Image Recognition, Classification and Analysis using Convolutional Neural Networks

Rakesh K R
Department of Electronics and Telecommunication
R V College of Engineering, Bangalore, India
rakeshkr@rvce.edu.in

G R Namita
Department of Electronics and Telecommunication
R V College of Engineering, Bangalore, India
grnamita.te17@rvce.edu.in

Rohit Kulkarni
Department of Electronics and Telecommunication
R V College of Engineering, Bangalore, India
rohitkulkarni.te17@rvce.edu.in

Abstract— **This paper presents a comprehensive discussion and analysis of the various architectures of Convolutional Neural Networks for image classification. This paper intends to implement and analyze the performance of AlexNet, VGG16, VGG19 and ResNet50 as Image Classifiers on the dataset CIFAR10. Accuracy, Loss and Confusion Matrix were used as metrics to analyze the performance.**

Keywords— ***Image Classification, Convolutional Neural Network, CNN, AlexNet, VGGNet, ResNet.***

I. INTRODUCTION

Image classification has several applications in the current world ranging from MRIs to Driverless cars. To achieve these, Deep Learning and Neural Networks play an important role. Using Deep Learning and Neural Networks, we are able to mimic the human brain to process and solve complex problems. Convolutional Neural Network (CNN) models are used to perform image recognition and classifying. CNNs are preferred over other types of Neural Networks due to their ability to detect low level and high level features from images.

A CNN model has both convolutional and non-convolutional. The convolutional layers form the basis of these neural networks. It takes an input from the previous layer, transforms it by performing the convolution operation. It is able to detect features, forming the basis of image classification. To detect these features, the convolutional layers have filters which detect these patterns. The deeper the network goes, the more sophisticated these filters become. The non-convolutional layers include: Pooling layers which are used to increase/decrease the size of the image, Fully Connected Layers which are used to perform the non linear combinations of the detected features to classify the images.

This paper discusses trained deep CNN models to perform image classification. The models discussed in this paper are AlexNet, VGGNet: VGG16, VGG19 and ResNet50. The dataset used for training and testing the network is the CIFAR10 [7].

II. LITERATURE REVIEW

Deep learning algorithms based on convolutional neural networks can be used to classify images. It is observed that the model with greater number of layers classifies images with greater accuracy due to its ability to detect and extract features [1].

There are specific applications which require deeper feature extraction to classify objects in images at high accuracy. Convolution Neural Networks are used compared to other Neural Network architectures such as ANN or CNN due to the ability to extract low-level and high level features [2].

Classifiers built on CNN models can be applied on large scale datasets for recognition and classification. The learned images are transferred and reused in the classification with the aim of building a robust classifier [3].

With the addition in the number of layers, accuracy of Image classifiers was expected to increase. However with deeper neural networks, the accuracy began to saturate and decrease with the more layers due to the Vanishing Gradient problem. To solve this, Residual Networks were developed which provided the original image as a reference to each layer and thus improved the learnability, thus allowing us to build deeper CNN models at higher accuracy [4].

III. PROPOSED ARCHITECTURE

Image Classification is performed using the 5 steps (Fig 1):

1) Pre-processing: The image pixel values are converted from 0-255 to 0-1 and the labels are one hot encoded.

2) Data Partitioning: The images and their labels are split into training and testing data, at a 80:20 split.

3) Building the model: Tensorflow and Keras functions are used to build the models.

4) Training and Testing: The model is trained on the training data following which it is implemented on the testing data.

5) Analysis: The model is evaluated based on its performance on both the training and testing data.

