CNN-Based Object Classification using CIFAR-10 with Classical Machine Learning Comparison

1st Sankar Shailesh*, 2nd Shrijith V*, 3rd Ahamed Jaseem J**
*Department of CSE, Sri Siva Subramanyia Nadar College of Engineering, Chennai, India
shailesh2410305@ssn.edu.in,shrijith2410051@ssn.edu.in, ahamedjaseem2410548@ssn.edu.in

***Abstract*—Image classification is a fundamental problem in computer vision. In this work, a lightweight Convolutional Neural Network (CNN) is designed and implemented using PyTorch for object classification on the CIFAR-10 dataset. The performance of the CNN is compared with classical machine learning models such as Support Vector Machine (SVM) and Logistic Regression. The proposed CNN model demonstrates significantly better performance due to its ability to learn spatial and hierarchical features from images. Experimental results show that the CNN achieves higher classification accuracy compared to classical models, highlighting the importance of deep learning techniques in image recognition tasks.**

***Index Terms*—Convolutional Neural Network, CIFAR-10, Image Classification, Deep Learning, Support Vector Machine**

I. Introduction

Image classification plays a major role in many real-world applications such as object detection, autonomous systems, and medical imaging. Traditional machine learning approaches rely heavily on manually extracted features and often struggle with high-dimensional image data.

Convolutional Neural Networks (CNNs) have become the standard approach for image-related tasks due to their ability to automatically extract spatial features using convolution operations. In this project, a lightweight CNN model is developed and evaluated on the CIFAR-10 dataset. The model's performance is compared with classical machine learning models to analyze the effectiveness of deep learning in image classification.

The main contributions of this work are as follows:

- Design and implementation of a lightweight CNN for CIFAR-10 classification.
- Application of data normalization and augmentation techniques.
- Performance comparison between CNN and classical machine learning models.
- Visualization of learned feature maps to understand model behavior.

II. Literature Survey

Early approaches to image classification relied heavily on handcrafted feature extraction techniques. Popular feature descriptors such as Scale-Invariant Feature Transform (SIFT) and Histogram of Oriented Gradients (HOG) were commonly used to capture edge and texture information from images. These handcrafted features were then combined with classifiers such as Support Vector Machines (SVM) to perform object recognition. Although such methods achieved reasonable accuracy, they required extensive domain knowledge and manual feature engineering. The performance of these models was highly dependent on the quality of the extracted features, making them less adaptable to complex and large-scale image datasets.

The emergence of deep learning significantly transformed the field of computer vision. Convolutional Neural Networks (CNNs), initially introduced for document recognition tasks [1], demonstrated that hierarchical feature extraction could be learned automatically from raw pixel data. Unlike traditional methods, CNNs use convolutional layers to preserve spatial relationships between neighboring pixels while progressively learning higher-level representations.

With the advancement of computational power and availability of large datasets, deep CNN architectures became more practical and effective. Modern deep learning models are capable of learning low-level features such as edges and textures in early layers and more complex object-specific features in deeper layers. As discussed in [2], deep neural networks eliminate the need for handcrafted features by learning representations directly from data through backpropagation.

Over the past decade, CNN-based approaches have consistently outperformed classical machine learning models in benchmark image classification tasks. Their ability to automatically extract spatial and hierarchical features makes them particularly well-suited for visual recognition problems, including object classification on datasets such as CIFAR-10.

III. Methodology

The project follows a structured pipeline consisting of dataset loading, preprocessing, model design, training, and evaluation.

A. Dataset Organization

The CIFAR-10 dataset consists of 60,000 color images of size 32×32 pixels across 10 different object classes. The dataset is divided into 50,000 training images and 10,000 testing images. The dataset was loaded using the torchvision library in PyTorch.

B. Pre-processing and Input Construction

The input images were converted into tensors and normalized to improve training stability. Normalization ensures that pixel values are scaled appropriately for faster convergence.

Data augmentation techniques such as random horizontal flipping and random rotation were applied to the training dataset to improve generalization and reduce overfitting.

For classical machine learning models, the images were flattened into one-dimensional vectors before training.

