

Towards classification of images by using block-based CNN

Retaj Matroud Jasim, Tayseer Salman Atia

Department of Computer Engineering, Al Iraqia University, Baghdad, Iraq

Article Info

Article history:

Received Sep 6, 2022

Revised Oct 6, 2022

Accepted Oct 21, 2022

Keywords:

Classifier

Convolutional neural network

DenseNet

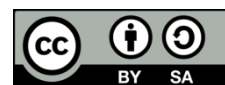
Multi-layer perceptron

ResNet

ABSTRACT

Image classification is the process of assigning labeling to the input images to a fixed set of categories; however, assigning labels to the image is difficult by using the traditional method because of the large number of images. To solve this problem, we will resort to deep learning techniques. Which is enables computers to recognize and extract visual characteristics. The convolutional neural network (CNN) is a deep neural network used for many purposes, such as image classification, detection, and face recognition, due to its high-performance accuracy in classification and detection tasks. In this paper, we develop CNN based on the transfer learning approach for image classification. The network comprises two types of transfer learning, ResNet and DenseNet, as building blocks of the network with an multilayer perceptron (MLP) classifier. The proposed method does not need to preprocess before these datasets that input into the network. It was train on two datasets: the Cifar-10 and the Sign-Traffic datasets. We conclude that the proposed method achieves the best performance compared with other states of the art. The accuracy gained is 97.45% and 99.45%, respectively, where the proposed CNN increased the accuracy compared to other methods by 3%.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Retaj Matroud Jasim

Department of Computer Engineering, Iraqi University

Baghdad, Iraq

Email: retaj.m.jasim@students.aliraqia.edu.iq

1. INTRODUCTION

A large volume of published studies classifying images is a significant problem in artificial vision systems and has been for decades. This area aims to provide a label to a picture based on the information visible [1]. Researchers may benefit from image classification since it allows them to organize images according to their shared characteristics. For example, if images A and B have specific characteristics, we may classify and label them as part of the same set. Their studies tested the algorithms in various ways and made comparisons [2]. Object detection and classification are among the most challenging tasks regarding image processing. Several object classification methods have been suggested throughout the years to address these issues [3].

Researchers and developers are now able to approach bigger models to handle complicated issues, something that was previously impossible with traditional artificial neural networks (ANNs) [4]. Most studies have used hand-crafted features like histogram of oriented gradients (HoG) [5] or scale-invariant feature transform (SIFT) [6] to characterize a picture with discriminatory power. Next, the collected features are fed into a learnable classifier, such as a support vector machine, a random forest, or a decision tree.

However, it becomes a highly challenging challenge to discover characteristics from a large number of provided photos. For these and other reasons, a new model based on deep neural networks is in the future. Convolutional neural network (CNN) is widely used for image identification and is one of the most well-known deep neural networks. Many computer vision and natural processing applications, such as image

identification, and object identification, have benefited from its utilization [3]. Furthermore, CNN provides outstanding efficiency in solving machine learning issues. For instance, a complete image classification dataset is useful for programs with images. In the previous decade, CNN has seen widespread use in the effort to enhance picture categorization precision. Since CNN permits the cooperative learning of features and classifiers, it can provide superior classification accuracy for big data sets [7]. The bag-of-features pipeline has recently been used in picture classification approaches. Clustering is performed using SIFT descriptors [8]. Features are collected via spatial pooling [9] histogram encoding [10] and, most recently, fisher vector encoding [11]. While these representations have been shown to provide workable outcomes, it is not immediately clear whether or not they are ideal for the job at hand. This requires a lot of time and effort, not to mention the cost of hiring specialized personnel. The AlexNet deep CNN by Krizhevsky *et al.* [12] stands out as the most novel of these networks (i.e., graphics processing unit (GPU), an intense network of 60 million and 650,000 neurons). AlexNet embraced the challenge, outperforming its rivals and achieving a top-five error rate of just 15.3%. The error rate in the top 5 spots was close to 26.2%; this was not a CNN variant. Gehring *et al.* [13] developed a CNN architecture for learning in sequence. The model outperforms the recurrent models, which failed to understand the compositional nature of the sequences. In addition, all of the components may be parallelized entirely during training for more efficient calculations.

