I.J. Image, Graphics and Signal Processing, **2019, 10, 30-39**
Published Online October 2019 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijigsp.2019.10.05

An Optimized Architecture of Image Classification Using Convolutional Neural Network

Muhammad Aamir[1], Ziaur Rahman[1]
[1] College of Computer Science
Sichuan University, No.24 South Section 1, Yihuan Road, Chengdu, China, 610065
Email: aamirshaikh86@hotmail.com, ziaurrahman167@yahoo.com

Waheed Ahmed Abro[2]
[2]School of Computer Science and Engineering
Southeast University Sipailou No.2, Nanjing, China,210096
Email: engr.waheedabro@gmail.com

Muhammad Tahir[3], Syed Mustajar Ahmed[4]
[3]School of Software Technology, [4]School of Computer Science and Electrical Engineering
Dalian University of Technology, Dalian, China,116620
Email: muhammad.tahir.shaikh@gmail.com, itssyed@mail.dlut.edu.cn

Received: 22 July 2019; Accepted: 23 August 2019; Published: 08 October 2019

Abstract—The convolutional neural network (CNN) is the type of deep neural networks which has been widely used in visual recognition. Over the years, CNN has gained lots of attention due to its high capability to appropriately classifying the images and feature learning. However, there are many factors such as the number of layers and their depth, number of features map, kernel size, batch size, etc. They must be analyzed to determine how they influence the performance of network. In this paper, the performance evaluation of CNN is conducted by designing a simple architecture for image classification. We evaluated the performance of our proposed network on the most famous image repository name CIFAR-10 used for the detection and classification task. The experiment results show that the proposed network yields the best classification accuracy as compared to existing techniques. Besides, this paper will help the researchers to better understand the CNN models for a variety of image classification task. Moreover, this paper provides a brief introduction to CNN, their applications in image processing, and discuss recent advances in region-based CNN for the past few years.

Index Terms—Convolutional neural network, deep learning, image classification, precision, recall.

I. Introduction

CNN is one of the most popular deep neural network architectures, which comes in numerous variations. There are several features such as convolutional operation, characteristic of parameter sharing, and shift-invariant, which makes them typical deep learning model in computer vision. The underlying CNN architecture has formed by stacking three types of layers on top of each other, that are a convolutional layer[1], pooling layer, and fully connected layer, also known as a dense layer[2] [3]. A simplified CNN architecture for dog classification is illustrated in Fig.1.

Moreover, CNN offers exceptional performance in machine learning problems. In particular, applications that deal with image data, such as a complete image classification dataset. In the last decade, the accuracy of image classification has been improved with the advance of deep learning, especially concerning CNN. This increase in the efficiency of image classification has led researchers and developers to approach larger models to solve complex problems, which was not possible with classical artificial neural networks (ANNs) [4].CNN has been effectively applied to a variety of deep learning problems, such as object recognition, object classification, speech recognition, and, in particular, problems associated with massive image data. The first CNN is introduced by LeCun et al. [5] in 1990, and its improved version developed in 1998 [6]. A multi-layer ANN which can be trained with the backpropagation algorithm called LeNet-5 has been designed to classify handwritten digits [7]. The network can transform the original image into useful representations so that it can recognize visual patterns directly from unprocessed pixels without much preprocessing [8]. However, due to the lack of extensive training data and computing power at this time, LeNet-5 cannot achieve excellent results in more complex matters, such as: In the classification of images and videos on a large scale. After this number of methods

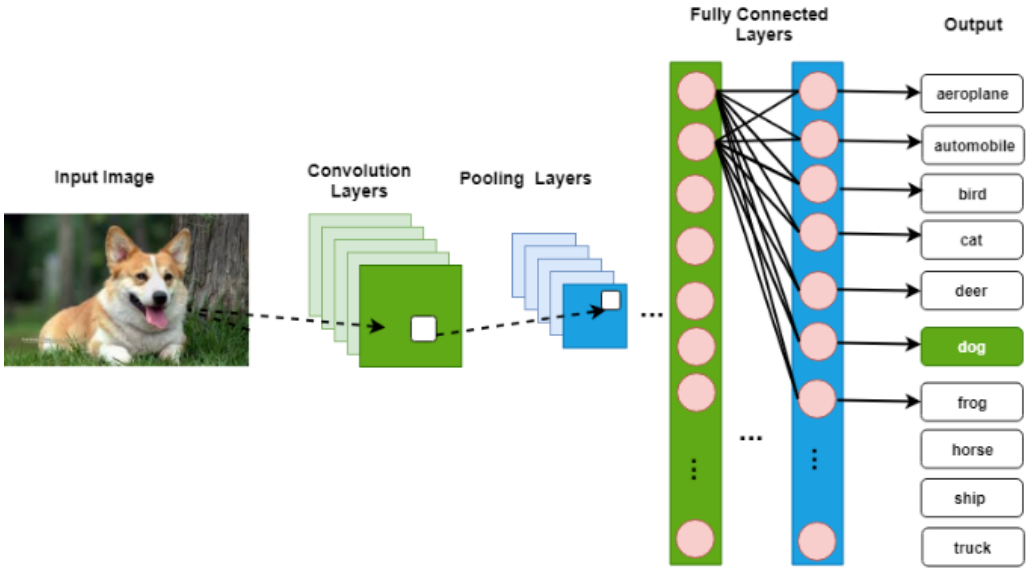

Fig.1. A simplified CNN architecture for dog Classification

have been proposed to address the difficulties in training the deep neural network.

