Indonesian Journal of Electrical Engineering and Computer Science
Vol. 30, No. 1, April 2023, pp. 491~500
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v30.i1.pp491-500

An abbreviated review of deep learning-based image classification models

Zaman Talal Abbood[1], Mohammed Nasser Hussain Al-Turfi[1], Layth kamil Adday Almajmaie[2]
[1]College of Engineering, Al-Iraqia University, Baghdad, Iraq
[2]Department of Computer Engineering, University of Technology, Baghdad, Iraq

Article Info

Article history:

Received Oct 3, 2022
Revised Dec 7, 2022
Accepted Dec 13, 2022

Keywords:

Abbreviated review
Benchmark datasets
Deep learning algorithms
Image classification
Pre-trained convolutional neural networks

ABSTRACT

Image classification is an extensively researched sub-fields of computer vision implemented in face recognition, self-driving, medical image segmentation, biological identification, and others. Traditional models of image classification require manual construction of feature extraction techniques and classification accuracy which are closely associated with these utilized techniques. During the rapid progress of multimedia technologies, the number of images that require classification got bigger, and this led to making image classification more complicated, hence, the manual construction of feature extraction techniques consumes more time and provides lower accuracy. In the recent decade, deep learning-based models have appeared in various applications. These models hold the merits of an effective extraction of image features, low-weight features filtering, a large capacity for processing, and higher classification speed and accuracy. Thus, lots of researchers have attempted to utilize deep learning algorithms, especially convolutional neural networks (CNNs) for image classification. Therefore, this paper concentrates on providing an abbreviated review of deep learning-based image classification models, by covering the recently utilized deep learning algorithms, comparing various related works and benchmark datasets mentioned in this paper, and summarizing the fundamental analysis and discussion.

Corresponding Author:

Zaman Talal Abbood
College of Engineering, Al-Iraqia University
Baghdad, Iraq
Email: saadalbawi@gmail.com

1. INTRODUCTION

Image classification has continually been an active scientific research trend all over the world, and the appearance of artificial intelligence has encouraged the development of this domain [1]. In the last few years, the tasks of image classification were mostly carried out using conventional machine learning algorithms. Machine learning algorithms represent one of the rapidly increasing fields of artificial intelligence, carried considerable technological and economic progress [2]. Generally, machine learning-based models involve unsupervised and supervised learning, and both have exhibited outstanding performance in the tasks of image classification [3]. However, these models are not easy to exploit practically, are small in capacity, and are easy to have overfitting. Furthermore, even though the machine learning-based image classification models perform well in the tasks of two-class, their advancement extent in multiple classes is restricted. The variety of practical tasks is large, and the number of datasets is getting bigger and bigger, which needs models having a higher learning capacity. Additionally, conventional methods of feature extraction hold low effectiveness, insufficient extraction of features, and critical loss of resources. Accordingly, researchers started to utilize deep learning algorithms for image classification research [4].

In deep learning, a convolutional neural network (CNN) has been broadly implemented in image classification tasks. Dissimilar to the conventional image classification models, a CNN does not require image features adjustment and extraction artificially, since it auto-learns the close-related features of the image with the results of classification using datasets of training. Consequently, CNNs efficiently handle the issues of artificial extracting of features and the low accuracy of classification concerning conventional models [5]. The typical construction of CNN includes distinct convolutional, pooling, and fully connected layers demonstrated in Figure 1.

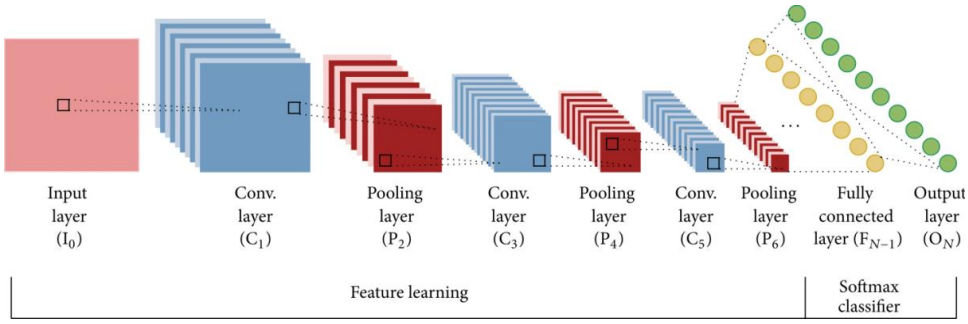

Figure 1. The typical construction of CNN

The two-dimensional image raw pixels could be input to CNNs directly. Then, these image pixels are convolved with multi-learned kernels utilizing shared weights. The main parameters of the convolutional layer are the size of the kernel, and the number and size of maps. After that, the pooling layer is utilized for reducing the image's size (minimizing the feature maps' resolution) while maintaining the included information. The outcome of the pooling layer is obtained via mean, maximum, and stochastic activation, concerning mean pooling, max pooling, and stochastic pooling, for non-overlapping rectangular areas. The pairs of convolutional and pooling layers form the phase of feature extraction. Afterward, these obtained features are weighted and merged into a fully connected layer (one or more layers), and this process indicates the phase of classification. Lastly, in the task of classification, the output layer holds one neuron for each category [6]. When the utilized activation function is a soft max, then the output of each neuron indicates the probability of the posterior category [7]. CNNs are able to filter input noise, collect crucial features, and achieve higher model accuracy [8]. However, the low performance and model over fitting represent common issues in the tasks of image classification.

Dissimilar to other existing review researches, the fundamental contribution of this paper can be outlined as follows: firstly; present the recent and widely utilized pre-trained CNNs in image classification tasks. Secondly; compare the performance of various deep learning-based image classification models and benchmark datasets. Finally, provide insightful analysis and discussion.

2. PRE-TRAINED CONVOLUTIONAL NEURAL NETWORKS

Since the difficulty of classifying images increases. The models of pre-trained CNNs require increasing the number of layers, and this may lead to an extreme increase in model training difficulty. This section presents the fundamental concepts of pre-trained CNNs that are commonly utilized in image classification models.