To further facilitate a more organic optimization, nonlinearities are made constant and independent of the input length in. Ye *et al.* [14], developed an alternative approach to CNN. They detailed the pixel-by-pixel operation of the CNN and showed its use in several contexts. Since other CNN kinds have more features and processing capacity, they are employed by many academics as primary image classifiers in their studies. The results of a comparison of the suggested approach with others indicated that it was superior. To categorize pictures more quickly and accurately than previous models, Han *et al.* [15] suggested a novel CNN approach, which they tested on six distinct small-sized datasets to verify their findings. After analyzing the outcomes, they concluded that this strategy is simple for small datasets. The novel-based model for image classification and multi-label method was given by Song *et al.* [16]. The basic premise of this study was to train a model using various data sources, including multi-label picture data. When many labels are needed, this study might be helpful. According to the paper's declared accuracy parameters, its offered model beats other presented models. In a recent study, authors M. A method for classifying images on embedded systems was developed and proven in a study by Çalik and Demirci [17]. The researchers employed CNN to achieve an accuracy of 85.9% on the Cifar-10 dataset. In their paper "Empirical study of the output of common convolution neural networks for object identification in real-time video feeds," Sharma *et al.* [18] published the results of such a study.

The following is the structure of the paper: section 2 summarizes the components used to construct the suggested model. Section 3 discusses the proposed classification strategy for photographs. Section 4 describes the experimental setup, which includes the datasets that are used to train the proposed model, the results, the evaluation metrics that evaluate the accuracy and loss of the model, the dissection of the results, and the comparison of the accuracy of the proposed model to that of other previous models. Section 5 concludes the model proposal.

2. METHOD

2.1. DenseNets and ResNets blocks of suggested framework

ResNet [19] and DenseNet [20] are two CNNs suggested as cutting-edge in recent years. ResNet and DenseNet are two successful deep learning architectures primarily related to their respective building components, ResNet blocks (RBs) and DenseNet blocks (DBs). Figure 1 shows an example of RB composed of three convolutional layers and one skip connection. The names of the convolutional layers are Conv1, Conv2, and Conv3. On Conv1, a reduced number of filters with a size of 1×1 minimizes the spatial dimension of the input in order to reduce the problematic computational of Conv2. On Conv2, filters with a larger size, such as 3×3 , are used to learn spatially identical characteristics. On Conv3, a filter size of 1×1 is employed again, and this time the spatial dimension is raised so that more characters may be generated. The output of Conv3 is combined with the input to form the output of the RD. In case the input and Conv3's output spatial sizes are different, a series of convolutional operations with 1×1 sized filters are performed on the input to attain the same dimensionality as the result of Conv3 for the sum. Figure 2 displays an example of a DB. In the interest of simplicity, the DB has just four convolutional layers. In practice, the number of convolutional layers in the DB may be adjusted by the user. Each convolutional layer in the DB takes inputs from the input data and the output of all previously convolutional layers. Attempts in [21], [22] have investigated the mechanism underlying the success of running backs and safeties, revealing that RBs and DBs can decrease the negative effect of the gradient's vanishing problem [23], based on which a deep architecture is possible to efficiently learn the classification tasks of the training dataset and subsequently

enhance the classification precision. Additionally, it has been suggested that dense connections in DBs may reuse low-level attributes to improve the acquired discrimination of characteristics in the upper layers of CNNs [20]. The suggested method selects RBs and DBs as the basis primarily due to their positive features.