Most unusual is a popular deep CNN, developed byKrizhevshy et al. [9] was AlexNet.The network introduced in 2012 included the availability of high computing devices (i.e., GPU, a very deep network of 60 million and 650,000 neurons, etc.). AlexNet outperformed all previous competitors and accepted the challenge by reducing the error of the top 5 to 15.3%. The error rate of the top 5 positions, which is not a variation of CNN, was around 26.2%. With the popularity of AlexNet, in 2013, Matthew et al. [10] developed a model to visualize and understand the convolutional network, attempting to outdo the model developed by Krizhevsky et al. After the visualize model of Krizhevsky et al. it is observed that the small changes in architecture improved classification performance. The only disadvantage of AlexNet is that the model had too many parameters.

Extending our discussion about CNN, NIN et al. [11] team developed the network which utilized a fewer number of parameters. The model had 7.5 million parameters, as compared to AlexNet's 60 million parameters. Furthermore, Google team purposed a new, deep CNN model, called the Inception [12]. This model reduced the network parameters to 4 million as compared to AlexNet's 60 million parameters. Besides, for object detection, the model used a similar approach to R-CNN, but, for proposals generation, the model combined selective search and multi-box methods, with 50% of proposals taken from the selective-search and 200 proposals received from the multi-box [13]. Furthermore, an improvement of Inception- ResNet is proposed by Dai et al. [14], which introduces deformable convolution and deformable region of interest (ROI) pooling. Moreover, VGG et al. [15] team developed an even more in-depth CNN. The team observed that the depth of the convolutional system has a great deal of impact on image detection. For their mode, the VGG team utilized a small 3×3, convolutional filters and set the convolutional stride to one. Therefore, no information got lost, all while using has 19 weighted layers.

On the other hand, numerous methods have been proposed to improve CNN and to apply CNN effectively in computer vision task. Researchers are trying to find solutions to a variety of problems by adopting CNN and its components, such as: For example, layer design, regularization, loss function, fast computation, and activation function. SqueezeNet et al. [16] presented a small DNN architecture that achieves an AlexNet level precision in ImageNet with 50 times fewer parameters. Also, the proposed method can be compressed to a size of 0.5 MB using the compression techniques of the model, which is smaller than that of AlexNet. Besides, Joseph et al. [17] proposed a CNN architecture called YOLO (You Only Look Once) for the unified recognition of objects in real-time. The network has 24 convolutional layers, followed by two fully connected layers. By alternating the convolution layer 1×1, the space of the features is reduced from the previous layers. Convolutional layers are pre-trained on the ImageNet classification task by setting half the resolution of an input image that is 224×224, and then the resolution has doubled for detection. Jonas et al. [18] proposed a CNN architecture for learning sequence by sequence. The model exceeds the recurrent models that could not recognize the structure of the composition in the sequences. Moreover, during the training process, all the elements can be fully parallelized for the better computations. Besides, non-linearities are made fixed and independent of the input length for more natural optimization. Aayush et al. [19] proposed PixelNet, using pixels for representations.

Based on these research patterns, to the best of the authors' knowledge, the problem of image classification using CNN is still in its infancy in the experimental findings and needs to be addressed appropriately. Therefore, to meet this goal in this research effort, the authors have introduced a network named "a simple CNN," which classify the different objects in CIFAR-10

imaginary and compare its performance with other existing image classification methods. The proposed network used in this study is an improvement over the traditional approaches to image classification. Moreover, the proposed network takes the fewer parameters, requires less memory and robustly minimizes the cost of iterations to classify the objects more quickly, which distinguishes this study from past research paradigms.

Furthermore, this paper is organized as follows: Section II presents the theoretical background of the study in details. Section III contains the proposed architecture. Section IV reports the experimental results. Section V offers discussion of the results. Finally, Section VI presents the conclusion and future work.

II. Theoretical Background

CNN has been used successfully in a wide range of tasks of computer vision and natural processing, such as the classification of images, the recognition of objects, the locating of objects, the tracking of objects, the generation of images, the estimation of human postures , the recognition and detection of texts, the visual question and answering , the recognition of actions, the visual saliency detection and the labeling of scenes. Its ability to handle a wide variety of tasks has given a breakthrough to deliver cutting-edge performance, especially in the field of computer vision and image processing. Besides, recent developments and advancements in CNN have led to outstanding performance, and its ability to be used in a variety of applications has revolutionized the world and its community. CNN is one of the most popular ANN architectures and has been used for a long time in image classification. The most popular use of CNN is to improve the accuracy of image classification. Compared to other methods, CNN achieves better classification accuracy for large data sets, as it allows the joint learning of features and classifiers [7]. Following the success of AlexNet, several contributions have significantly improved the accuracy of the classification by reducing the size of the filter [10] or by expanding the depth of the network [15,12].

However, the most successful approaches to object detection are currently extensions of image classification models. Detecting and classifying objects is one of the main problems in image processing. Over the years, many approaches to object detection have been proposed to solve these problems [20]. Also, object detection techniques have been categorized into two categories: CNN-based and non-CNN-based approaches. Approaches to recognizing objects that are not based on CNN, such as HOG, SVM, DPM, etc., have been widely used to classify object proposals into corresponding object categories [21].

Within the past several years, multiple attempts have been made to use CNN for object detection. The most prominent methods such as R-CNN [22], Fast R-CNN [23], Faster R-CNN [24], R-FCN [25], single-shot multi-box detector (SSD) [26], and others to locate and classify objects, not only label the class of an object, but also draw a bounding box around the position of the object in the image. This efficient detection capability makes the detection of objects a much more difficult task than the classification of images from conventional computer vision techniques.