C. CNN Architecture

The proposed CNN consists of two convolutional layers followed by max pooling layers. The first convolution layer uses 32 filters of size 3×3 to capture low-level features such as edges and textures. The second convolution layer uses 64 filters to capture higher-level patterns.

Each convolution layer is followed by a ReLU activation function to introduce non-linearity. Max pooling is applied to reduce spatial dimensions and computational complexity.

The extracted feature maps are flattened and passed to fully connected layers for classification. The final layer consists of 10 neurons corresponding to the CIFAR-10 classes.

D. Training Details

The model was trained using the Adam optimizer with a learning rate of 0.001. The batch size was set to 64, and the model was trained for multiple epochs.

Data augmentation techniques such as random horizontal flipping and random rotation were applied to improve generalization. The dataset was normalized to stabilize training and improve convergence speed.

IV. Results and Discussion

The performance of the proposed CNN model was compared with classical machine learning models including Support Vector Machine (SVM) and Logistic Regression. All models were trained and evaluated on the same dataset to ensure a fair comparison. For the classical models, the images were flattened into one-dimensional feature vectors before training.

The CNN achieved an accuracy of approximately 75% on the test dataset. In comparison, the SVM model achieved around 45% accuracy, while Logistic Regression achieved approximately 30%. These results clearly indicate that the CNN significantly outperforms traditional machine learning approaches for image classification on the CIFAR-10 dataset.

The superior performance of the CNN can be attributed to its ability to preserve spatial information and learn hierarchical features through convolution operations. Unlike classical models, which treat images as independent pixel values in a flattened vector, CNNs exploit local connectivity and shared weights to extract meaningful patterns. This allows the model to learn edges, textures, shapes, and object-level features progressively across layers.

In contrast, classical machine learning models rely solely on the raw pixel intensities after flattening, which leads to loss of spatial structure and contextual information. As a result, their decision boundaries are less effective in distinguishing visually similar classes such as cats and dogs or automobiles and trucks.

Feature map visualization further demonstrated that early convolutional layers capture low-level features such as edges and color gradients, while deeper layers respond to more abstract and complex patterns. This hierarchical learning mechanism enables CNNs to generalize better and achieve higher classification accuracy compared to classical approaches.

TABLE I
Model Performance Comparison

Model	Test Accuracy (%)
CNN	73.9
SVM	46.65
Logistic Regression	28.15

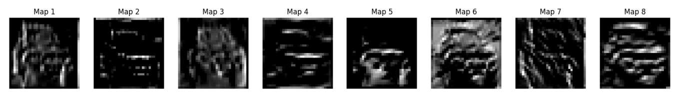

Fig. 1. Feature Map Visualization from First Convolution Layer

V. Conclusion

In this work, a lightweight Convolutional Neural Network (CNN) was implemented for image classification on the CIFAR-10 dataset and its performance was compared with classical machine learning models such as Support Vector Machine and Logistic Regression. The experimental results clearly demonstrate that the CNN significantly outperforms the classical approaches. The improvement in performance can be attributed to the CNN's ability to preserve spatial relationships between pixels and to learn hierarchical feature representations directly from raw image data.

The comparison highlights an important limitation of traditional machine learning models when applied to image data. Flattening images into one-dimensional vectors removes spatial structure, which reduces the model's ability to distinguish between visually similar classes. In contrast, convolution operations allow the CNN to progressively learn edges, textures, and higher-level object features, leading to improved generalization and classification accuracy.

Overall, this study reinforces the effectiveness of deep learning methods for computer vision tasks. Future work may include experimenting with deeper architectures, applying transfer learning techniques, performing hyperparameter tuning, and analyzing class-wise performance in greater detail. Such improvements could further enhance model accuracy and robustness.

VI. Project Repository

The source code and implementation details for this project are publicly available at:

`https://github.com/shailesh1606/CNN-CIFAR10/`

References

[1] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[2] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep learning*. MIT press Cambridge, 2016, vol. 1, no. 2.
[3] A. Krizhevsky, G. Hinton *et al.*, "Learning multiple layers of features from tiny images," 2009.
[4] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, F. Pereira, C. Burges, L. Bottou, and K. Weinberger, Eds., vol. 25. Curran Associates, Inc., 2012.