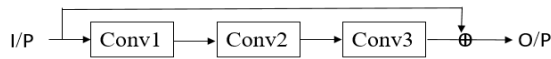


Figure 1. RB

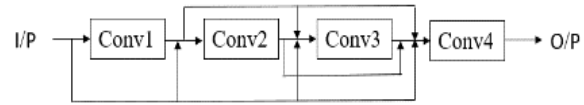


Figure 2. DB

2.2. Suggested framework for image classification

We suggest a framework based on a model of deep neural networks with four blocks. After a fully connected layer and a multi-layer perceptron as a classifier using softmax as the activation function, the network consists of the first two blocks from ResNet50 and the second two blocks from DenseNet121. By default, these blocks share the same configuration parameters. In both the Cifar-10 and the Sign-Traffic datasets, this CNN has shown effective during training. In every test, we split our dataset into two halves: training and validation. The best model is selected in CNN training after the first 30 iterations have the lowest validation loss. The architecture accepts images of varying sizes as input, and the input images are zero-padded. Figure 3 illustrates the suggested network and Table 1 illustrates the settings of the blocks used in design the architecture.

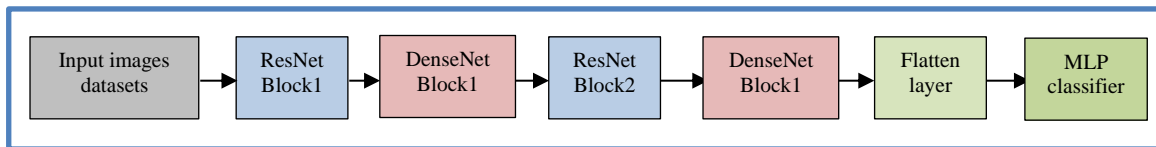


Figure 3. The proposed architecture of the model

Table 1. The settings of building blocks of the model

Model	Building block	Parameter	Classifier
R1-D1-R2-D2	R1	Normlization (batch normalization, local response normalization)	MLP
	Conv2D $\left\{ \begin{matrix} (1 \times 1,64) \\ (3 \times 3,64) \\ (1 \times 1,256) \end{matrix} \right\} \times 3$		
	R2	Activation function (leaky relu) Kernal size (7×7) Stride (2,2) Filter (64)	
	Conv2D $\left\{ \begin{matrix} (1,1,128) \\ (3,3,128) \\ (1,1,512) \end{matrix} \right\} \times 3$		
	Pooling (average pooling)		
	D1	Kernal size (2×2) Stride (2×2) Filter1 (56) Filter2 (28)	
	Conv2D $\left\{ \begin{matrix} (1 \times 1) \\ (3 \times 3) \end{matrix} \right\} \times 6$		
	D2	Padding (zero padding) Regulerization (dropout=0.5) Optimizer (Adam)	
	Conv2D $\left\{ \begin{matrix} (1 \times 1) \\ (3 \times 3) \end{matrix} \right\} \times 12$		
	Pooling (average pooling)		

3. BENCHMARK DATA SETS

3.1. Cifar-10

Cifar-10 is a dataset of natural RGB images of 32×32 pixels [24]. It contains 10 classes with 50,000 training images and 10,000 test images. All of these images have different backgrounds with different light

sources. Objects in the image are not restricted to the one at center, and these objects have different sizes that range in orders of magnitude. The dataset Cifar-10 contains 60,000 color images, with a training set comprising of 50,000 images, a test set containing 10,000 images, all within twenty object classes in ten broad categories: airplane, automobile, bird, cat, deer, dog, frog, horse, ship and truck as shown in Figure 4 with images of size 32×32 pixels.