Moreover, comprehensive review of the works which utilizing CNN to achieve robust performance is discussed as follows:

A. Object Tracking

Object tracking has always been a challenging research problem and has played an essential role in a variety of machine vision applications. In recent years, the tracking of objects has been widely used in the fields of transport, medical, military, and others [3]. Targeting forgone objects is a great challenge. There may be a variety of problems that affect the performance of object tracking algorithms, such as background noise, deformation, occlusion, sudden movement, and illumination variation. CNN has achieved great success and prevalent in many areas of computer vision. CNN has drawn a lot of attention among the vision community, along with visual object tracking. Several attempts have been made to achieve a robust visual tracking by combining deep neural networks. D. Li et al. [3] have developed an online tracking algorithm based on CNN. The framework combines CNN with kernelized correlation filters (KCF). Further to object tracking, the method achieves satisfactory tracking accuracy and robustness. Zhang et al. [27] developed a technique which solves the problem of multiple object tracking. The method combines CNN with the frame-pair input method for multi-object tracking also improves the tracking performance compared with previous deep neural networks-based trackers. Nevertheless, deeper systems are challenging to learn, and the spatial information is weaker in deeper layers, which positively affects the performance to localize the target. Moreover, Kokul et al. [28] have developed system learn discriminative features while reducing the spatial information lost.

B. Text Detection and Recognition

The text detection from images and character recognition is a challenging problem, which has achieved the significant amount of attention due to the many variations and uncontrollable factors, such as, variations in texts background, lighting, texture, font, size, style noise, and geometric distortions. However, various methods have been proposed to overcome these problems. The use of CNN has significantly improved both text detection and character recognition. Ren et al. [29] have proposed a novel scene-text detection algorithm based on CNN. The method uses a modified MSER detector I-MSER that extracts non-overlapped areas of the images to capture the text. Besides, the algorithm achieves high recall and excellent results. However, it is necessary to improve the speed of text detection and detection consistency. Nagaoka et al. [30] proposed the Multi-RPN Faster R-CNN model for text detection, which enhances

the detection score dramatically and more variety of texts generated in natural scene images.

C. Visual Silence Detection

In recent years, visual saliency detection has been one of the main problems in the field of computer vision, which has received increasing attention in many applications for complex tasks such as cognitive psychology, neurobiology and image processing. Visual saliency is the ability of an image processing system to locate the most relevant areas of the image quickly. CNN has already achieved significant results in the task of visual recognition, such as recognition of objects, paring of scenes, and classification of images. CNN is considered the correct option for this task. As a result, numerous computational models have been developed to detect visual saliency using CNN [31]. H. Misaghi et al. [32] have proposed a CNN-based visual saliency method which can identify multiple salient regions to any input size. G. Li et al. [33] have developed a high-quality visual saliency method, which extracts more robust multiscale features using CNN. The system considered as one of the successful visual recognition methods. Besides, a deep saliency multitasking model based on CNN has been proposed to investigate the feature sharing properties of salient object detection with a significant reduction in feature redundancy [34].

D. Visual Question Answering

Visual question answering is a technique to automatically produce the correct answer to arbitrary human-specific text-based questions from the image. The recent achievements of CNN in a wide variety of task have transformed them into one of the most useful architectures of current times. Therefore, CNN, have also been successfully applied to visual question answering. Singh et al. have proposed a system which combines CNN and long short-term memory networks (LSTM) to answer open-ended questions that are grounded in images [35].

Furthermore, Noh et al. [36] have developed the method, which joins CCN and parameter prediction network (PPN) for image questioning answering. The proposed method achieves robust performance on all possible public image questioning and answering benchmarks.

E. Human Pose Estimation

Human pose estimation remains the most challenging problem in computer vision until today, which has been receiving lots of consideration for well over 15 years. Despite many years of research, pose estimation still well-thought-out as a challenging task. However, until now, some approaches have been developed to achieve dynamic performance. Though, researchers are unable to get satisfactory results due to the variability of non-linear effects, such as variation in the visual appearance of a human in images, change in the lighting conditions, and variability in human physique. Besides, the occlusions in the images, the complexity of the human skeletal structure, and the high dimensionality of the posture can significantly affect performance.

Due to the above challenges, the estimation of the posture remains a challenging and mostly unresolved problem. CNN has been used with great success in a variety of visual tasks, such as: For example, the classification and detection of objects, facial recognition, text recognition, video action recognition and much more. As with other visual recognition tasks, the use of CNN in estimating human posture has dramatically improved performance due to CNN's abundant learning ability. Bearman et al. [37] has designed a CNN for the estimation of the human pose and the activity classification. The proposed method addressed the problem of regression of the estimation of joint human location and reached a PDJ value of around 60%. In addition, they achieved a classification accuracy of 80.51% in 20 activity categories. Toshev et al. [38] proposed a method for human pose estimation based on the deep neural network. In this model, pose estimation is formulated as a joint regression problem. The proposed method estimates the human pose based on the deep neural network. In this model, posture estimation is formulated as a joint regression problem. The approach leads to estimates of very high precision poses, and it's because the posture can be grounded in a holistic manner. The model is considered simple and powerful and shows the cascade of DNN regressors.

Liu et al. [39] has proposed a system based on CNN, which takes the RGB image as an input to estimate the head pose. The use of CNN dramatically improved the error rate and can get a test regression Euclidean loss of less than 0.0113. Tomas et al. [40] has proposed an efficient model for accurately estimating the pose of humans in gesture videos. The method evaluated on the BBC TV singing dataset. The achieved results show that pose predictions are significantly better, and the process is faster to compute than recent approaches. Furthermore, H. Vu et al. [41] has proposed a graphical model based on the estimation of human posture using CNN. The proposed CNN's configuration not only improves the accuracy of the existing network by up to 2% but also use fewer parameters, resulting in greater HPE accuracy and a more straightforward network structure.

