BALAMANOJ REDDY KOMMAREDDY

Application of CIFAR-100 for efficient object detection by using Deep Learning

The report deals with training and evaluating different deep learning models on the CIFAR-10 dataset. Here's a summary of the main aspects:

Objective:

The objective is to build, train, and evaluate several deep learning models to classify images from the CIFAR-10 dataset.
The models are designed to achieve high accuracy in image classification tasks.

Dataset:

The CIFAR-10 dataset consists of 60,000 32x32 color images in 10 classes, with 6,000 images per class.
The dataset is preprocessed by scaling pixel values to the range [0, 1].

Model Architectures:

- Five different models are experimented with:
- MLP with Batch Normalization.
- MLP with L2 regularization.
- Simple CNN.
- CNN with Data Augmentation and Dropout.
- ResNet.

Each model architecture is described in detail, highlighting the layers and parameters used.

Training:

- All models are compiled with the Adam optimizer and sparse categorical cross-entropy loss.
- Each model is trained for 10 epochs on the training data and validated on the test data.
- Training progress is printed for each model, including loss and accuracy.

Evaluation:

- After training, the test accuracy of each model is reported.
- Learning curves (accuracy and loss) are plotted for each model to visualize training progress.

Model Comparison:

- Various metrics like accuracy, precision, recall, and F1 score are calculated for each model.

- Models are compared based on these metrics to determine their performance.
- A bar chart is plotted to visually compare the performance of each model.
- Main Issues:
- The models' performance might be limited by their complexity or architecture.
- Overfitting may occur in some models, particularly those with a large number of parameters.
- Some models might require further optimization or tuning of hyperparameters to improve performance.

Key Findings:

- The ResNet model achieves the highest accuracy among the tested models, with an accuracy of approximately 81.4%.
- Data augmentation and dropout help improve the performance of the CNN model significantly.
- Simple CNN and CNN with Data Augmentation and Dropout models achieve good accuracy with less complexity compared to MLP-based models.
- The MLP models, despite incorporating Batch Normalization and L2 regularization, do not perform as well as the CNN-based models.

Introduction:

Image classification is a fundamental task in computer vision with applications ranging from medical imaging to autonomous driving. In this report, I focus on the problem of classifying images from the CIFAR-10 dataset, which contains 60,000 32x32 color images across 10 different classes. Each image in the dataset belongs to one of the following classes: airplane, automobile, bird, cat, deer, dog, frog, horse, ship, or truck.

The importance of accurate image classification cannot be overstated. In fields like healthcare, correctly identifying medical images can aid in the diagnosis of diseases. In autonomous vehicles, precise image classification is crucial for recognizing objects and making informed decisions about navigation. Additionally, image classification is used extensively in security systems, industrial automation, and many other domains.

The CIFAR-10 dataset, while relatively small compared to other image datasets like ImageNet, presents a challenging task due to the low resolution and diversity of objects. Therefore, it serves as an excellent benchmark for testing and evaluating different deep learning models.

In this study, I explore the effectiveness of various deep learning architectures in classifying CIFAR-10 images. I investigate the performance of models ranging from simple Multi-Layer Perceptrons (MLPs) to more complex Convolutional Neural Networks (CNNs) and compare their accuracy, precision, recall, and F1 score. By doing so, I aim to identify the most suitable model for image classification tasks on the CIFAR-10 dataset. This research is critical for advancing the state-of-the-art in image classification and providing insights into the performance of different neural network architectures.

Current Research :

The research paper "Object detection from images by convolutional neural networks for embedded systems using Cifar-10 images" [1] by Tushar Singh and Vinod Kumar aims to develop a model for embedded systems to detect objects from the CIFAR-10 images dataset using convolutional neural networks (CNNs). The main focus of the experiment is to use less memory size and train the model with good accuracy within a limited time so that the model can be utilized for embedded systems. CNNs are chosen for image classification because they are designed to handle input data in the form of a matrix, which is suitable for processing images, as images are also stored as matrices.