2.1. AlexNet network

AlexNet was presented via Krizhevsky *et al.* [9] in 2012. AlexNet is a large and complex neural network of 650 thousand neurons and sixty-million parameters, as shown in Figure 2, it includes five convolutional layers (some of these layers are accompanied by max-pooling layers for reducing features), and three fully-connected layers (these layers are accompanied by dropout layers for reducing joint adaptation among neurons, avoiding over fitting, and improving the robustness and generalization) with a last thousand-category soft max. It represents one of the most common architectures of CNNs for image or object classification and in general, it can be utilized on large datasets, like ImageNet which includes about fifty-million images [10]. Convolution is capable of reducing the network complexity via parameter sharing, and automatically learning features from the images of training. In this network, rectified linear unit (ReLU) is utilized as an activation function for accelerating the convergence of the model and reducing the gradient disappearing. AlexNet with the ReLU function converges more rapidly than the conventional activation functions since the ReLU's gradient is permanently one, when the input is more than zero. In neighboring fully-connected layers, the neurons are directly connected, and the soft max function works on activating neurons in the (0,1) range via restricting the output [11].

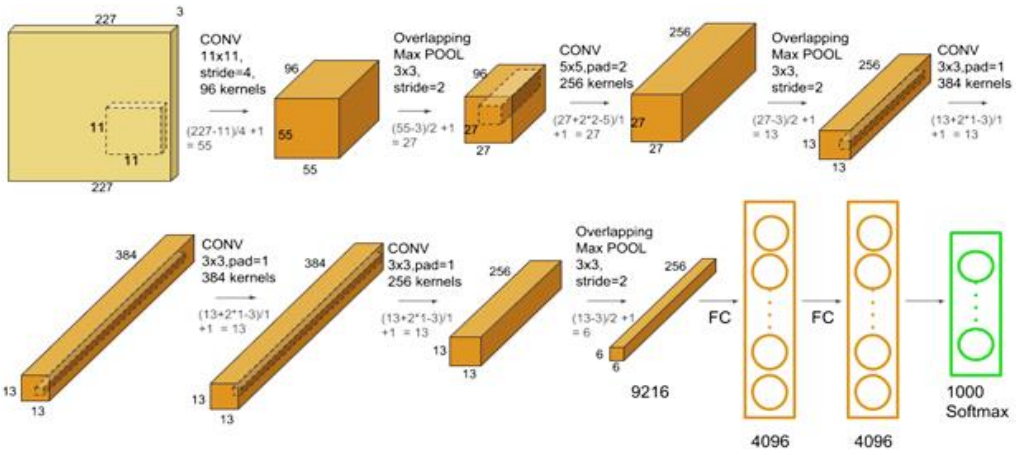

Figure 2. AlexNet network architecture

2.2. VGGNet

VGG was presented via Simonyan and Zisserman [12] in 2014. This pre-trained model of CNN was trained using an ILSVRC ImageNet dataset, including 1.3 million images. VGG network is comparable to the AlexNet network, which utilizes the construction of the convolution region succeeded by the fully connected region. VGG model utilized a number of successive convolutional layers accompanied by max-pooling layers. These convolutional layers keep the width and height of the input unaltered, whereas the pooling layers work on halving it. Lots of convolution filters (of size 3×3) were utilized in the VGG network that can guarantee the increase of network depth, and reduce the parameters of the model [13], [14]. There are various models of the VGG networks, such as VGG16 and VGG19. Figure 3 shows that VGG16 and VGG19 include a series of five blocks, two fully connected layers, and one output layer. Figure 3(a) shows that VGG16 includes 16 layers, thirteen convolutional and three fully connected layers, and utilizes the activation function named ReLU. While Figure 3(b) shows that VGG19 includes 19 layers which have provided better performance than VGG16 [15].

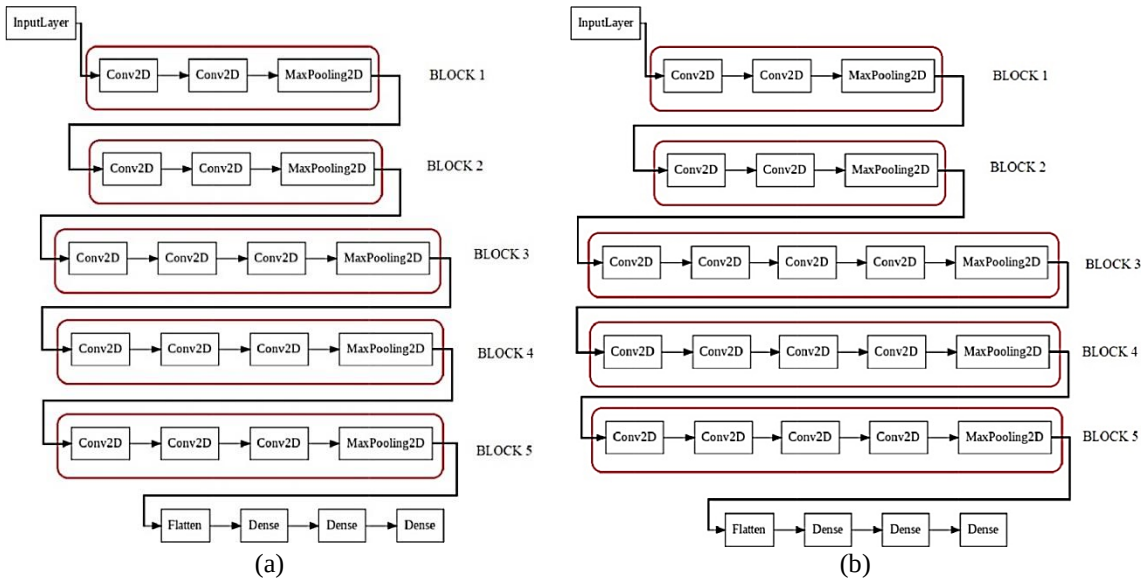

Figure 3. Layers in (a) VGG16 and (b) VGG19 architectures

2.3. InceptionV1-V4

GoogLeNet was presented via Szegedy *et al.* [16] in 2014. This model introduced the Inception model concept, and in successive years, several researchers worked on improving the performance of the Inception model. GoogLeNet includes twenty-two layers and holds approximately six million parameters. The fundamental module of GoogLeNet is the module of Inception. Figure 4 illustrates the various modules of Inception. Figure 4(a) explains the fundamental module which includes four branches in parallel. Several convolutional layers of various

sizes have been utilized in the first three branches for extracting information of various spatial sizes. Within these layers, convolution (of size 1×1) is capable of reducing the number of channels and compressing information to minimize the complexity of the model. In the final branch, the max-pooling was added to minimize the resolution, accompanied by convolution (of size 1×1) to set the depth following the pooling. Generally, the distinctive design of this module gets the network width bigger and increases the capability of adaptation to various resolutions and scales. Figure 4(b) and Figure 4(c) explain the modules of InceptionV2 and InceptionV3, respectively [17].