F. Human Action Recognition

The growing popularity of CNN and its use as a practical solution for many vision problems have led many researchers to apply CNN to the tasks of video analysis and video comprehension, such as the detection of human actions. Based on the adoption of the methods using CNN, the recognition of human actions has significantly improved recognition performance. The detection of human actions based on CNN has been widely applied to some real-world applications, such as the detection of smart surveillance events, human-computer interaction, and video retrieval.

However, recognizing human actions in videos without restrictions is a challenge due to some real conditions such as different viewing angles & speed of actions, light

variations, and occlusion. Cheng et al. [42] has proposed a video surveillance system for real-time detection of human actions that predict human actions using temporal and CNN images. The use of CNN has achieved superior performance, and the process can be performed at approximately 20 frames per second, which is an excellent efficiency compared to other existing approaches. Wang et al. [43] has proposed a human action recognition method using CNN and their results compared with the bag of words (BOW) algorithm method. The experimental results show that the use of CNN outperforms the existing BOW method. Ravanbakhsh et al. [44] has developed the technique for action recognition using image-based CNN features. Optical flow inspires the technique. The CNN-flow has introduced to achieve better performance on action recognition accuracy as compared to other traditional methods.

Furthermore, Lin et al. [45] has introduced another cascaded deep architecture approach to human action recognition. The method uses factorized spatiotemporal CNN. The network effectively learns the spatiotemporal features using standard back-propagation. Earnest et al. [46] have developed the hybrid deep neural network model for efficient human action recognition. The method uses the bank features of the action to achieve the diversity of the classifier by diversifying the input features and varying the initialization of the ANN weights. The proposed method leads to a high recognition accuracy of 99.68%.

G. Scene Labeling

Scene labeling is a well-studied problem in image processing. This is the most difficult step towards a complete understanding of the image. Deep learning is currently a field of modern and active research, especially CNN. CNN is one of the most effective architectures for labeling scenes. The numerous parametric and non-parametric models have been proposed for labeling scenes [47,48,49,50,51].

C. Farabet et al. [52] have introduced a multi-scale CNN to learn scale-invariant local features for scene labeling. However, the proposed made failed to ensure the global contextual coherence and spatial consistency. To address this problem, CNN is combined with some post-processing techniques such as superpixels, CRF, and segmentation trees. Later T.Kekec et al. [53] have proposed a method which combines two different CNN models. The technique can learn context information and visual features in separate networks. The technology has drastically improved the learning accuracy by carefully designing the pre-processing steps to help the learning. Islam et al. [54] has introduced the method for dense image labeling using CNN. The proposed approach combines CNN with support vector machine (SVM) classifier. The model is evaluated on the Stanford datasets for background data (semantic, geometric) and Pascal VOC 2012. The result shows that the proposed model outperforms as compared to other existing techniques. Ming et al. [47] has developed the method for scene labeling, which uses CNN with intra-layer recurrent connections. The process can perform local feature extraction along with context integration, simultaneously. The backpropagation through time (BPTT) algorithm is used to train the system. Furthermore, the experimental result shows the effectiveness and efficiency of the model, which is the best over two benchmark datasets.

III. Proposed Architecture

A simple architecture has been designed to improve the expression ability and the performance of the network. However, we aimed to develop a simple optimal system with less memory consumption — the network which is not as deep as possible. Additionally, a simple network which can be fit to different situations. The block diagram of the proposed network shown in Fig.2 and a description of the design of the layers can be seen in Table 1. In addition, the output of each layer & its shape and the number of parameters obtained in each layer are shown Table 2.

A. The network was designed as follow:

- Convolutional input layer, 32 feature maps with a size of 3×3, a rectifier activation function, and a weight constraint of max norm set to 3.
- Max Pool layer with size 2×2.
- Convolutional layer, 64 feature maps with a size of 3×3, a rectifier activation function, and a weight constraint of max norm set to 3.
- Max Pool layer with size 2×2.
- Dropout set to 20%.
- Convolutional layer, 128 feature maps with a size of 3×3, a rectifier activation function, and a weight constraint of max norm set to 3.
- Max Pool layer with size 2×2.
- Flatten layer.
- Fully connected layer with 512 units and a rectifier activation function.
- Dropout set to 50%.
- Fully connected output layer with 10 units and a SoftMax activation function.