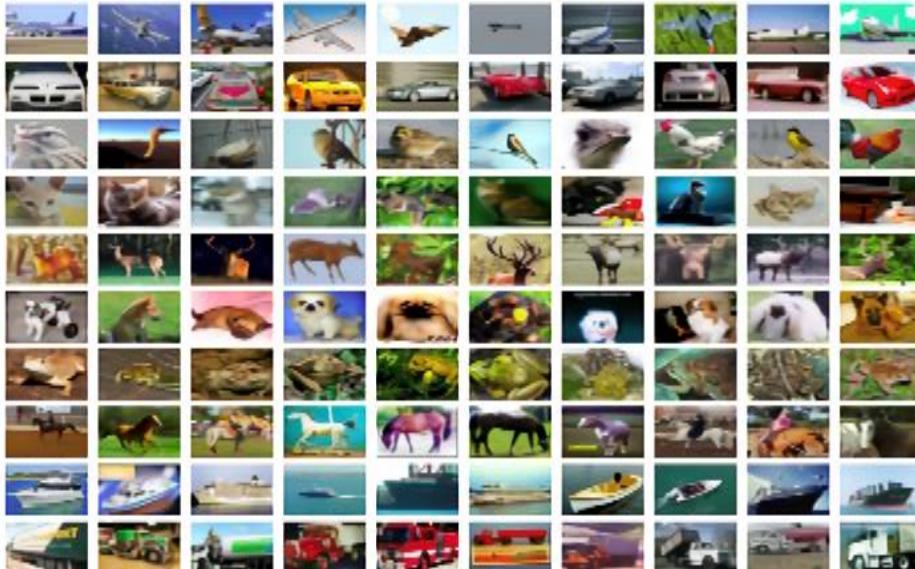


Figure 4. Some images of Cifar-10 dataset [24]

3.2. Sign signal

The traffic sign dataset [25] contains more than 360 images in total, divided into different classes. To avoid using the testing data, we leave 180 images from the training set for validation and 180 test images featuring among four classes "stop sign", "non stop sign", "green light" and "red light". Both training and testing data are distributed over these categories as shown in Figure 5.

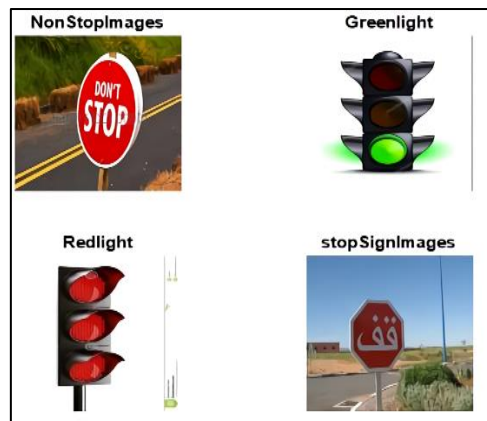


Figure 5. Some image from sign-traffic dataset [25]

4. EXPERIMENT RESULTS AND DISCUSSION

This experiment uses to classify multi class images. In this part, we compare the proposed method with other methods described for image classification in the literature, demonstrating that the suggested network has enough performance for our current needs. Cifar-10, widely utilized in detection and classification applications, and the sign-traffic dataset are used to evaluate the efficacy of the proposed

architecture. This model is built using the Keras library and the Google TensorFlow framework on a machine with 16 GB of RAM and an NVidia GEFORCE GTX 1,650. A learning rate of the Adam optimizer was utilized (0.001), and a batch size of 50 samples was also utilized. Moreover, the model uses the (categorical cross-entropy) loss function. The drop-out is (0.5) used to avoid overfitting.

We used MPL as a classifier, which is expected since it correlates the feature non-linearly to generate all possible patterns. The accuracy receiver operating characteristic (ROC) curve results of Cifar-10 dataset are shown in Figure 6(a), while the error rate ROC curve results of Cifar-10 are shown in Figure 6(b). Moreover, the accuracy ROC curve result of sign-traffic shown in Figure 7(a), and the error rate ROC curve results of sign-traffic shown in Figure 7(b). Table 2 shows the proposed network outperforms competing methods with a classification accuracy of 97.45% of Cifar-10 and 99.45% for sign-traffic datasets. The generated findings are reasonably stable and accurate, giving valuable insights into the classifying performance of the images.maps.