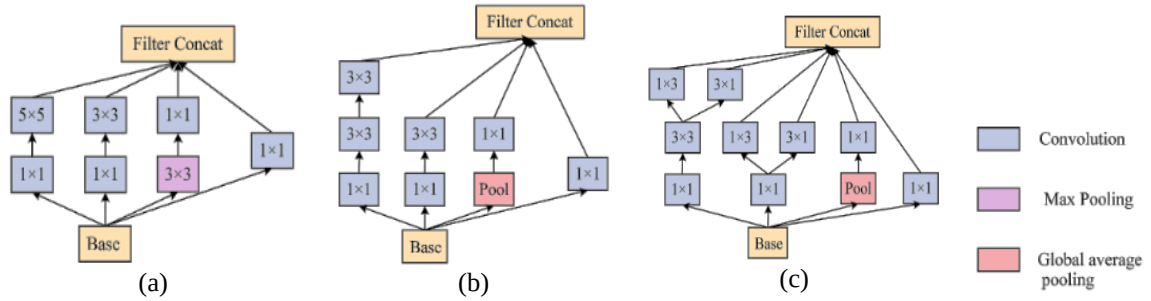

Figure 4. The module of (a) InceptionV1, (b) InceptionV2, and (c) InceptionV3

In comparison to InceptionV1, the improvements to inceptionV2 [18] are represented in the use of smaller convolution and batch normalization. Hence, the 3×3 convolutions are utilized rather than the 5×5 convolutions, and this led to increasing the speed of computation (i.e, decreasing computational time). Batch normalization is utilized for making the model more stable and faster via normalizing the variances and means of input layers [17].

Inception V3 [19] fundamentally concentrated on utilizing less power of computation via altering the previous architectures of Inception. The improvements to inceptionV3 are represented in the use of factorized (smaller and asymmetric) convolutions and reducing the grid size (using pooling processes). Factorized convolutions work on minimizing the number of parameters within a network by 33%, and this leads to minimizing the efficiency of computation. Here, the smaller convolutions are utilized rather than larger convolutions to obtain fast training, and 1×3 convolution accompanied by 3×1 convolution is utilized rather than 3×3 convolution to minimize the number of parameters.

The fundamental objective of Inception V4 [20] is to minimize the Inception V3 complexity that built a consolidated selection for every block of Inception. Figure 5 explains that the blocks of Inception encompass modules of Inception, Figure 5(a) explains the module of Inception-A, Figure 5(b) explains the module of Inception-B, and Figure 5(c) explains the module of Inception-C. The architecture of these models utilizes optimization of memory on back-propagation for reducing the requirement of memory.

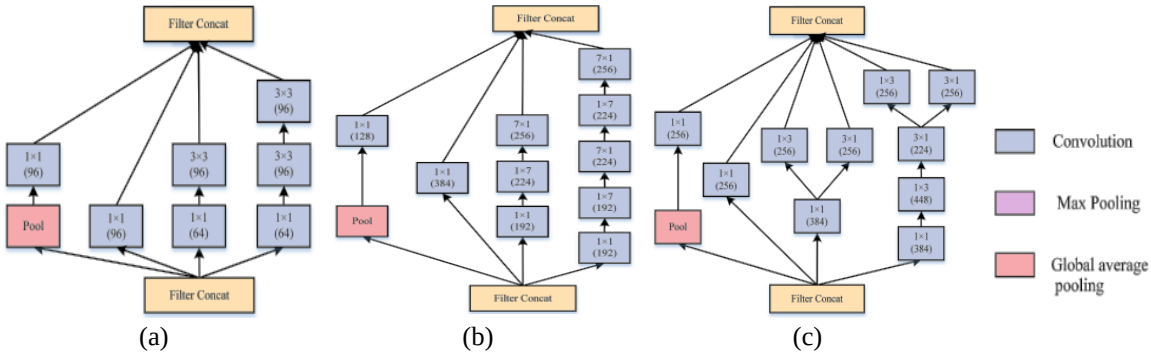

Figure 5. The blocks of InceptionV4, (a) Inception-A, (b) Inception-B, and (c) Inception-C

2.4. ResNet and DenseNet

ResNet was presented via He *et al.* [21] for training the considerably deeper networks easily and accurately. This model was evaluated using the ImageNet dataset with a depth of reaching 152-layer. It is plain to be optimized, however, the networks that easily stack layers demonstrate a higher error of training when the depth of networks gets bigger. Compared with the former models, ResNet is easily capable of gaining higher accuracy concerning considerably increased depth [22].

DenseNet was presented via Huang *et al.* [23]. It follows the same direction as ResNet in facing the gradient degradation and disappearance issue (ResNet only addressed some solutions to this issue). In the conventional convolution, the connection is established between every layer and the following layer. While in DenseNet, every layer is connected to another layer in a fashion of feed-forward. This manner enables every layer to directly reach the gradients from the function of loss and the input signal, causing implied deep supervision. Figure 6 demonstrates the content of the DenseNet block in which "x_0" indicates an image that is crossed over a convolutional network of "L" layers. Each layer carries out a nonlinear transformation "H_L" that represents operations like ReLU, batch normalization, convolution, and pooling [24].