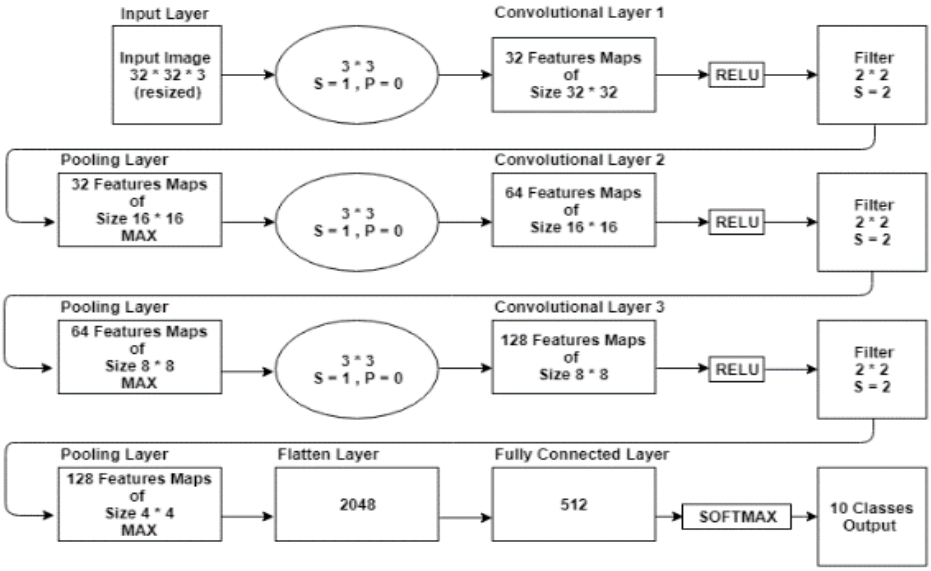

Fig.2. The architecture of proposed CNN

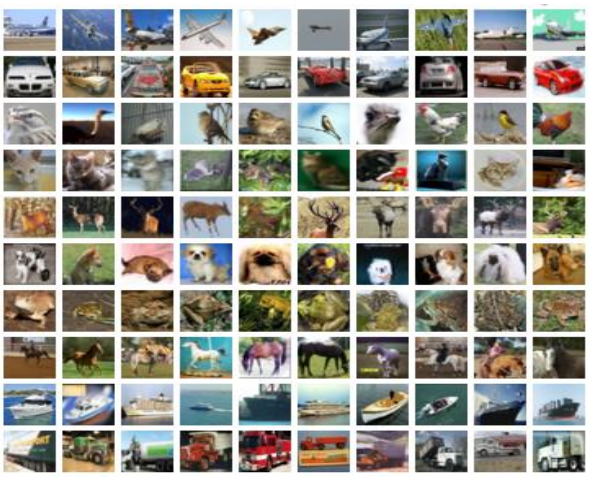

Fig.3. Examples of images from CIFAR-10 in all the ten different classes

Table 1. Description of the layers in the network

Layers	Type	Number of Maps	Dimension of Filters	Weight constraint of the max norm
0	Input	3	32×32	-
1	Convolutional	32	3×3	3
2	Pooling (MAX)	-	2×2	-
3	Convolutional	64	3×3	3
4	Pooling (MAX)	-	2×2	-
Dropout set to 20%				
5	Convolutional	128	3×3	3
6	Pooling (MAX)	-	2×2	-
Flatten layer				
11	Fully Connected	512	1	-
Dropout set to 50%				
12	SoftMax	10		-

Table 2. Summary representation of our model

Layer (type)	Output Shape	Param #
conv2d_1 (Conv2D)	(None, 32, 32, 32)	896
max_pooling2d_1 (MaxPooling2)	(None, 32, 16, 16)	0
conv2d_2 (Conv2D)	(None, 64, 16, 16)	18496
max_pooling2d_2 (MaxPooling2)	(None, 64, 8, 8)	0
conv2d_3 (Conv2D)	(None, 128, 8, 8)	73856
dropout_1 (Dropout)	(None, 128, 8, 8)	0
max_pooling2d_3 (MaxPooling2)	(None, 128, 4, 4)	0
flatten_1 (Flatten)	(None, 2048)	0
dropout_2 (Dropout)	(None, 2048)	0
dense_1 (Dense)	(None, 1024)	2098176
dropout_2 (Dropout)	(None, 1024)	0
dense_2 (Dense)	(None, 512)	524800
dropout_3 (Dropout)	(None, 512)	0
dense_3 (Dense)	(None, 10)	5130

IV. Evaluation and Results

To establish that the proposed network has a good performance for present purposes, a comparison between the proposed technique and previously reported techniques to image classification are presented in this section. The experimental setup was executed using Keras with an Intel(R) 2.66GHz CPU 4.0 GB RAM. The performance evaluation of proposed architecture is attained on the most famous image repository, CIFAR-10 [55], used for detection and classification tasks. 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: aeroplane, automobile, bird, cat, deer, dog, frog, horse, ship and truck as shown in Fig.3. The size of each image is 32×32 pixels. For

training the network, we choose the mini-batch Adam optimizer with a fixed learning rate of 0.001 and batch size of 32 samples. Furthermore, the network uses the binary cross-entropy (BCE) function, also called log loss for learning, as shown in Fig.4 and can be defined in Eq. (1) and Eq. (2). We train on 50,000 samples and test on 10,000 samples. Furthermore, the dropout rate is set to be 0.2 to avoid overfitting in the network.

$$BCE = -\sum_i^c t_i \log(f(s)_i) \tag{1}$$

$$f(s)_i = \frac{e^{s_i}}{\sum_j^c e^{s_j}} \tag{2}$$

Fig.4. The Cross-Entropy Loss Function

Moreover, the quality of the proposed network was evaluated. The precision (P), recall (R), f_1-score which is the harmonic mean of P and R) and overall accuracy of the network were derived from the following quantities: objects correctly identified (TP), objects labeled as unfavorable that are actually negative (TN), individuals incorrectly labeled as object (FP) and objects incorrectly labeled as not objects (FN), which were based on the quantities given in Eq. (3), Eq. (4) , Eq. (5) and Eq. (6) Mathematically, these were defined as:

$$P = \frac{\mathrm{TP}}{TP + FP} \times 100\% \tag{3}$$

$$R = \frac{TP}{TP + FN} \times 100\% \tag{4}$$

$$F_1 = 2 \times \frac{P \times R}{P + R} \tag{5}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \tag{6}$$

Table 3. Overall accuracy of the proposed network as compared to existing methods on CIFAR-10

