Image Classification on CIFAR-10 Using Deep Convolutional Neural Networks

N V Bharani Subramanya Kumar[1], C V Mahesh Reddy[2], CH. Samyana Reddy[3], Krishn Chand Kewat[4], Laxmi Narsimha Talluri[5], Shaik Mohammed Rahil Sarfaraz[6], and Sushama Rani Dutta[7]

[1,2,3,4,5,6]Department of Electronics and Communication Engineering,
[7]Department of Computer Science and Engineering
[1,2,3,4,5,6,7]Koneru Lakshmaiah Educational Foundation, Hyderabad 500075, India
[1]2210040083@klh.edu.in, [2]2210040077@klh.edu.in, [3]2210040032@klh.edu.in,
[4]2210040076@klh.edu.in, [5]2210040129@klh.edu.in, [6]2210040110@klh.edu.in,
[7]sushamarani.dutta@klh.edu.in

Abstract. This work showcases an improvement over existing methods by developing a novel deep convolutional neural network (CNN) architecture for image classification specifically targeting the images in the CIFAR-10 dataset [4] which consists of 60,000 color images (32 x 32 pixels size) divided into 10 classes. So far, the model architecture incorporates a number of convolution and pooling layers which are then followed by the fully connected layers to better learn the complex structure existing within the input spatial configuration. The typical challenge of overfitting is addressed by employing various techniques such as data augmentation and dropout regularization strategy. Immediately from the experimental evidence, it is clear that the deep CNN performs superior to other traditional models in the case of image recognition classifying problems and therefore the model has proved to be robust in discerning the differences that exist in the categories in the images within the CIFAR-10 dataset.

Keywords: CIFAR-10, Deep Convolutional Neural Network (CNN), Image Classification, Data Augmentation, Regularization, Machine Learning.

1 Introduction

The most powerful tool to be used for tasks of image classification has been recently demonstrated to be deep learn- ing, specifically CNNs [2]. Image classification is used as a procedure in which predefined labels are assigned to input images; however, this procedure holds more significance for applications, including autonomous driving, facial recognition, medical imaging, and object detection, among a few. As the real-world applications become more complex, the demand for such accurate and efficient models of classification is increasing.

Among the numerous image datasets, CIFAR-10 has emerged as the benchmark to test how well a given image classification model works.

The dataset of CIFAR-10 contains 60,000 color images divided into 10 classes, with each image 32x32 pixels in size. The categories involved in the CIFAR-10 include airplane, automobile, bird, cat, deer, dog, frog, horse, ship, and truck. Since CIFAR-10 has low image resolution, coupled with the natural class overlap, makes it a challenging dataset, thus an ideal testbed for gauging machine learning models' capability.
Traditional methods-like SVMs and k-NN algorithms-relied [8]heavily on handcrafted features to get somewhat acceptable performance. However, a complex dataset like CIFAR-10 is way beyond the scope of high-performance capabilities offered by traditional approaches. The breakthrough of CNNs in image classification has bypassed deep learning with the ability to learn hierarchical features directly from raw image pixels without feature engineering and significant boosts in accuracy.
The construction of CNN architectures drastically improved the field of image classification. The utilization of deeper net- works and residual connections through efficient optimization techniques leads to training on datasets like CIFAR-10 much faster with state-of-the-art architectures.

Table 1: Model Architecture Summary

Layer (type)	**Output Shape**	**Param #**	**Comments**
First Half of the Model			
Conv2D	(32, 32, 32)	896	First Convolutional Layer
MaxPool2D	(16, 16, 32)	0	MaxPooling Layer
Conv2D	(16, 16, 64)	18,496	Second Convolutional Layer
MaxPool2D	(8, 8, 64)	0	MaxPooling Layer
Second Half of the Model			
Conv2D	(8, 8, 128)	73,856	Third Convolutional Layer
MaxPool2D	(4, 4, 128)	0	MaxPooling Layer
Conv2D	(4, 4, 256)	295,168	Fourth Convolutional Layer
GlobalAveragePooling2D	(256)	0	Global Average Pooling
Dense	(128)	32,896	Fully Connected Layer
Dropout	(128)	0	Dropout Layer
Dense	(10)	1,290	Output Layer (Softmax)
Total Parameters			**422,602**