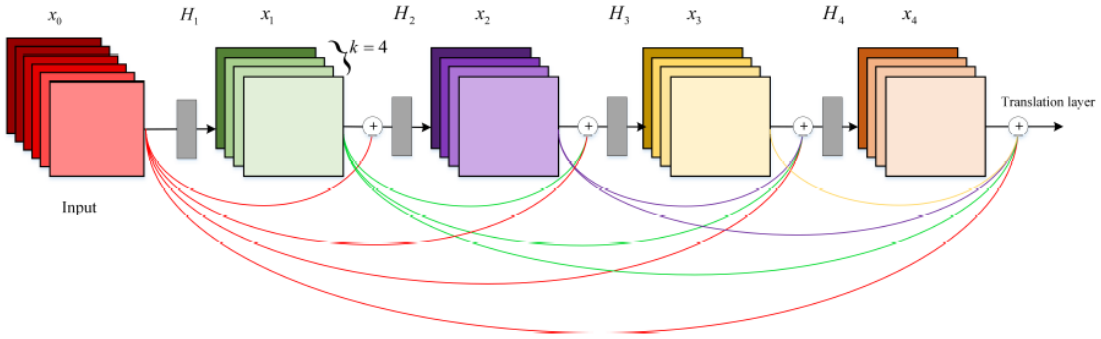

Figure 6. DenseNet block of five layers, with the rate of growth (k=4)

3. DEEP LEARNING BASED IMAGE CLASSIFICATION MODELS

Before presenting diverse image or object classification models, it is important to perceive various but similar tasks of computer vision such as object localization and object detection, for avoiding any unsureness that may appear later. Image or object classification appoints a class label for the entire image, and object localization puts a bounding box on every object existing within an image [25]. While object detection represents a composite of object categorization and localization tasks, hence, it appoints label for every object of interest after putting a bounding box [26]. In this section, we review some of the related approaches and methods using deep learning that is utilized for object or image classification models, some of them are described briefly.

Zhang *et al.* [27] presented a model dependent on an adapted extreme learning machine. In this CNN model, a well-trained network with 5 convolutional and 3 fully connected layers was implemented to obtain the deep features from the utilized datasets. Extreme learning machines, support vector machine (SVM), and nearest neighbor (NN) were utilized as classifiers for object categorization over convolutional activation deep feature representation. In particular, several standard object categorization datasets were adopted for evaluating different classifiers. Experiments demonstrate that this proposed model was effective in categorization tasks. Zhou *et al.* [28] combined CNNs with biomimetic pattern recognition to present a new model of image classification that successfully addressed the problem of soft max limited capacity. This model was evaluated on CIFAR-10 and MNIST datasets, and the obtained results of classification accuracies were 87.11% and 99.01%, respectively. Gu *et al.* [29] presented an improved VGG16-based image classification model using image data augmentation and hyper-parameter tuning for reducing model over fitting and increasing model performance. This presented model was evaluated using the CIFAR-10 dataset, and the obtained accuracy was 96%.

Shakarami and Tarrah [30] proposed a developed method for image categorization using several algorithms of machine and deep learning. This presented method included several stages; Firstly the images were input and resized, and every image was transmitted to an extractor of deep features (developed CNN-AlexNet), and handcrafted descriptors like local binary pattern descriptor and the histogram of oriented gradients descriptor concurrently. Then, the algorithm of principle component analysis (PCA) was utilized for reducing the produced features dimensions via the histogram of oriented gradients. After that, the vectors of deep features and Handcrafted-PCA features were merged. Finally, the classification was done utilizing three algorithms of machine learning (random forest (RF), SVM, and k-nearest neighbors (K-NN)). In all experiments, the proposed method utilizing RF classifier conducted on the Caltech-101 object dataset was more accurate than other classifiers with 89.74% of accuracy.

Qin *et al.* [31] presented an object categorization approach depending on pre-trained CNNs (InceptionV3 and DenseNet201). This approach included several stages; Firstly, this method utilized the appropriately enlarged images as inputs to each network for improving the accuracy of classification. Secondly, the last modules were substituted by a block of inverted residuals with fewer parameters for minimizing the cost of computation. Finally, the convolution layer, Batch normalization layer, rectified linear unit (ReLU),

average pooling, and the function of soft max were supplemented to build the networks. This proposed method was substantially evaluated on various renowned standard datasets, and the obtained results demonstrated that this approach showed promising performance in object categorization and minimized the parameters of the network and the cost of computation. However, it is optimistic to obtain a further effective and lightweight convolution approach that can efficiently minimize the network parameters.

Rashid *et al.* [32] proposed a supportable architecture of deep learning that uses multiple layers of deep feature combining and selection to obtain accurate image categorization. This approach comprised several stages; Firstly, by utilizing two deep learning architectures (VGG19 and InceptionV3), the features were extracted on the basis of transfer learning. Secondly, the combination for the whole extracted vectors of features was conducted via the approach of parallel maximum covariance. Thirdly, the preferable features were chosen by utilizing the method of multi-logistic regression controlled entropy variances. Finally, in the classification stage, the classifier of the ensemble subspace discriminant was utilized. The experimental procedure was performed utilizing several publicly available datasets, and over the Caltech-101 dataset, the obtained accuracy was 95.5%. The fundamental limitation of this approach was the features' quality, via utilizing low-quality images, it is not feasible to obtain robust features.

Akshaya and Kala [33] proposed a CNN-based image/object classification method. This proposed deep CNN included six convolutional layers (supplied with convolution, exponential linear unit (ELU), and batch normalization functions), three max-pooling layers with a dropout layer placed after each max-pooling layer to minimize the network over fitting, one flatten layer, and one output (fully connected) layer. This model was conducted on the actually a complex dataset named CIFAR10 image dataset and the obtained result of accuracy was 89.87%. However, this proposed method requires more improvement to obtain accurate results of categorization.