Methods	***Accuracy %***
Logistic Regression (Softmax) [56]	40.76
KNN classifier [56,57]	22.84, 38.6
Patternet [58]	41.325
SVM [56,59]	55.22, 33.54
md et al. [60]	75.86
Author's	**78.29**

Table 4. Performance of the proposed network on 10 classes of CIFAR

Name	***Precision (Correctness)***	***Recall (Completeness)***	$f_1 - score$
Airplane	82.00	82.5	82.2
Automobile	87.83	88.8	88.3
Bird	79.43	61.4	69.2
Cat	62.53	61.1	61.8
Deer	67.61	80.8	73.6
Dog	71.33	68.2	69.7
Frog	82.42	84.9	83.6
Horse	81.72	81.4	81.5
Ship	87.48	88.1	87.7
Truck	81.69	85.7	83.6

In summary, Table 3 and Table 4 shows the performance evaluation of the proposed network. It can be seen from Table 3; The proposed network yields the best performance result compared to other approaches with a classification accuracy of 78.29 %. The results achieved are quite robust and reliable, which provides meaningful insight into the classification efficiency of the images. Furthermore, Table 4 depicts the quantitative stats to measure quality performance evaluation of the proposed network. The network performance measured by criterions for precision, recall rate, and f_1 -score. The said measures of measurement criterion are more indicative of examining "How efficiently the network can classify objects in comparison to other quality assessment measures addressed in the existing literature. Experimental results of the proposed network achieved high performance for most of the classes between three well-known quality assessment measurement criterions, i.e., precision, recall, and f_1 -score, respectively.

V. Discussion

Based on these research insights on image classification from CIFAR-10 imagery, this paper introduces a simple deep neural network. The proposed system carries the ability to improve the overall accuracy of image classification process as compared to previous methods (Table 3). Whereas the experimental results are presented in Section IV, the benefits and effects of the proposed model are highlighted while considering the previous methods. Furthermore, experimental results gave more insights into image classification algorithms and recognized as the deep neural network is the most suitable classifier for CIFAR-10. However, it is hard to say which classifier is best and selection of the classifier can be different according to the requirment. Depending on the classification problem the selection of the classifier may differ since there are lots of factors that can lower the classification accuracy.

In contrast to SoftMax logistic regression, the proposed system performs well for Image classification on CIFAR-10. Besides, it achieved a classification accuracy of 78.29%. The reason behind that the proposed network works well for non-linear and more extensive features obtained from images. Also, it doesn't require non-linear features transformation compared to logistic regression. Furthermore, in comparison to KNN, our network is much efficient in terms of overall classification accuracy. KNN is a simple architecture, but it requires high computational cost to train features. However, the proposed system also requires more computational cost but is very cheap to classify the images once the training of paramerers is acomplished.

Moreover, in comparison to SVM, our network can be trained efficiently, require less memory, run smoothly and easy to tune parameters. Likewise, our system performance is well for the problems with many examples or features, whereas this is considered as a drawback for SVM. Nevertheless, proposed architecture achieved high classification accuracy as compared to SVM.Besides, it is difficult for the SVM to paramtise. However, it outperforms other methods in the Table 3. Extending our results analysis about patternet, it outperforms the techniques discussed in [56,57,59]. Though, it obtained reduced performance considering proposed network. Furthermore, Mark et al [60] proposed an efficient model to classify images on CIFAR-10. The model achieves excellent performance in training at relatively high speed. The technique requires few tunable parameters as compared to others. Yet, in terms of efficacy on classification tasks, our method surpasses the classification accuracy.

As the main objective of this study revolves around the effectiveness in the classification accuracy, the above-discussed approaches have specific advantages and disadvantages. Still, the proposed model is efficient mainly in-terms of classification accuracy. Based on these research patterns, it is observed that the classification accuracy of the above-discussed methods can be varied for different problems in different situations. To conclude, it is challenging to decide which classifier performs well, which is totally dependent on type of data, size of images, parameter tuning, and so on.

VI. Conclusion

This paper briefly discusses the basic concepts of CNNs, and their rapid advancement over the past few years in a variety of computer vision applications include image classification, object recognition, object tracking, posture estimation, and so on. Based on these research findings, it is observed that the image classification using deep neural networks can achieve robust performance. However, there are still some issues that need to be addressed appropriately.

Therefore, in this paper, an effort has been made to propose a simple network to address the problem of image classification. The proposed network required less computational cost and consumed low memory. The network improves the classification accuracy and achieves good recognition results as compared to traditional approaches. Furthermore, the performance evaluation of the network depicts that it can be useful for designing a much faster classifier. The proposed network has the ability to solve different problems for a wide variety of applications, where the input to the network is an image. To conclude, this paper help researches, practitioners, and readers to better understand the problem of image classification, and choose the appropriate solution to it.

Furthermore, image classification speed and accuracy can be improved using the application of fractional calculus. In the future, we extend our work to design a network based on a fractional-order differential equation to achieve robust results in image classification.

References

[1] J. Gareth, W. Daniela, H. Trevor, and T. Rober, *An Introduction to Statistical Learning with Applications in R*. 2000.

[2] A. M. Andrew, "Second-order Methods for Neural Networks: Fast and Reliable Training Methods for Multi-Layer Perceptrons (Perspectives in Neural Computing Series)," *Kybernetes*. 1998.

[3] D. Li and W. Chen, "Object tracking with convolutional neural networks and kernelized correlation filters," in *Proceedings of the 29th Chinese Control and Decision Conference, CCDC 2017*, 2017.