Although these advancements brought about huge progress, CIFAR-10 is still considered an arduous dataset because images are of a small size and there are larger intra-class variations. For example, the category "bird" consists of sev- eral species, and there are variations in models, angles, and conditions of "car." Moreover, classes like "frog" and "cat" have shapes and features that differ less, which renders the task of accurate classification quite difficult. State-of-the-art models cannot be said to be precise even now when it comes to perfect accuracy on CIFAR-10.
There are a lot of techniques that are used to overcome these problems. Data augmentation techniques such as random cropping and flipping and rotation may extend the training set, and in this way, make the models

more robust over possible overfitting. Regularization techniques, such as dropout and batch normalization, prevent overfitting and ensure that models generalize well to new data. Transfer learning has also proven effective, as the CIFAR-10 can fine-tune on top of the learned model trained on large datasets like ImageNet to use learned features and boost performance.

Optimisations like Adam [5] allow the training of such deep CNNs at a faster rate of convergence and also minimize getting stuck in local minima. Along with the above optimisation, hyperparameter tuning is concerning the learning rates, batch size, and finally network architectures, which affects the performance model in very significant ways.

Advanced techniques brought great improvements in the classification accuracy on CIFAR-10, but it is still very much a ways away from the state of human performance on this very dataset. This paper discusses some realizations with the implementations of deep CNN architectures on CIFAR-10 as a challenge to further push the boundaries of classifier accuracy and how this can be used in the exploration of new approaches to further improvement.

The rest of the paper is organized into the following: Section 2 is the related work overview concerning the classification of CIFAR-10 in the context of CNN-based methods and associated challenges. Section 3 introduces the methodology applied for the design and training of our CNN model in terms of architecture, data preprocessing techniques adopted, and the hyperparameter optimization strategy. Section 4 presents the evaluation of our model by comparing its performance against previous methods. Conclusion of Section 5 This paper concludes the development of CIFAR-10 classification and gives a direction for future work to overcome the challenges posed by this problem.

2 Literature Survey

The CIFAR-10 dataset has been the benchmark for any successful image classification task for a long time. The dataset, extensively used to check the performance of machine learning and deep learning models, consists of 60,000 32x32 color images spread over 10 classes: airplane, automobile, bird, cat, deer, dog, frog, horse, ship, and truck. The images are much lower resolution than in the very large datasets like ImageNet, so it would actually be a good test case for developing classification models that can classify small, high-variance images.

2.1 Traditional Machine Learning Approaches

Before the deep learning breakthrough, other classical ma- chine learning methods people used to apply on CIFAR-10 in- clude SVMs and k-NN. One common feature or characteristic of many of these models is that they relied on handcrafted fea- tures: Histogram of Oriented Gradients (HOG), Scale-Invariant Feature Transform (SIFT). However, these approaches were inherently limited by the dependence on manually engineered features, which significantly reduced their performance for highly complex datasets such as CIFAR-10. Since this process of extracting the features was not optimized for data, such models had terrible classification performances on harder tasks.

2.2 Convolutional Neural Networks (CNNs) for CIFAR-10

The advent of CNN was a paradigm shift in image classi- fication, at least as far as CIFAR-10 datasets were concerned. CNN can automatically learn relevant features from the raw image pixels, thus avoiding the manually engineered features. The early work in deep learning models such as AlexNet in the year 2012 had already presented the power of CNNs in large-scale image classification tasks but went on to prove that a deeper network was necessary for CIFAR-10's challenges. Therefore, it was pushed further by other networks, namely VGGNet in 2014; the deeper models have better accuracy than shallow ones. There was a simpler architecture and smaller size of convolutional filter in VGGNet, but that is how it could extract richer features. With increased depth, however, came increased computational cost, making the training slower and more resource-hungry.

2.3 Data Augmentation and Regularization

One technique that has helped to make CNNs better at tackling CIFAR-10 is data augmentation [1]. Data augmen- tation techniques artificially enlarge the size of the training set through such processes as random cropping, flipping, and rotation. This lowers the risk of overfitting and makes models more robust. It also allows the network to see variations of the same object, hence improving the model's ability to generalise. Others are L2 regularization, also referred to as weight decay. These are primarily penalties to large weights and reduce the tendency for models overfitting.