Zhu *et al.* [34] proposed a unified innovative framework, called attention-aware perceptual enhancement networks that integrated an attention mechanism and perceptual improvement in a comprehensive way for categorizing low-resolution images. This framework included five fundamental components; Firstly, the super-resolved image was created from their corresponding low-resolution image using the perceptual improvement network. Secondly, a map of attention was generated using the attention creation network. This map is capable of indicating informative regions of super-resolved images and promoting semantic information. Thirdly, the feature extraction network obtained a 1D-feature representation. Fourthly, the one-dimensional feature was rectified with the addition of element-wise using the network of feature rectification. Finally, the classification process was done using (VGG-16 and ResNet-34) classifiers. Extensive experiments were performed on three datasets and the best obtainable accuracy was 95.94 achieved by ResNet-34 based framework over the Stanford Dogs datasets. However, this network was fundamentally built for accurate categorization, and the super-resolution images that were created did not fulfill the expectations of human visuals. This is fundamentally produced via the large perceptual loss. Thus, there is a requirement for alleviating this issue to provide more realistic perceptually images.

Basha *et al.* [35] attempted to automatically tune several CNN models (DenseNet-121, ResNet-50, and VGG-16) to construct appropriate models for image categorization task. In order to accomplish this goal, the following objectives were achieved; Firstly, the layer of soft max for the CNN models was dropped via substituting it with another layer of new soft max holding the neurons equivalent to the number of categories in the given datasets. Then, the layers of CNN (from initial to final layers) were automatically tuned using Bayesian optimization. The experiments were performed on several standard datasets (Stanford Dogs, CalTech-256, and CalTech-101). The obtained results of categorization using the presented auto-tuned DenseNet-model outperformed the benchmark model of transfer learning via realizing 84.67%, 86.54%, and 95.92% accuracy through Stanford Dogs, CalTech-256, CalTech-101, respectively.

Zhang *et al.* [36] developed Hierarchical Bilinear CNN model by combining convolutional networks with multitask learning on the hierarchical visual structures. Hierarchical bilinear CNN utilized VGG16 for constructing a network of inner output branches. It combined the CNNs' hierarchy with prior object categories structure for strengthening the capability of classification, it leveraged the module of bilinear for extracting more selective information, and it also utilized the label trees as inner guides for boosting model performance. This developed model was evaluated using CIFAR-10 and CIFAR-100 datasets, and the obtained values of accuracy were 91.75% and 66.03%, respectively.

Jung *et al.* [37] proposed an active method of weighted mapping for improving the ResNet performance, by changing the weight values in accordance with the input image class. In this method, the ablation test (using the method of linear discriminant analysis) was utilized for proving that the deduced weight values are capable of classifying visual objects. This proposed method was implemented using the basic DenseNet and ResNet, and validated using CIFAR-10 and CIFAR-100 datasets.

Hassanzadeh *et al.* [38] designed an evolved deep CNN-based image classification model. This network is made up of several convolutional blocks for extracting features, and a layer of classification with an output layer for generating the last prediction. In this proposed model, a genetic algorithm (GA) was utilized to evolve the convolutional blocks by finding the optimal numbers of blocks, convolutional layers, filters, pooling, dropouts, batch

normalization, and activation functions. GA was also utilized to evolve the classification layer by finding the optimal classification layer kind, dropouts, number of nodes, batch normalization, and activation function. This evolved model was trained using the CIFAR10 and MNIST datasets, and concerning the CIFAR10 dataset the optimal obtained values of accuracy, F1 score, recall, and precision were 0.98%, 0.95%, 0.95%, and 0.95%, respectively.

Xu *et al.* [39] presented a selective kernel network model for enhancing the effect of feature training in the networks of deep learning. This model was utilized for learning the image data characteristics. And a classifier of the Gaussian process with a function of a multi-layers convolution kernel was utilized for performing image classification. This model was trained using the CIFAR10 and MNIST datasets, and obtained values of accuracy were 94.12% and 99.35%, respectively.

4. PERFORMANCE ANALYSIS AND DISCUSSION

4.1. Image classification datasets

There are various commonly utilized benchmark image (or object) classification datasets. Table 1 shows some of these publicly available datasets with the rising difficulty of categorization; Caltech-101, Caltech-256 which represents an improvement to its former with new characteristics like bigger class sizes and new clutter classes, Flower-102, CIFAR-10, CIFAR-100 categorizes get into 20 super-classes of 5 classes each, Stanford Dogs, and Food-101.

Table 1. Publicly available image classification datasets

Dataset	No. of Classes	No. of Images	Year	Description
Caltech-101 [40]	102	9,144	2003	101 common object categories; The size of each image is roughly 300×200 pixels.
Caltech-256 [41]	257	30,307	2006	An extension for Caltech-101
Flower-102 [42]	103	8,189 Each class includes (40 to 250)	2008	Dataset of flower recognition including 103-flower classes prevalent in the UK
CIFAR-10 [43]	10	Each class includes 6,000	2009	32×32 pixel tiny color images
CIFAR-100 [43]	100	Each class includes 600	2009	More challenging for object classification, 32×32 RGB
Stanford Dogs [44]	120	22,000	2011	One object in each image dataset for recognizing the breed of dogs
MNIST [45]	10	70,000	2011	Dataset of handwritten numbers from 0 to 9 written by 250 individuals, 28×28 gray scale image
Food-101 [46]	101	Each category includes 1,000	2014	Categorized food images

4.2. Evaluation measures

For the purpose of diving within the common measures utilized for evaluating the deep learning-based image classification models, a number of fundamental descriptions require to be initially established:

- Firstly; true negative (TNegative) indicates the right prediction of the condition's absence.
- Secondly; false negative (FNegative) indicates the non-right prediction of the condition's absence when it exists.
- Thirdly; true positive (TPositive) indicates the right prediction of the current condition.
- Fourthly; false positive (FPositive) indicates the non-right prediction of the condition when it does not exist.