[4] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," in *Proceedings of 2017 International Conference on Engineering and Technology, ICET 2017*, 2018.

[5] L. Cun *et al.*, "Handwritten Digit Recognition with a Back-Propagation Network," in *Advances in Neural Information Processing Systems 2*, 1990.

[6] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, 1998.

[7] J. Gu *et al.*, "Recent advances in convolutional neural networks," *Pattern Recognit.*, 2018.

[8] R. Hecht-Nielsen, "Theory of the backpropagation neural network," *Neural Networks*, 1988.

[9] A. Krizhevsky *et al.*, "ImageNet Classification with Deep Convolutional Neural Networks Alex," *Proc. 31st Int. Conf. Mach. Learn.*, 2012.

[10] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2014.

[11] M. Lin, Q. Chen, and S. Yan, "Network In Network," pp. 1–10.

[12] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2015.

[13] D. A. Dumitru Erhan, Christian Szegedy, Alexander Toshev, "Scalable Object Detection using Deep Neural Networks," *IEEE Conf. Comput. Vis. Pattern Recognit.*, vol. 2155–2162, pp. 787–790, 2014.

[14] J. Dai *et al.*, "Deformable Convolutional Networks," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017.

[15] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image

Recognition," pp. 1–14, 2014.
[16] F. N. Iandola, M. W. Moskewicz, K. Ashraf, S. Han, W. J. Dally, and K. Keutzer, "SqueezeNet," *arXiv*, 2016.
[17] J. Redmon *et al.*, "You Only Look Once: Unified, Real-Time Object Detection," *Adv. Neural Inf. Process. Syst. 27*, 2015.
[18] J. Gehring, M. Auli, D. Grangier, D. Yarats, and Y. N. Dauphin, "Convolutional Sequence to Sequence Learning," 2017.
[19] X. Chen and B. Russell, "PixelNet: Representation of the pixels, by the pixels, and for the pixels."
[20] M. Aamir, Y. F. Pu, W. A. Abro, H. Naeem, and Z. Rahman, "A hybrid approach for object proposal generation," in *Lecture Notes in Electrical Engineering*, 2019.
[21] H. Lee, S. Eum, and H. Kwon, "ME R-CNN : Multi-Expert R-CNN for Object Detection."
[22] R. Girshick, J. Donahue, T. Darrell, U. C. Berkeley, and J. Malik, "R-CNN," *1311.2524v5*, 2014.
[23] R. Girshick, "Fast R-CNN," *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2015 Inter, pp. 1440–1448, 2015.
[24] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN," *arixv*, 2015.
[25] R. F. C. Networks and J. Dai, "R-FCN : Object Detection via," *arXiv Prepr.*, 2016.
[26] W. Liu *et al.*, "SSD: Single shot multibox detector," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2016.
[27] Y. Zhang, Y. Tang, B. Fang, and Z. Shang, "Fast multi-object tracking using convolutional neural networks with tracklets updating," in *2017 International Conference on Security, Pattern Analysis, and Cybernetics, SPAC 2017*, 2018.
[28] T. Kokul, C. Fookes, S. Sridharan, A. Ramanan, and U. A. J. Pinidiyaarachchi, "Gate connected convolutional neural network for object tracking," in *Proceedings - International Conference on Image Processing, ICIP*, 2018.
[29] X. Ren, K. Chen, X. Yang, Y. Zhou, J. He, and J. Sun, "A novel scene text detection algorithm based on convolutional neural network," in *VCIP 2016 - 30th Anniversary of Visual Communication and Image Processing*, 2017.
[30] Y. Nagaoka, T. Miyazaki, Y. Sugaya, and S. Omachi, "Text Detection by Faster R-CNN with Multiple Region Proposal Networks," in *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*, 2018.
[31] Z. Rahman, Y. F. Pu, M. Aamir, and F. Ullah, "A framework for fast automatic image cropping based on deep saliency map detection and gaussian filter," *Int. J. Comput. Appl.*, 2019.
[32] H. Misaghi, R. A. Moghadam, and K. Madani, "Convolutional neural network for saliency detection in images," *2018 6th Iran. Jt. Congr. Fuzzy Intell. Syst. CFIS 2018*, vol. 2018-Janua, no. February, pp. 17–19, 2018.
[33] G. Li and Y. Yu, "Visual saliency detection based on multiscale deep CNN features," *IEEE Trans. Image Process.*, 2016.
[34] X. Li *et al.*, "DeepSaliency: Multi-Task Deep Neural Network Model for Salient Object Detection," *IEEE Trans. Image Process.*, 2016.
[35] M. Iyyer, "Deep Learning for Visual Question Answering," *Slides*, no. November, pp. 1–7, 2015.
[36] H. Noh, P. H. Seo, and B. Han, "Image question answering using convolutional neural network with dynamic parameter prediction," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016.
[37] A. Bearman and C. Dong, "Human Pose Estimation and Activity Classification Using Convolutional Neural Networks," *Stanford CS231n*, 2015.
[38] A. Toshev and C. Szegedy, "DeepPose: Human pose estimation via deep neural networks," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2014.
[39] X. Liu, "Head pose Estimation Using Convolutional Neural Networks," 2016.
[40] T. Pfister, K. Simonyan, J. Charles, and A. Zisserman, "Deep convolutional neural networks for efficient pose estimation in gesture videos," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2015.
[41] H. Vu, E. Cheng, R. Wilkinson, and M. Lech, "On the use of convolutional neural networks for graphical model-based human pose estimation," in *Proceedings - 2017 International Conference on Recent Advances in Signal Processing, Telecommunications and Computing, SigTelCom 2016*, 2017.
[42] C. Bin Jin, S. Li, T. D. Do, and H. Kim, "Real-time human action recognition using CNN over temporal images for static video surveillance cameras," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2015.
[43] M. Wang, "Human Action Recognition Using CNN and BoW Methods t-Distributed Stochastic Neighbor Embedding," vol. 2012, 2016.
[44] M. Ravanbakhsh, H. Mousavi, M. Rastegari, V. Murino, and L. S. Davis, "Action Recognition with Image Based CNN Features," 2015.
[45] L. Sun, K. Jia, D. Y. Yeung, and B. E. Shi, "Human action recognition using factorized spatio-temporal convolutional networks," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015.
[46] E. P. Ijjina and C. Krishna Mohan, "Hybrid deep neural network model for human action recognition," *Appl. Soft Comput. J.*, 2016.
[47] M. Liang, X. Hu, and B. Zhang, "Convolutional Neural Networks with Intra-layer Recurrent Connections for Scene Labeling," *Adv. Neural Inf. Process. Syst.*, 2015.
[48] D. Eigen and R. Fergus, "Nonparametric image parsing using adaptive neighbor sets," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2012.
[49] C. Liu, J. Yuen, and A. Torralba, "Nonparametric scene parsing via label transfer," in *Dense Image Correspondences for Computer Vision*, 2015.
[50] S. Gould, R. Fulton, and D. Koller, "Decomposing a scene into geometric and semantically consistent regions," in *Proceedings of the IEEE International Conference on Computer Vision*, 2009.
[51] V. S. Lempitsky, A. Vedaldi, and A. Zisserman, "A Pylon Model for Semantic Segmentation," *Nips'11*, 2011.
[52] C. Farabet, C. Couprie, L. Najman, and Y. Lecun, "Learning hierarchical features for scene labeling," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2013.
[53] E. Fromont, R. Emonet, T. Kekec, A. Trémeau, and C. Wolf, "Contextually Constrained Deep Networks for Scene Labeling," 2015.
[54] M. A. Islam, N. Bruce, and Y. Wang, "Dense image labeling using Deep Convolutional Neural Networks," in