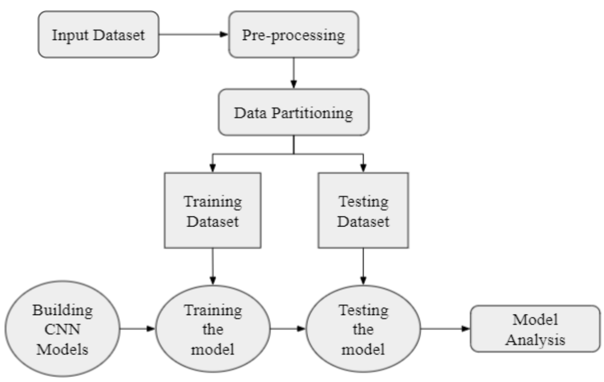

Fig 1: Proposed Architecture

2022 First International Conference on Electrical, Electronics, Information and Communication Technologies (ICEEICT) | 978-1-6654-3647-2/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICEEICT53079.2022.9768474

IV. CNN ARCHITECTURES

A. AlexNet

AlexNet architecture (Fig 2) consists of 5 convolutional layers, 3 max-pooling layers, 2 normalization layers, 2 fully connected layers, and 1 softmax layer. The activation function used is ReLU.

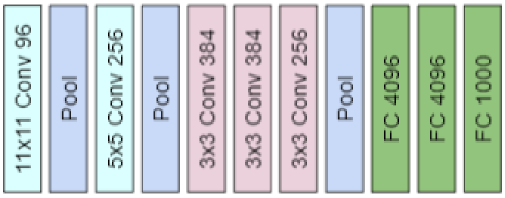

Fig 2: AlexNet Architecture

B. VGG16

Unlike AlexNet, VGG16 does not focus on hyper parameters. It uses 3 x 3 filters with stride 1 in the convolution layers and uses ‘same padding’ in pooling layers with stride 2. VGG16 (Fig 3) has 13 convolution and pooling layers followed by 3 FC layers. The activation function used is ReLU and the classifier used is SoftMax.

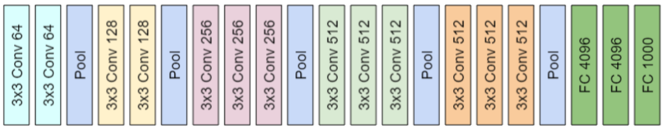

Fig 3: VGG16 Architecture

C. VGG19

VGG19 architecture (Fig 4) is similar to VGG16 with the only difference being that there are 3 additional convolution layers with a total of 19 layers. The pooling type used was max-pooling with ReLU as the activation function and SoftMax as the classifier.

.

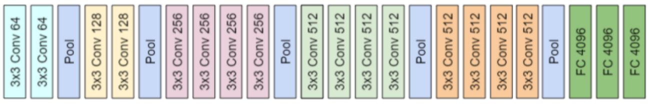

Fig 4: VGG19 Architecture

D. ResNet50

To solve the Vanishing Gradient problems in Deeper CNN models, ResNets were developed. It uses skipped to provide an alternative path for the gradient. Convolution Blocks (Fig 5) and Identity blocks (Fig 6) are used to implement ResNet models. Convolution blocks are used when the convolved image and the input image aren’t of the same size. ResNet50 has 4 Convolution and 12 Identity Blocks followed by FC layers to implement a 50 layer architecture (Fig 7).