Having described the previous terms, it is possible to describe the recall and precision measures. The measure of Sensitivity or Recall (Rec) indicates the number of positive conditions that have been rightly predicted by the model. A higher value of sensitivity demonstrates a lower level of non-detected positive conditions. This measure is given as follows [47]:

$$Rec = \frac{T_{Positive}}{T_{Positive}+F_{Negative}} \quad (1)$$

while precision (*Pre*) indicates the accuracy measure of the positive condition predictions. A higher value of precision demonstrates a lower level of non-right predictions ($F_{Positive}$). This measure is given as follows [48]:

$$Pre = \frac{T_{Positive}}{T_{Positive}+F_{Positive}} \quad (2)$$

furthermore, there are other significant measures that are utilized for models' evaluation which are Specificity, Accuracy, and F1-Score. The measure of Specificity (*Spe*) indicates the rate of $T_{Negative}$, and the measure of Accuracy (*ACC*) indicates the ratio of rightly predicted instances to the entire existing instances. While the measure of F1-Score ($F1_S$) works on combining the values of Recall and Precision measures into harmonic mean. These measures are given as follows [49], [50]:

$$Spe = \frac{T_{Negative}}{T_{Negative}+F_{Positive}} \quad (3)$$

$$Acc = \frac{T_{Negative}+T_{Positive}}{F_{Negative}+T_{Negative}+F_{Positive}+T_{Positive}} \quad (4)$$

$$F1_S = 2\frac{Rec \times Pre}{Rec+Pre} \quad (5)$$

4.3. Comparison of image classification models

There are lots of researchers who have attempted to utilize deep learning algorithms, especially CNNs for image classification. Therefore, a comparison of the categorization performance is presented concerning some related deep learning-based approaches that are utilized for object or image classification models. Table 2 illustrates a comparison of the categorization performance obtained with state-of-the-art models.

Table 2. Comparison of the categorization performance obtained with state-of-the-art methods

Authors Name, Ref., Year	Utilized Datasets	Feature Extraction Methods	Classification Methods	Accuracy %
Zhang *et al.* [27], 2017	Caltech 256	Well-trained CNN	NN, SVM, and Extreme Learning Machine	76.2, 83.2, and 85.0
Zhou *et al.* [28], 2017	CIFAR-10 MNIST	CNNs	Biomimetic Pattern Recognition	87.11 99.01
Gu *et al.* [29], 2019	CIFAR-10	VGG16	Soft max	96%
Shakarami and Tarrah [30], 2020	Caltech-101	AlexNet CNN and (Local Binary Pattern and Histogram of Oriented Gradients)-PCA	RF	89.74
Qin *et al.* [31], 2020	Caltech-101, Caltech-256 Flower-102	Improved InceptionV3 Improved DenseNet201 Improved InceptionV3 Improved DenseNet201	Soft max	94.18, 86.62 94.31, 86.24 96.58 97.32
Rashid *et al.* [32], 2020	Caltech-101 CIFAR-100	VGG19 and InceptionV3 + Multi Logistic Regression controlled Entropy-Variances method	Ensemble Sub-space Discriminant classifier	95.5 68.80
Akshaya and Kala, [33], 2020	CIFAR-10	Deep CNN model	Soft max	89.87
Zhu *et al.* [34], 2020	Food-101 Caltech-256 Stanford Dogs	Attention-aware Perceptual Enhancement Networks	VGG-16, ResNet-34	93.43, 94.06 90.21, 92.99 93.17, 95.94
Basha *et al.* [35], 2021	Caltech-101 Caltech-256 Stanford Dogs	AutoTune (Bayesian Optimization) + DenseNet-121	Soft max	95.92 86.54 84.67
Zhang *et al.* [36], 2021	CIFAR-10 CIFAR-100	Hierarchical Bilinear CNN model using VGG16	Soft max	91.75 66.03
Jung *et al.* [37], 2021	CIFAR-10 CIFAR-100	DenseNet and ResNet	Sigmoid	-
Hassanzadeh *et al.* [38], 2022	CIFAR-10	An Evolved Deep CNN model	Soft max	98
Xu *et al.* [39], 2022	CIFAR-10 MNIST	Selective Kernel Network Model	Gaussian Process with a Function of Multi-layers Convolution Kernel	94.12 99.35

5. CONCLUSION

Generally, image classification works on describing the entire image via extracting features manually or utilizing methods of feature learning, and subsequently using the classifier to specify the category of object. Accordingly, the process of extracting features from images represents a significant issue. However, the appearance of deep learning algorithms holds a succession of innovations in the image classification field and presented an outstanding performance on visual tasks. The substantial achievement of deep CNNs is assigned to their strong capability of feature learning. Dissimilar to the conventional models of image classification, the deep CNNs-based image classification models provide a comprehensive process of learning, in which the input is an image, and the process of training and prediction is performed within the neural network, finally, the outcomes are produced. These models give up the methods of manual image features extraction and smash the deadlock of conventional image classification models. The accessibility of public datasets and codes is necessary for any paper concerning deep learning-based image classification models in order that researchers are capable of deploying and testing improved models in future directions.

REFERENCES

[1] N. Wambugu *et al.*, "Hyperspectral image classification on insufficient-sample and feature learning using deep neural networks: A review", *International Journal of Applied Earth Observation and Geoinformation*, vol. 105, p. 102603, 2021, doi: 10.1016/j.jag.2021.102603.

[2] B. Li, "Hearing loss classification via AlexNet and extreme learning machine," *International Journal of Cognitive Computing in Engineering*, vol. 2, pp. 144-153, 2021, doi: 10.1016/j.ijcce.2021.09.002.

[3] H. Q. Flayyih, J. Waleed and S. Albawi, "A systematic mapping study on brain tumors recognition based on machine learning algorithms," *2020 3rd International Conference on Engineering Technology and its Applications (IICETA)*, 2020, pp. 191-197, doi: 10.1109/IICETA50496.2020.9318886.

[4] B. J. Khadhim, Q. K. Kadhim, W. K. Shams, S. T. Ahmed, and W. A.W. Alsiadi, "Diagnose COVID-19 by using hybrid CNN-RNN for chest X-ray," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 29, no. 2, pp. 852-860, 2023, doi: 10.11591/ijeecs.v29.i2.pp852-860.

[5] H. Lin and J. J. Yang, "Ensemble cross-stage partial attention network for image classification," *IET Image Processing*, vol. 16, no. 1, pp. 102-112, 2022, doi: 10.1049/ipr2.12335.