Table 2. The result of the proposed model on two datasets

Model	Datasets	Accuracy (%)	Error rate
Proposed model	Cifar-10	97.45	0.14
Proposed model	Sign-Traffic	99.45	1.72×10^{-4}

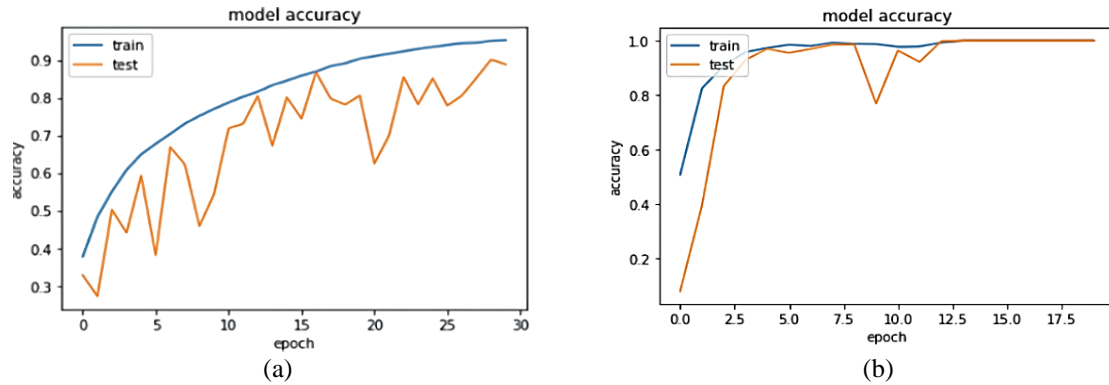


Figure 6. The result of the two datasets (a) the model accuracy of Cifar-10 dataset and (b) the model accuracy of sign-traffic dataset

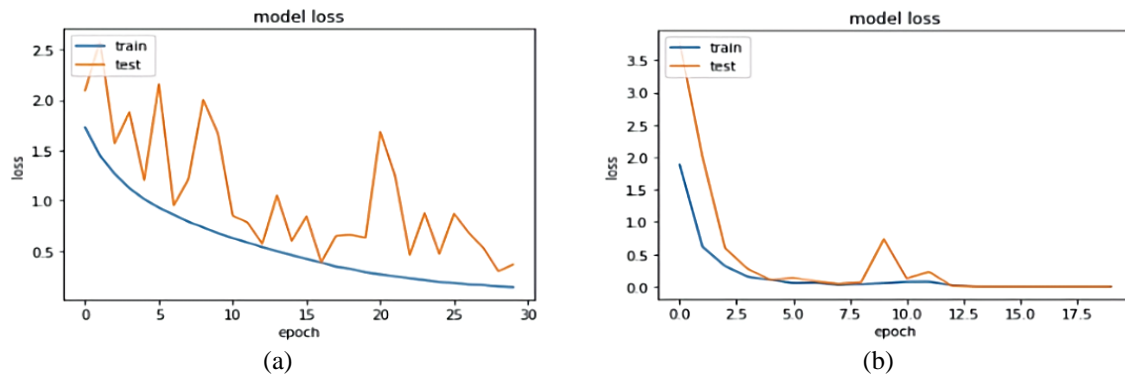


Figure 7. The result of the two datasets (a) the loss of Cifar-10 datasets and (b) the loss of sign-traffic datasets

4.1. Comparisons with state-of-the-art works

To check further the performance of the proposed CNN model. The comparisons are compared among proposed CNN and some state-of-the-art works. Note that handcrafted-AE-CNN by Sun, also compared with CNN proposed by Yim *et al.* [1], and with proposed network by Aamir *et al.* [3] are designed for Cifar-10 dataset classification tasks, so they cannot converge for our forensics task. In addition compare the results of sign-traffic datasets with the proposed CNN by Jmour and Zayen. Table 3 report the accuracy of multi classification on these dataset. We can see that proposed CNN can obtain the best results in multi classification tasks.