2.4 Optimization

Optimization of deep neural network learning process is very important, and the choice of optimization algorithm may play a great role in achieving the success of the model. The Adam optimizer has gained much popularity in deep learning since it combines the property of both RMSprop and Momentum-based Gradient Descent with adaptive learning rates, improving efficiency with convergence. In particular, Adam outperforms in CIFAR-10 classification tasks with con- verging networks that do not overshoot and are stable in the training procedure. Since Adam is controlling the learning rate using first and second moments of the gradient, it contributes to faster convergence as well as avoids being stuck into the local minima, especially at small image size and complex intra-class variations.

2.5 Challenges with CIFAR-10

Although many architectures and training methods of CNNs are developed, the CIFAR-10 dataset still contains intra-class variances and inter-class similarities and remains challenging [3]. For instance, shapes in classes like "airplane" and "ship" may look similar; in contrast, classes such as "frog" and "cat" have overlapping body structure and color points and do not seem to delimit clear boundaries between the classes. Moreover, the low resolution of the images used in the CIFAR- 10 dataset complicates the goal of capturing finer details. Even state-of-the-art models ResNet-110 and EfficientNet are impressive for achieving high accuracy rates but, at the same time, fail to perfect classification. Such types of errors mostly occur on images where the features are not perfectly distin- guishable from each other to classify between classes

3 METHODOLOGY

This section describes the methodology used to develop and train the deep Convolutional Neural Network, CNN model, for image classification utilizing the CIFAR-10 dataset. The process involved a number of steps: data preprocessing, design of the model, training, and evaluation.

3.1 Dataset Preprocessing

The CIFAR-10 dataset comprises 60,000 color images, each measuring 32x32 pixels, and is divided into 10 distinct classes. A total of all the images forms two subsets comprising 50,000 training images and 10,000 test images. All the images are in RGB format and belong to one of the 10 classes.
To preprocess this dataset: *'Normalisation Scaling all pixel values in images to the range [0, 1] using division by 255. This operation also makes it possible to process images efficiently for the neural network.
Data Augmentation To reduce the risk of overfitting and improve generalization, real-time data augmentation tech- niques are utilized during training. Such techniques include random horizontal flips and rotations, plus random cropping. This strategy increases the diversity of the training set without requiring any additional manually labeled data

3.2 Model Architecture

The model used is a deep Convolutional Neural Network (CNN) designed to capture spatial hierarchies that would exist in images found in the CIFAR-10 set. The architecture consists of layers of convolutional layers, with pooling layers between them, finally ending in a fully connected layer at the end. More details about the architecture are given below:
* Convolutional Layers: The model starts with a series of two-dimensional convolutional layers, where max-pooling layers follow each to reduce the spatial dimensions. The convolutional layers use ReLU activation to introduce non-linearity in and benefit the ability of such a model to pick

up intricate patterns. The number of filters starts at 32 and increases to 256 progressively in subsequent layers, thereby allowing the network to increase its capability to learn more sophisticated features as the depth expands.
* Global Average Pooling: After convolutional and pooling layers, a Global Average Pooling layer reduces spatial dimen- sions down to a 1x1 feature map. It will avoid overfitting and reduce the number of parameters in a fully connected layer.
* The fully connected layer:It has a dense configuration of 128 units, processing the output received from the pooling layer using ReLU activation. Dropout also has been applied at a rate of 0.5 to further enhance regularization of the model and to reduce the possible overfitting.
* Output Layer: This is the last layer, and it implements a softmax activation function with 10 units representing the ten classes of the CIFAR-10. Here, the probability distribution for all the above respective classes is generated.

3.3 Model training

The model is trained with this following configuration [7]: The categorical cross-entropy loss function is optimized with an Adam optimizer. It is adaptive to one that adjusts the learning rate dynamically with regards to the first and second moment of the gradients.
* Loss Function: Since this is a multi-class classification, the loss function chosen is categorical cross-entropy.
* Batch Size: The training employs a batch size of 64 throughout. This is based on experimentation with results that find an optimal compromise between training time and performance.
* Epochs: The model is trained on for 50 epochs with early stopping based on validation loss, meaning the training stops as soon as the performance of the model starts degrading for a validation set.
* Evaluation Metrics:The accuracy of the model is analyzed using accuracy, recall, precision and F1 score to ensure to generalize well on classes.