Proceedings - 2016 13th Conference on Computer and Robot Vision, CRV 2016, 2016.
[55] A. Krizhevsky, "Learning Multiple Layers of Features from Tiny Images," *Ttechnical Rreport, Dep. Comput. Sci. Univ. Toronto*, 2009.
[56] "Logistic Regression SoftMax." [Online]. Available: https://github.com/wikiabhi/Cifar-10.
[57] "K-nearest neighbor classification." [Online]. Available: http://cs231n.github.io/classification/.
[58] "Pattern Recognition Network." [Online]. Available: https://www.mathworks.com/help/nnet/ref/patternnet.html.
[59] "Support Vector Machine." [Online]. Available: https://houxianxu.github.io/implementation/SVM.html.
[60] M. D. McDonnell and T. Vladusich, "Enhanced image classification with a fast-learning shallow convolutional neural network," in *Proceedings of the International Joint Conference on Neural Networks*, 2015.

Authors' Profiles

Muhammad Aamir received his Bachelor of Engineering Degree in Computer Systems Engineering from the Mehran University of Engineering & Technology Jamshoro, Sindh, Pakistan in (2008). And Master of Engineering Degree in Software Engineering from CHONGQING University P.R. China in (2014). Currently, he is a Ph.D. Research Student in Sichuan University P.R. China. His research interest includes Pattern Recognition, Computer Vision, Image processing, deep learning and fractional calculus.

Ziaur Rahman received MS degree in software engineering in 2017 from Chongqing University, Chongqing, China. Currently, He is pursuing PhD degree from Sichuan University, Chengdu, China. His research includes are image processing, deep learning and fractional calculus.

Waheed Ahmed Abro received his Bachelor of Engineering Degree in Computer System Engineering from the Mehran University of Engineering & Technology, Jamshoro, Sindh, Pakistan in (2008). And Master of Science in Computer system from National University of Computer and Emerging Sciences, Islamabad, Pakistan in (2015). Currently he is Ph. D. Research student in Southeast University P.R. china. His research interest includes Natural Language Processing, Spoken Language Understanding, Computer Vision and Pattern Recognition.

Muhammad Tahir received the B.S. degree in software engineering from the University of Sindh, Jamshoro Sindh, Pakistan, in 2008, and the M.S. degree in software engineering from the School of Software Engineering, Chongqing University, China, in 2014. He is currently pursuing the Ph.D. degree in software engineering with the School of Software Technology, Dalian University of Technology, China. He is on Ph.D. Study leave from Lecturer position with the Department of Computer Science, COMSATS University Islamabad, Sahiwal Campus, Pakistan. He has authored/coauthored publications in World renowned journals. His research interests include network security, web application performance tuning, mobile edge computing, game theory, artificial intelligence, and machine learning.

Syed Mustajar Ahmed received BS degree from Xidian University Xi'an, China in 2016, he is currently pursuing the Master's degree at Dalian University of Technology, department of Computer Science & Electrical Engineering, his research interest includes Machine Learning, Deep learning, NLP and Image processing.

How to cite this paper: Muhammad Aamir, Ziaur Rahman, Waheed Ahmed Abro, Muhammad Tahir, Syed Mustajar Ahmed, "An Optimized Architecture of Image Classification Using Convolutional Neural Network", International Journal of Image, Graphics and Signal Processing(IJIGSP), Vol.11, No.10, pp. 30-39, 2019.DOI: 10.5815/ijigsp.2019.10.05