The research article "Deep Convolutional Neural Network Compression based on the Intrinsic Dimension of the Training Data" [2] by Abir Mohammad Hadi and Kwanghee Won, published in ACM SIGAPP Applied Computing Review, addresses the challenge of selecting the optimal deep learning architecture for a specific task and dataset.

Traditionally, this process involves exhaustive searches for neural network architectures or multi-phase optimization, including initial training, compression or pruning, and fine-tuning steps. In this study, the authors propose an approach that employs a deep reinforcement learning-based agent to dynamically compress a deep convolutional neural network (CNN) during its training process.

The key innovation of their approach lies in integrating the concept of the intrinsic dimension of the training data, providing the agent with insights into the complexity of the task. The agent utilizes two distinct ranking criteria, L1-norm-based and attention-based measures, to selectively prune filters from each layer as it deems necessary.

In their experiments, the authors used the CIFAR-10 dataset and its subsets (2-class and 5-class subsets) to model the task complexity. They demonstrated that the agent learns different policies depending on the intrinsic dimension. On average, the agent pruned off 78.48%, 77.9%, and 83.12% filters from all layers of the VGG-16 network for CIFAR-10 full, 5-class, and 2-class subsets, respectively.

The paper "AdaGossip: Adaptive Consensus Step-size for Decentralized Deep Learning with Communication Compression" [3] proposes a novel technique to address the communication overhead in decentralized learning setups. Their method, AdaGossip, adaptively adjusts the consensus step-size based on compressed model differences between neighboring agents, eliminating the need for manual tuning of hyper-parameters. Through extensive experiments on various Computer Vision datasets and network topologies, AdaGossip achieves superior performance compared to the current state-of-the-art methods, showing a 0-2% improvement in test accuracy. This approach significantly reduces communication overhead, making on-device learning over large distributed datasets more practical and efficient.

Data Collection:

For this project, I utilized the CIFAR-10 dataset, which is commonly used as a benchmark dataset for image classification tasks. The CIFAR-10 dataset consists of 60,000 32x32 color images across 10 different classes, with 6,000 images per class. These images are evenly distributed among the following categories: airplane, automobile, bird, cat, deer, dog, frog, horse, ship, and truck.

The dataset is divided into a training set and a test set, with 50,000 images for training and 10,000 images for testing. The training set is used to train the model, while the test set is used to evaluate its performance on unseen data.
Each image in the CIFAR-10 dataset is represented as a 32x32 pixel grid with three color channels (RGB), resulting in a total of 3,072 features per image. The pixel values are scaled between 0 and 255, representing the intensity of the red, green, and blue channels.

Model Development:

To develop the image classification models, I used various deep learning architectures implemented using the TensorFlow and Keras libraries in Python. Specifically, I experimented with the following types of models:

Multi-Layer Perceptron (MLP): A simple feedforward neural network consisting of multiple fully connected layers. Each pixel in the input image is treated as a separate feature, and the network learns to classify images based on these features.

Convolutional Neural Network (CNN): A deep learning architecture specifically designed for image classification tasks. CNNs use convolutional layers to automatically learn hierarchical representations of images, capturing spatial patterns and features.

Transfer Learning: Leveraging pre-trained CNN models, such as VGG16, ResNet50, and MobileNet, trained on large image datasets like ImageNet. By fine-tuning these pre-trained models, I can adapt them to the CIFAR-10 dataset, potentially achieving better performance with less training time and data.

I trained each model using the training set of CIFAR-10 images and evaluated their performance using the test set. By comparing the accuracy, precision, recall, and F1 score of these models, I aimed to identify the most effective approach for image classification on the CIFAR-10 dataset.