3.4 Hyperparameter Tuning

For parameter tuning [6], the model needs to run with the grid search and manual experimentation for optimization. Good working parameters, including learning rates, batch size, and dropout rate, are tuned to optimize the performance of the model. In addition, several regularization techniques such as dropout and data augmentation are used to increase the robustness of the model and avoid overfitting.

4 EVALUATION AND RESULTS ANALYSIS

4.1 Dataset

The CIFAR-10 dataset is another very widely benchmarked set, especially for image classification tasks, in the field of computer vision and machine learning. It contains 60,000 col- ored images with each image of resolution 32x32 pixels. The dataset is divided into 10 distinct classes: airplane, automobile, bird, cat, deer, dog, frog, horse, ship, and truck. The class is well balanced, with 6,000 images per class. Further, this is then subdivided into 50,000 images for training and 10,000 images for testing. The method employed is multi-class classification because, as stated before, a single image has a class label. Being relatively smaller in size, CIFAR-10 is computationally inexpensive, thus researchers can experiment easily with mul- tiple machine learning algorithms and archi- tectures.The size is compact and simple, allowing CIFAR-10 to be used for testing not only classical methods like SVMs and k-NN, but also more recent deep architectures made up of CNNs.The loss evolution and the accuracy evolution graph has been shown in the fig 2.

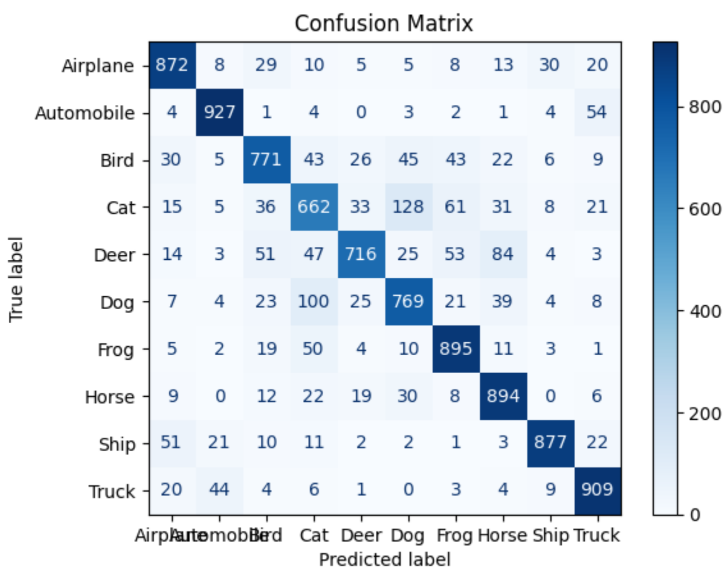

Fig. 1: Confusion Matrix of Deep CNN for CIFAR-10 Dataset

4.2 Performance of other models on CIFAR-10 Dataset

Per Model Performance Using the Testing Accuracy and Test Loss on numerous machine learning models, it was noted that SVM Polynomial and HOG could deliver up to 40 percent. Meanwhile, CNN presented a similar test loss at 74 percent with lower accuracy at 74 percent. The performance of SVM linear was average with an accuracy of 36 percent together with a test loss of 44 percent, while the lowest accuracy was recorded by KNN at 34 percent with much higher loss at 66 percent.

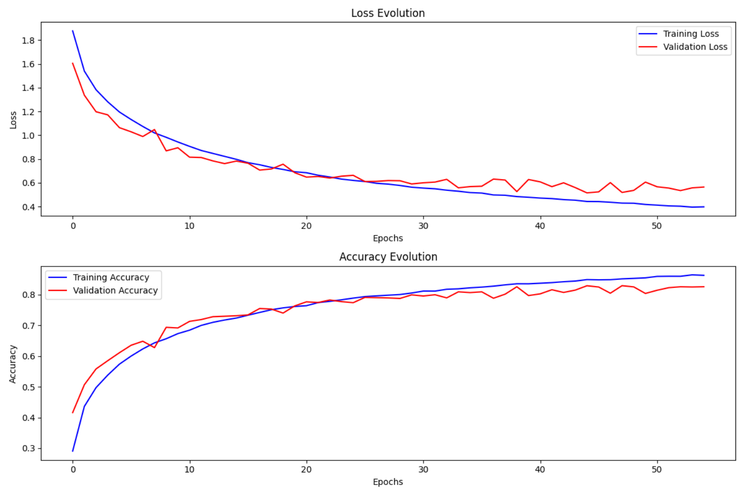

Fig. 2: Loss Evolution and Accuracy Evolution

Our proposed Deep CNN achieved much higher accuracy 88 percent CIFAR than others, implying that Deep CNN is better for using in the classification of images which can be observed in the fig 3.

