2005.177

[2] Cortes, C. and Vapnik, V., (1995). Support-vector networks. Machine learning, 20(3):273-297. https://doi.org/10.1007/bf00994018
[3] Cover, T. and Hart, P., (1967). Nearest neighbor pattern classification. IEEE transactions on information theory, 13(1):21-27. https://doi.org/10.1109/TIT.1967.1053964
[4] Alaeddine, H. and Jihene, M., (2021). Deep residual network in network. Computational Intelligence and Neuroscience, 2021. https://doi.org/10.1155/2021/6659083
[5] Abouelnaga, Y., Ali, O. S., Rady, H., and Moustafa, M., (2016). Cifar-10: Knn-based ensemble of classifiers. In 2016 International Conference on Computational Science and Computational Intelligence (CSCI), Las Vegas, NV, USA, pp:1192-1195. https://doi.org/10.1109/csci.2016.0225
[6] Thakkar, V., Tewary, S., and Chakraborty, C., (2018). Batch Normalization in Convolutional Neural Networks—A comparative study with CIFAR-10 data. In 2018 fifth international conference on emerging applications of information technology (EAIT), Kolkata, India, pp:1-5. https://doi.org/10.1109/eait.2018.8470438
[7] Coates, A., Ng, A., and Lee, H., (2011). An analysis of single-layer networks in unsupervised feature learning. In Proceedings of the fourteenth international conference on artificial intelligence and statistics, pp:215-223.
[8] Xu, B., Wang, N., Chen, T., and Li, M., (2015). Empirical evaluation of rectified activations in convolutional network. arXiv preprint arXiv:1505.00853.
[9] Giuste, F.O. and Vizcarra, J.C., (2020). CIFAR-10 Image Classification Using Feature Ensembles. arXiv preprint arXiv:2002.03846.
[10] Romanuke, V.V., (2017). Appropriateness of DropOut layers and allocation of their 0.5 rates across convolutional neural networks for CIFAR-10, EEACL26, and NORB datasets. Applied Computer Systems, 22(1):54-63. https://doi.org/10.1515/acss-2017-0018
[11] Tang, Y., (2013). Deep learning using linear support vector machines. arXiv preprint arXiv:1306.0239.
[12] Krizhevsky, A. and Hinton, G., (2009). Learning multiple layers of features from tiny images (Technical Report). University of Toronto.
[13] Rudnicky, A.I., Brennan, R.A., and Polifroni, J.H., (1988). Interactive problem solving with speech. The Journal of the Acoustical Society of America, 84(S1):S213-S213. https://doi.org/10.1121/1.2026163.