[6] J. Waleed, T. Abbas, and T. M. Hasan, "Facemask wearing detection based on deep CNN to control COVID-19 transmission," *Muthanna International Conference on Engineering Science and Technology*, 2022, pp. 158-161, doi: 10.1109/MICEST54286.2022.9790197.

[7] M. S and E. Karthikeyan, "Classification of image using deep neural networks and soft max classifier with CIFAR datasets," *2022 6th International Conference on Intelligent Computing and Control Systems*, 2022, pp. 1132-1135, doi: 10.1109/ICICCS53718.2022.9788359.

[8] G. C. Sekhar and A. Rajendran, "A secure framework of blockchain technology using CNN long short-term memory hybrid deep learning model," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 28, no. 3, pp. 1786-1795, 2022, doi: 10.11591/ijeecs.v28.i3.pp1786-1795.

[9] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, pp. 1097-1105, 2012.

[10] T. Lu, B. Han, and F. Yu, "Detection and classification of marine mammal sounds using AlexNet with transfer learning," *Ecological Informatics*, vol. 62, p. 101277, 2021, doi: 10.1016/j.ecoinf.2021.101277.

[11] T. Lu, F. Yu, C. Xue, and B. Han, "Identification, classification, and quantification of three physical mechanisms in oil-in-water emulsions using AlexNet with transfer learning," *Journal of Food Engineering*, vol. 288, p. 110220, 2021, doi: 10.1016/j.jfoodeng.2020.110220.

[12] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[13] A. V. Ikechukwu, S. Murali, R. Deepu, and R. C. Shivamurthy, "ResNet-50 vs VGG-19 vs training from scratch: a comparative analysis of the segmentation and classification of Pneumonia from chest X-ray images," *Global Transitions Proceedings*, vol. 2, no. 2, pp. 375-381, 2021, doi: 10.1016/j.gltp.2021.08.027.

[14] A. S. Hatem, M. S. Altememe, and M. A. Fadhel, "Identifying corn leaves diseases by extensive use of transfer learning: a comparative study," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 29, no. 2, pp. 1030-1038, 2023, doi: 10.11591/ijeecs.v29.i2.pp1030-1038.

[15] N. Begum and M. K. Hazarika, "Maturity detection of tomatoes using transfer learning," *Measurement: Food*, vol. 7, p. 100038, 2022, doi: 10.1016/j.meafoo.2022.100038.

[16] C. Szegedy *et al.*, "Going deeper with convolutions," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, 2015, pp. 1-9, doi: 10.1109/CVPR.2015.7298594.

[17] R. Suresh and N. Keshava, "A survey of popular image and text analysis techniques," *2019 4th International Conference on Computational Systems and Information Technology for Sustainable Solution*, 2019, pp. 1-8, doi: 10.1109/CSITSS47250.2019.9031023.

[18] S. Ioffe and C. Szegedy, "Batch normalization: accelerating deep network training by reducing internal covariate shift," *Proceedings of the 32nd International Conference on Machine Learning*, PMLR 37, pp. 448-456, 2015, doi: 10.5555/3045118.3045167.

[19] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna, "Rethinking the inception architecture for computer vision," 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 2818-2826, doi: 10.1109/CVPR.2016.308.

[20] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-ResNet and the impact of residual connections on learning," San Francisco, CA, USA, 2016, doi: 10.1609/aaai.v31i1.11231.

[21] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 770-778, doi: 10.1109/CVPR.2016.90.

[22] T. Sun, S. Ding, and L. Guo, "Low-degree term first in ResNet, its variants and the whole neural network family," *Neural Networks*, vol. 148, pp. 155-165, 2022, doi: 10.1016/j.neunet.2022.01.012.

[23] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 2017, doi: 10.1109/CVPR.2017.243.

[24] K. Wang, P. Jiang, J. Meng, and X. Jiang, "Attention-based densenet for pneumonia classification," *IRBM*, vol. 43, no. 5, pp. 479-485, 2022, doi: 10.1016/j.irbm.2021.12.004.

[25] R. A. Sobbahi and J. Tekli, "Comparing deep learning models for low-light natural scene image enhancement and their impact on object detection and classification: Overview, empirical evaluation, and challenges," *Signal Processing: Image Communication*, vol. 109, p. 116848, 2022, doi: 10.1016/j.image.2022.116848.

[26] A. Salari, A. Djavadifar, X. Liu, and H. Najjaran, "Object recognition datasets and challenges: a review," *Neurocomputing*, 2022, doi: 10.1016/j.neucom.2022.01.022.

[27] L. Zhang, Z. He, and Y. Liu, "Deep object recognition across domains based on adaptive extreme learning machine," *Neurocomputing*, vol. 239, pp. 194-203, 2017, doi: 10.1016/j.neucom.2017.02.016.

[28] L. Zhou, Q. Li, G. Huo, and Y. Zhou, "Image classification using biomimetic pattern recognition with convolutional neural networks features," *Computational Intelligence and Neuroscience*, vol. 2017, pp. 1-12, 2017, doi: 10.1155/2017/3792805.

[29] S. Gu, M. Pednekar, and R. Slater, "Improve image classification using data augmentation and neural networks," *SMU Data Science Review*, vol. 2, no. 2, 2019.

[30] A. Shakarami and H. Tarrah, "An efficient image descriptor for image classification and CBIR," *Optik*, vol. 214, p. 164833, 2020, doi: 10.1016/j.ijleo.2020.164833.

[31] J. Qin, W. Pan, X. Xiang, Y. Tan, and G. Hou, "A biological image classification method based on improved CNN," *Ecological Informatics*, vol. 58, 101093, 2020, doi: 10.1016/j.ecoinf.2020.101093.

[32] M. Rashid *et al.*, "A sustainable deep learning framework for object recognition using multi-layers deep features fusion and selection," *Sustainability*, vol. 12, no. 12, p. 5037, 2020, doi: 10.3390/su12125037.

[33] B. Akshaya and M. T. Kala, "Convolutional neural network based image classification and new class detection," *2020 International Conference on Power, Instrumentation, Control and Computing (PICC)*, 2020, pp. 1-6, doi: 10.1109/PICC51425.2020.9362375.