Table 3. Comparison of proposed model with other models

Model	Datasets	Accuracy (%)	Error rate
AE-CNN	Cifar-10	95.03	0.04
CNN	Cifar-10	78.29	0.22
CNN	Cifar-10	95.53	0.04
Proposed model	Cifar-10	97.45	0.03
AlexNet	Sign-Traffic	93.33	0.07
Proposed model	Sign-Traffic	99.45	0.005

4.2. Evaluation metrics

We consider our work's accuracy (ACC) and error rate (ERR) metrics to evaluate the model's efficiency. The accuracy helps to know the errors in the measurement values of the models. Accuracy and the error rate are inversely related. High accuracy refers to a low error rate, and a high error rate refers to low accuracy. ACC is derived by dividing the total number of accurate predictions by the total number of observations in the dataset (shown in (1)). ERR is computed by dividing the total number of inaccurate predictions by the total dataset (shown in (2)).

$$ACC = \frac{TP+TN}{P+N} \quad (1)$$

$$ERR = \frac{FP+FN}{P+N} \quad (2)$$

Where (TP + TN) the correct prediction, (FP + FN) the incorrect prediction, (P + N) the total number of the datasets. Which (TP, TN) are taken from confusion matrix are shown in Table 4.

Table 4. Confusion matrix for showing results of classification

Actual	Predictions	
	Real	Fake
Real	True positive (TP)	False negative (FN)
Fake	False positive (FP)	True negative (TN)

5. CONCLUSION AND FUTURE WORK

In a study, we developed a technique that employs a deep neural network and consists of two blocks of two transfer learning approaches, namely ResNet and DenseNet, followed by a fully connected layer. This approach is implemented on Cifar-10 and sign-traffic datasets for training and testing. This kind of learning, namely the CNN, is used to identify image data. We have shown that fine-tuning settings is a crucial and beneficial training strategy. Based on these results, it can be concluded that image categorization using deep neural networks can achieve high performance. The suggested network needed fewer processing resources and less memory. The network enhances classification accuracy and yields acceptable identification outcomes compared to conventional methods. In addition, the network's performance assessment indicates that it may be used to construct a considerably better classifier. In future work, the researchers can use another transfer learning network instead of ResNet and DenseNet for image classification.




REFERENCES

- [1] J. Yim, J. Ju, H. Jung, and J. Kim, "Image classification using convolutional neural networks with multi-stage feature," *Robot Intelligence Technology and Applications* 3, vol. 345, pp. 587–594, 2015, doi: 10.1007/978-3-319-16841-8_52.
- [2] A. Sharma and G. Phonsa, "Image classification using CNN," *International Conference on Innovative Computing and Communication (ICICC 2021)*, 2021.
- [3] M. Aamir, Z. Rahman, W. A. Abro, M. Tahir, and S. M. Ahmed, "An optimized architecture of image classification using convolutional neural network - ProQuest," *International Journal of Image, Graphics and Signal Processing*, vol. 10, no. 10, pp. 30–39, Oct. 2019, doi: 10.5815/ijigsp.2019.10.05.
- [4] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," in *2017 International Conference on Engineering and Technology (ICET)*, Aug. 2017, pp. 1–6, doi: 10.1109/ICEngTechnol.2017.8308186.
- [5] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, Jun. 2005, vol. 1, pp. 886–893, doi: 10.1109/CVPR.2005.177.
- [6] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, Nov. 2004, doi: 10.1023/B:VISI.0000029664.99615.94.
- [7] P. Payal and M. M. Goyani, "A comprehensive study on face recognition: methods and challenges," *The Imaging Science Journal*, vol. 68, no. 2, pp. 114–127, Feb. 2020, doi: 10.1080/13682199.2020.1738741.