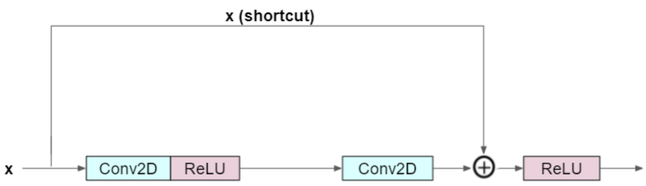

Fig 5: Identity Block

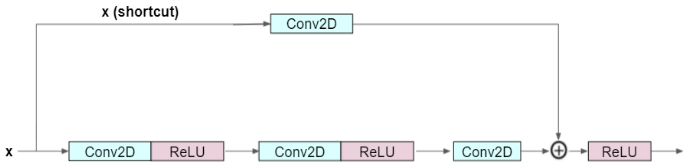

Fig 6: Convolution Block

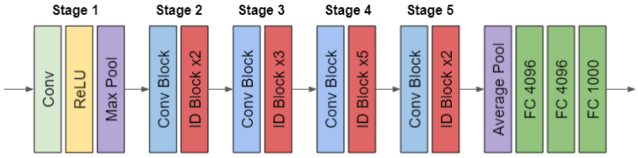

Fig 7: ResNet50 Architecture

V. IMPLEMENTATION

The dataset used in this paper for the purpose of training all the models is CIFAR10 [7]. It contains 60,000 images divided into 10 classes. 50,000 images comprise the train set and the remaining 10,000 are the test images. To compile all the models, the optimizer used is “Adam” (Eqn-1), loss function being “categorical cross-entropy” and the metric being “Accuracy”. The batch-size used was 256 and the model was evaluated for 50 epochs.

$$w_t = w_{t-1} - \eta \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon}$$

Eqn-1: Adam Optimizer.

η = Learning Rate

w_t = Weight

m = Moment of the vector

$$\text{Loss} = -\sum_{i=1}^{\text{output size}} y_i \cdot \log \hat{y}_i$$

Eqn-2: Categorical Cross Entropy

$\hat{Y}_i$ = ith scalar value of the input

y_i = Corresponding target scalar value.

On completion of the 50th epoch, training and testing loss and accuracy for each model is noted. The confusion matrices were obtained to see the number of incorrect predictions

VI. RESULTS AND ANALYSIS

On implementing AlexNet, VGG16, VGG19 and ResNet50 on the CIFAR10 dataset, a comparative analysis was performed. The metrics used to evaluate each model’s performance was Accuracy (Training and Testing), Loss and Confusion Matrices to evaluate the classifiers.

On observing the results in Table 1, it can be observed that AlexNet performs poorly compared to other models with lower test accuracy and higher loss. The remaining 3 models perform efficient classification on the test set. VGG19 shows the highest accuracy whereas ResNet50 offers the minimum loss.

Table 1: Results

Model	Training Accuracy (%)	Training Loss	Testing Accuracy (%)	Testing Loss
AlexNet	99.20	0.0310	76.83	1.5065
VGG16	99.87	0.0048	91.74	0.4847
VGG19	99.72	0.0095	92.46	0.4354
ResNet50	99.55	0.0042	90.63	0.1815

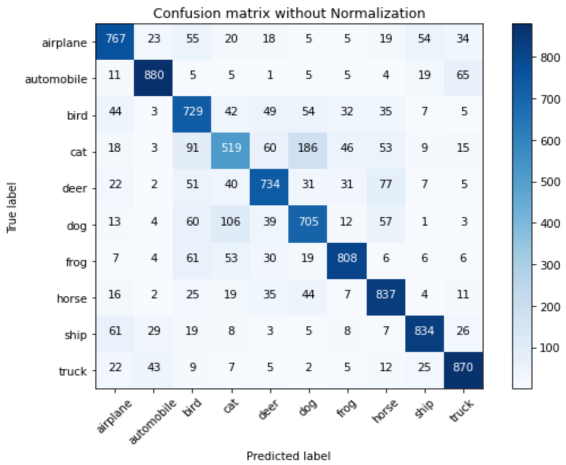

Fig 8: Confusion Matrix-AlexNet

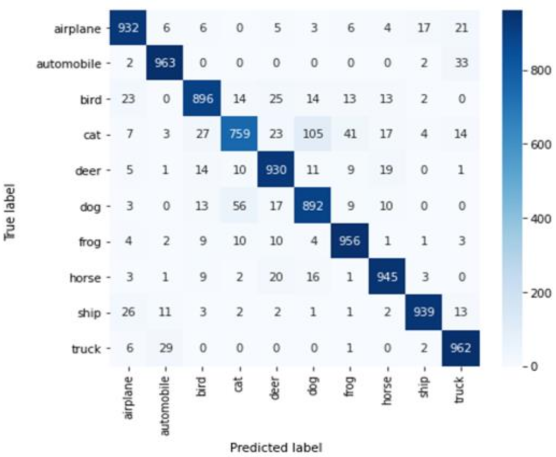

Fig 9: Confusion Matrix-VGG16

On observing the confusion matrices, it is seen that all 4 models had a large number of correct predictions proving that they are efficient classifiers. On performing a comparative analysis, AlexNet (Fig 8) had the most incorrect predictions out of all the 4 models and VGG19 (Fig 10) had the least number of incorrect predictions.