[34] X. Zhu, Z. Li, X. Li, S. Li, and F. Dai, "Attention-aware perceptual enhancement nets for low-resolution image classification," *Information Sciences*, vol. 515, pp. 233-247, 2020, doi: 10.1016/j.ins.2019.12.013.
[35] S. H. S. Basha, S. K. Vinakota, V. Pulabaigari, S. Mukherjee, and S. R. Dubey, "AutoTune: automatically tuning convolutional neural networks for improved transfer learning," *Neural Networks*, vol. 133, pp. 112-122, 2021, doi: 10.1016/j.neunet.2020.10.009.
[36] X. Zhang *et al.*, "Hierarchical bilinear convolutional neural network for image classification," *IET Computer Vision*, vol. 15, no. 3, pp. 197-207, 2021, doi: 10.1049/cvi2.12023.
[37] H. Jung, R. Lee, S.-H. Lee, and W. Hwang, "Active weighted mapping-based residual convolutional neural network for image classification," *Multimedia Tools and Applications*, vol. 80, pp. 33139-33153, 2021, doi: 10.1007/s11042-021-11538-z.
[38] T. Hassanzadeh, D. Essam, and R. Sarker, "EvoDCNN: An evolutionary deep convolutional neural network for image classification," Neurocomputing, vol. 488, pp. 271-283, 2022, doi: 10.1016/j.neucom.2022.02.003.
[39] L. Xu *et al.*, "Gaussian process image classification based on multi-layer convolution kernel function," *Neurocomputing*, vol. 480, pp. 99-109, 2022, doi: 10.1016/j.neucom.2022.01.048.
[40] L. Fei-Fei, R. Fergus, and P. Perona, "Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories," *2004 Conference on Computer Vision and Pattern Recognition Workshop*, 2004, pp. 178-178, doi: 10.1109/CVPR.2004.383.
[41] G. Griffin, A. Holub, and P. Perona, "Caltech-256 object category dataset," 2007.
[42] M.-E. Nilsback and A. Zisserman, "Automated flower classification over a large number of classes," *2008 Sixth Indian Conference on Computer Vision, Graphics & Image Processing*, 2008, pp. 722-729, doi: 10.1109/ICVGIP.2008.47.
[43] A. Krizhevsky and G. Hinton, "Learning multiple layers of features from tiny images," University of Toronto: Toronto, Canada, 2009.
[44] A. Khosla, N. Jayadevaprakash, B. Yao, and F.-F. Li, "Novel dataset for finegrained image categorization: stanford dogs," *in: Proc. CVPR Workshop on Fine-Grained Visual Categorization (FGVC)*, vol. 2, 2011.
[45] G. M. De León, A. Moreno-Báez, R. Magallanes-Quintanar, and R. D. V.-Cepeda, "Assessment in subsets of MNIST handwritten digits and their effect in the recognition rate," *Journal of Pattern Recognition Research*, vol. 6, no. 2, pp. 244-252, 2011, doi: 10.13176/11.348.
[46] L. Bossard, M. Guillaumin, and L. V. Gool, "Food-101–mining discriminative components with random forests," *in: European Conference on Computer Vision*, ECCV 2014, 2014, pp. 446–461, doi: 10.1007/978-3-319-10599-4_29.
[47] J. Waleed, S. Albawi, H. Q. Flayyih, and A. Alkhayyat, "An effective and accurate CNN model for detecting tomato leaves diseases," *2021 4th International Iraqi Conference on Engineering Technology and Their Applications (IICETA)*, 2021, pp. 33-37, doi: 10.1109/IICETA51758.2021.9717816.
[48] M. Mehmood, A. Shahzad, B. Zafar, A. Shabbir, and N. Ali, "Remote sensing image classification: a comprehensive review and applications," *Mathematical Problems in Engineering*, vol. 2022, pp. 1-24, 2022, doi: 10.1155/2022/5880959.
[49] T. M. Hasan, S. D. Mohammed, and J. Waleed, "Development of breast cancer diagnosis system based on fuzzy logic and probabilistic neural network," *Eastern-European Journal of Enterprise Technologies*, vol. 4, no. 9, 2020, doi: 10.15587/1729-4061.2020.202820.
[50] N. F. B. A. Halim, R. A. B. Ramlee, M. Z. B. Mas'ud, and A. Jamaludin, "Enhancement of automatic classification of arcus senilis-nonarcus senilis using convolutional neural network," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 28, no. 1, pp. 209-219, 2022, doi: 10.11591/ijeecs.v28.i1.pp209-219.

BIOGRAPHIES OF AUTHORS

Zaman Talal Abboud his Bachelor's degree in Computer Engineering from Al-Rafidain University College, Iraq 2012-2016 and works as an employee in the Ministry of Health in the Engineering Department. In 2020, she was accepted into the Iraqi University/College of Engineering to study for a master's degree, and she is now in the research stage. She can be contacted at email: zamantaa89@gmail.com, saadalbawi@gmail.com.

Mohammed Nasser Hussain Al-Turfi received BSc. Degree in Electrical Engineering–College of Engineering Al Mustansirya University (July, 1997). M.Sc. Degree in Computer and Control–Electrical Engineering Department College of Engineering-University of Baghdad, (October, 2000). Ph.D Degree in Computer communication and Control-Electrical Engineering Department-College of Engineering-University of Baghdad (June 2004). Certified and recommended by Prof. Dr. Ronny Veljanovski-School of Electrical Engineering, Faculty of Science Engineering and Technology, Victoria University, Melbourne, Australia. He can be contacted at email: mohammed_alturfi@yahoo.com.

Layth kamil Adday Almajmaie is a faculty member at the University of Technology, Department of Computer Engineering for more than 14 years, He was awarded a master degree from University of SAM India, in 2012, and Ph.D. from Department of Computer Engineering, University of Ultinbas, Turkey. His research interest is in network, wireless network, network security, data mining-cloud computing, distributed database, and image processing. He can be contacted at email: layth.adday@uotechnology.edu.iq.