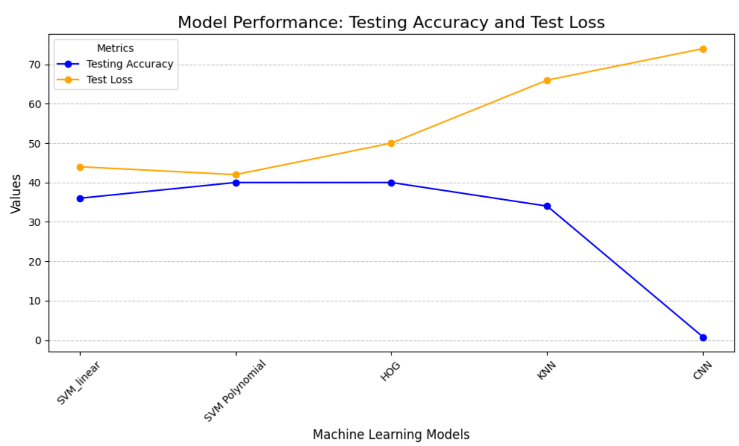

Fig. 3: Test accuracy and test loss on other models

5 CONCLUSION

This work proposed a completely new CNN architecture for the classification task within CIFAR-10 images. In this, the model learned to extract fairly complex spatial features in the right order from inside the data, by effectively making use of the convolution and pooling layers followed by the dense layers. Along with dropout regularization, data augmentation made overfitting impossible over some measures. Those ex- perimental results impressively reflect the deep CNN's per- formance as introduced, significantly outperforming classical models in tasks described with recognition of images and validating the idea of adopting the proposed method in an application for correctness differentiation between categories within the CIFAR-10 dataset.

References

1. Randall Balestriero, Leon Bottou, and Yann LeCun. The effects of regularization and data augmentation are class dependent. *Advances in Neural Information Processing Systems*, 35:37878–37891, 2022.
2. Md Anwar Hossain and Md Shahriar Alam Sajib. Classification of image using convolutional neural network (cnn). *Global Journal of Computer Science and Technology*, 19(2):13–14, 2019.
3. Soumya Joshi, Dhirendra Kumar Verma, Gaurav Saxena, and Amit Paraye. Issues in training a convolutional neural network model for image classification. In *Advances in Computing and Data Sciences: Third International Conference, ICACDS 2019, Ghaziabad, India, April 12–13, 2019, Revised Selected Papers, Part II*, volume 3, pages 282–293. Springer, 2019.
4. Benjamin Recht, Rebecca Roelofs, Ludwig Schmidt, and Vaishaal Shankar. Do CIFAR-10 classifiers generalize to CIFAR-10? *arXiv preprint arXiv:1806.00451*, 2018.
5. Mohamed Reyad, Amany M Sarhan, and Mohammad Arafa. A modified Adam algorithm for deep neural network optimization. *Neural Computing and Applications*, 35(23):17095–17112, 2023.
6. Patrick Schratz, Jannes Muenchow, Eugenia Iturritxa, Jakob Richter, and Alexander Brenning. Hyperparameter tuning and performance assessment of statistical and machine-learning algorithms using spatial data. *Ecological Modelling*, 406:109–120, 2019.
7. Kenji Suzuki, Hiroyuki Abe, Heber MacMahon, and Kunio Doi. Image-processing technique for suppressing ribs in chest radiographs by means of massive training artificial neural network (mtann). *IEEE Transactions on Medical Imaging*, 25(4):406–416, 2006.
8. Pin Wang, En Fan, and Peng Wang. Comparative analysis of image classification algorithms based on traditional machine learning and deep learning. *Pattern Recognition Letters*, 141:61–67, 2021.