- [8] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. Lecun, "OverFeat: integrated recognition, localization and detection using convolutional networks," *arXiv: 2nd International Conference on Learning Representations, ICLR 2014*, pp. 1–16, 2014.
- [9] H. B. Burke, "Artificial neural networks for cancer research: outcome prediction," *Semin Surg Oncol*, vol. 10, no. 1, pp. 73–79, Feb. 1994, doi: 10.1002/ssu.2980100111.
- [10] H. B. Burke *et al.*, "Artificial neural networks improve the accuracy of cancer survival prediction," *Cancer*, vol. 79, no. 4, pp. 857–862, Feb. 1997, doi: 10.1002/(sici)1097-0142(19970215)79:4<857::aid-cnrc24>3.0.co;2-y.
- [11] J. Lampinen, S. Smolander, and M. Korhonen, "Wood surface inspection system based on generic visual features," *Industrial Applications of Neural Networks*, pp. 35–42, 1998, doi: 10.1142/9789812816955_0005.
- [12] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017, doi: 10.1145/3065386.
- [13] J. Gehring, M. Auli, D. Grangier, D. Yarats, and Y. N. Dauphin, "Convolutional sequence to sequence learning," *arXiv:1705.03122*, May 2017, doi: 10.48550/arXiv.1705.03122.
- [14] S. Ye, R. Zhao, and X. Fang, "An ensemble learning method for dialect classification - IOPscience," *IOP Conference Series: Materials Science and Engineering*, vol. 569, no. 5, doi: 10.1088/1757-899X/569/5/052064/meta.
- [15] D. Han, Q. Liu, and W. Fan, "A new image classification method using CNN transfer learning and web data augmentation," *Expert Systems with Applications*, vol. 95, pp. 43–56, Apr. 2018, doi: 10.1016/j.eswa.2017.11.028.
- [16] L. Song *et al.*, "A deep multi-modal CNN for multi-instance multi-label image classification," *IEEE Transactions on Image Processing*, vol. 27, no. 12, pp. 6025–6038, Dec. 2018, doi: 10.1109/TIP.2018.2864920.
- [17] R. C. Çalik and M. F. Demirci, "Cifar-10 image classification with convolutional neural networks for embedded systems," *2018 IEEE/ACS 15th International Conference on Computer Systems and Applications (AICCSA)*, pp. 1–2, 2018, doi: 10.1109/AICCSA.2018.8612873.
- [18] N. Sharma, V. Jain, and A. Mishra, "An analysis of convolutional neural networks for image classification-sciencedirect," *Procedia Computer Science*, vol. 132, pp. 377–384, 2018, doi: 10.1016/j.procs.2018.05.198.
- [19] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, Jun. 2016, doi: 10.1109/CVPR.2016.90.
- [20] G. Huang, Z. Liu, L. V. D. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2261–2269, Jul. 2017, doi: 10.1109/CVPR.2017.243.
- [21] R. K. Srivastava, K. Greff, and J. Schmidhuber, "Training very deep networks," in *Advances in Neural Information Processing Systems*, vol. 28, 2015.
- [22] A. A. Ahmed and S. M. Darwish, "A meta-heuristic automatic CNN architecture design approach based on ensemble learning," *IEEE Access*, vol. 9, pp. 16975–16987, 2021, doi: 10.1109/ACCESS.2021.3054117.
- [23] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.
- [24] "CIFAR-10 PNGs in folders," [Online]. Available: <https://www.kaggle.com/datasets/swaroopkml/cifar10-pngs-in-folders> (accessed Jan. 10, 2022).
- [25] "Traffic sign dataset - classification," [Online]. Available: <https://www.kaggle.com/datasets/ahemateja19bec1025/traffic-sign-dataset-classification> (accessed Jan. 10, 2022).

BIOGRAPHIES OF AUTHORS



Retaj Matroud Jasim    was born in kut, Iraq, in 1995. She received a B.Sc. degree in computer engineering from AL-Imam AL Kadum University Collage, Wasit, 2017. She is currently studying M.Sc. in Computer Engineering at Al Iraqia University College of engineering, Iraq. She can be contacted at email: retaj.m.jasim@students.aliraqia.edu.iq.



Tayseer Salman Atia    is a professor at the Department of Computer Engineering, Al Iraqia University, Iraq. Where she has been a faculty member since 2012. From 2013-2014 she was the head of the computer-engineering department. From 2014-2015 she was the dean's assistant for scientific affairs. Tayseer graduated with the first class B.Sc. degree in computer science in 2004 and a M.Sc. in data security in 2007 from the University of Technology, Iraq. She completed her Ph.D. in computer science from Al Mosul University, Iraq. Her research interests are data security and artificial intelligence, especially computational intelligence techniques. She can be contacted at email: tayseer.salman@aliraqia.edu.iq.