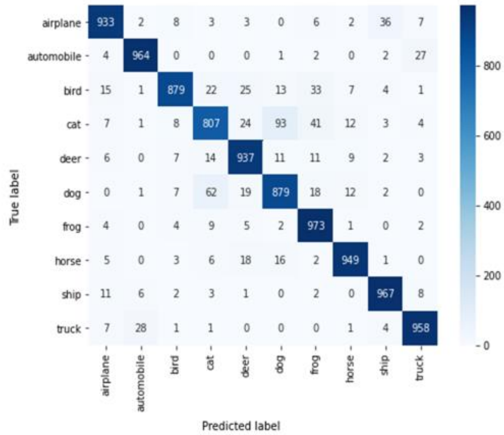

Fig 10: Confusion Matrix-VGG19

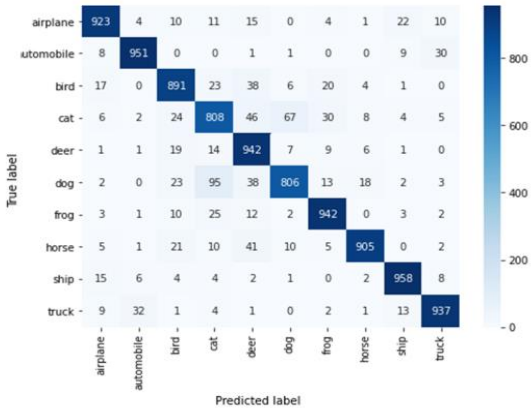

Fig 11: Confusion Matrix-ResNet50

VII. CONCLUSION

In this paper, Image Classification using Convolutional Neural Networks was presented. The main reason CNN is chosen is because of its ability to extract low and high level image features and perform classification at high accuracy.

The 4 CNN models implemented were AlexNet, VGG16, VGG19 and ResNet50 and the dataset used is CIFAR10 Each model was evaluated for 50 epochs at the end of which the Accuracy and Loss and Confusion Matrix were metrics used for evaluation. It is observed that all 4 models were efficient at classifying images at high accuracy. On performing a comparative analysis, VGG19 was the best model and AlexNet was the worst.

REFERENCES

[1] H. A. Shiddieqy, F. I. Hariadi and T. Adiono, "Implementation of deep-learning based image classification on single board computer," 2017, *International Symposium on Electronics and Smart Devices (ISESD), Yogyakarta, Indonesia*, pp. 133-137, doi: 10.1109/ISESD.2017.8253319.

[2] N. Jmour, S. Zayen and A. Abdelkrim, "Convolutional neural networks for image classification," 2018

International Conference on Advanced Systems and Electric Technologies (IC_ASET), Hammamet, Tunisia, pp. 397-402, doi: 10.1109/ASET.2018.8379889.
[3] M. T. Islam, B. M. N. Karim Siddique, S. Rahman and T. Jabid, "Image Recognition with Deep Learning," 2018 *International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS)*, Bangkok, Thailand, pp. 106-110, doi: 10.1109/ICIIBMS.2018.8550021.
[4] T. Haryanto, I. Wasito and H. Suhartanto, "Convolutional Neural Network (CNN) for gland images classification," 2017 *11th International Conference on Information & Communication Technology and System (ICTS)*, Surabaya, Indonesia, pp. 55-60, doi: 10.1109/ICTS.2017.8265646.
[5] S. Tripathi and R. Kumar, "Image Classification using small Convolutional Neural Network," 2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, pp. 483-487, doi: 10.1109/CONFLUENCE.2019.8776982.
[6] Learning Multiple Layers of Features from Tiny Images, Alex Krizhevsky, 2009.