Analysis :

Here are the findings of the analysis:

Model	Accuracy	Precision	Recall	F1 Score
MLP with BatchNorm	0.5116	0.536297	0.5116	0.511579
MLP with L2	0.3970	0.390330	0.3970	0.379442
Simple CNN	0.6909	0.692682	0.6909	0.687796
CNN with DA and Dropout	0.6956	0.695649	0.6956	0.692129
ResNet	0.8137	0.819477	0.8137	0.814185

Table1. Comparison of metrics of Trained Model

From the analysis, several key findings emerge:

ResNet outperforms other models: ResNet achieved the highest accuracy of 81.37% among all the models tested, indicating its superior performance in image classification on the CIFAR-10 dataset.

CNNs outperform MLP: Both Simple CNN and CNN with Data Augmentation (DA) and Dropout achieved higher accuracy compared to MLP-based models. This suggests that CNNs are better suited for image classification tasks due to their ability to capture spatial hierarchies and local patterns in images.

Data Augmentation and Dropout improve performance: The CNN model with Data Augmentation and Dropout achieved slightly higher accuracy compared to the Simple CNN, indicating that techniques like data augmentation and dropout regularization can help improve the model's generalization and performance.

Regularization techniques affect performance: The MLP model with L2 regularization achieved lower accuracy compared to the one with Batch Normalization. This suggests that Batch Normalization helps stabilize training and improve the convergence of the model.

Overall, the results suggest that deep learning models, especially CNNs like ResNet, are effective for image classification tasks on the CIFAR-10 dataset. Regularization techniques like Batch Normalization, Data Augmentation, and Dropout further improve the model's performance and generalization ability.

Summary and Conclusion :

In summary, the project focused on studying various deep learning models for image classification on the CIFAR-10 dataset. Here are the key findings and conclusions:

Model Performance: Five different models were evaluated: MLP with Batch Normalization, MLP with L2 regularization, Simple CNN, CNN with Data Augmentation (DA) and Dropout, and ResNet. Among these, ResNet achieved the highest accuracy of 81.37%, outperforming all other models.

Effectiveness of CNNs: CNN-based models (Simple CNN, CNN with DA and Dropout, and ResNet) consistently outperformed MLP-based models (MLP with Batch Normalization and MLP with L2 regularization). This suggests that CNNs are more suitable for image classification tasks due to their ability to capture spatial hierarchies and local patterns in images.

Regularization Techniques: Regularization techniques such as Batch Normalization, L2 regularization, Dropout, and Data Augmentation were employed to improve model performance. Among these, models with Batch Normalization and Data Augmentation showed better performance compared to models without these techniques.

Generalization and Robustness: Models with regularization techniques demonstrated better generalization and robustness against overfitting. Data Augmentation and Dropout proved particularly effective in improving the model's ability to generalize to unseen data.

Implications: These findings have practical implications for image classification tasks, indicating that deep learning models, especially CNNs like ResNet, are effective for such tasks. Employing regularization techniques like Batch Normalization, Data Augmentation, and Dropout can further enhance model performance and generalization.

In conclusion, the study highlights the importance of selecting appropriate deep learning architectures and regularization techniques for image classification tasks. Future work could explore more advanced architectures and fine-tuning strategies to achieve even higher performance on the CIFAR-10 dataset and other image classification benchmarks.

References

[1] Singh, T., & Kumar, V. (2023). Object detection from images by convolutional neural networks for embedded systems using Cifar-10 images. In Artificial Intelligence, Blockchain, Computing and Security (Vol. 1, pp. 5). CRC Press.

[2] Hadi, A. M., & Won, K. (2024). Deep Convolutional Neural Network Compression based on the Intrinsic Dimension of the Training Data. ACM SIGAPP Applied Computing Review, 24(1), 14–23. https://doi.org/10.1145/3663652.3663654

[3] Aketi, S. A., Hashemi, A., & Roy, K. (2024). AdaGossip: Adaptive Consensus Step-size for Decentralized Deep Learning with Communication Compression. arXiv preprint arXiv:2404.05919. https://doi.org/10.48550/arXiv.2404.05919