Project Report

Title: Image Classification Using InceptionV3 on Intel Image Dataset
Author: Rakat Murshed
Email: rakatmurshed13@gmail.com

---

1. Introduction

In the field of computer vision, image classification has become a significant area of research and application. The ability to classify images into predefined categories is crucial for various applications such as autonomous vehicles, medical diagnostics, and security systems. This project focuses on the implementation of a Convolutional Neural Network (CNN) using a pre-trained InceptionV3 model for classifying images from the Intel Image Dataset. The primary goal is to leverage transfer learning to improve classification accuracy on a relatively small dataset.

Objective:
The objective of this project is to develop a robust image classification model that can accurately categorize images from the Intel Image Dataset into six different categories: buildings, forest, glacier, mountain, sea, and street.

---

2. Previous Research

Previous research in image classification has extensively utilized Convolutional Neural Networks (CNNs) due to their ability to capture spatial hierarchies in images. Models like AlexNet, VGGNet, and ResNet have set benchmarks in the field, but InceptionV3, introduced by Google, has proven to be one of the most effective models due to its inception modules that allow for the combination of different convolutional filter sizes.

Related Work:

- **AlexNet**: One of the earliest deep CNNs that demonstrated the power of deep learning in the ImageNet challenge.
- **VGGNet**: Known for its simplicity and depth, VGGNet improved classification accuracy by using very small (3x3) convolution filters.
- **ResNet**: Introduced residual learning which made it possible to train very deep networks by addressing the vanishing gradient problem.
- **InceptionV3**: The model used in this project, known for its efficiency in computational cost while maintaining high accuracy, employs a novel approach of factorizing convolutions and using auxiliary classifiers for faster convergence.

---

3. Methodology

Dataset: The Intel Image Dataset used in this project consists of natural scene images classified into six categories: buildings, forest, glacier, mountain, sea, and street. The dataset is split into training and testing sets.

Model Architecture:

- **InceptionV3**: A pre-trained model on ImageNet was used for transfer learning. The top layers of the InceptionV3 model were removed, and custom fully connected layers were added to adapt the model to our specific classification task.
- **Custom Layers**: A Global Average Pooling layer followed by a fully connected layer with 1024 units and ReLU activation, and finally a softmax layer for outputting probabilities across the six classes.

Training:

- **Data Augmentation**: Techniques such as horizontal flipping, zooming, and shearing were applied to the training data to enhance the model's ability to generalize.
- **Optimizer**: Adam optimizer was used due to its efficient handling of sparse gradients on noisy data.
- **Loss Function**: Categorical Crossentropy was used, as it is well-suited for multi-class classification problems.

Training Strategy:

- Initially, the pre-trained layers of InceptionV3 were frozen, and only the custom layers were trained. This approach is beneficial when the dataset is small, allowing the model to leverage learned features from a large dataset like ImageNet.
- In later stages, the top layers of InceptionV3 were unfrozen, and the entire model was fine-tuned with a lower learning rate.

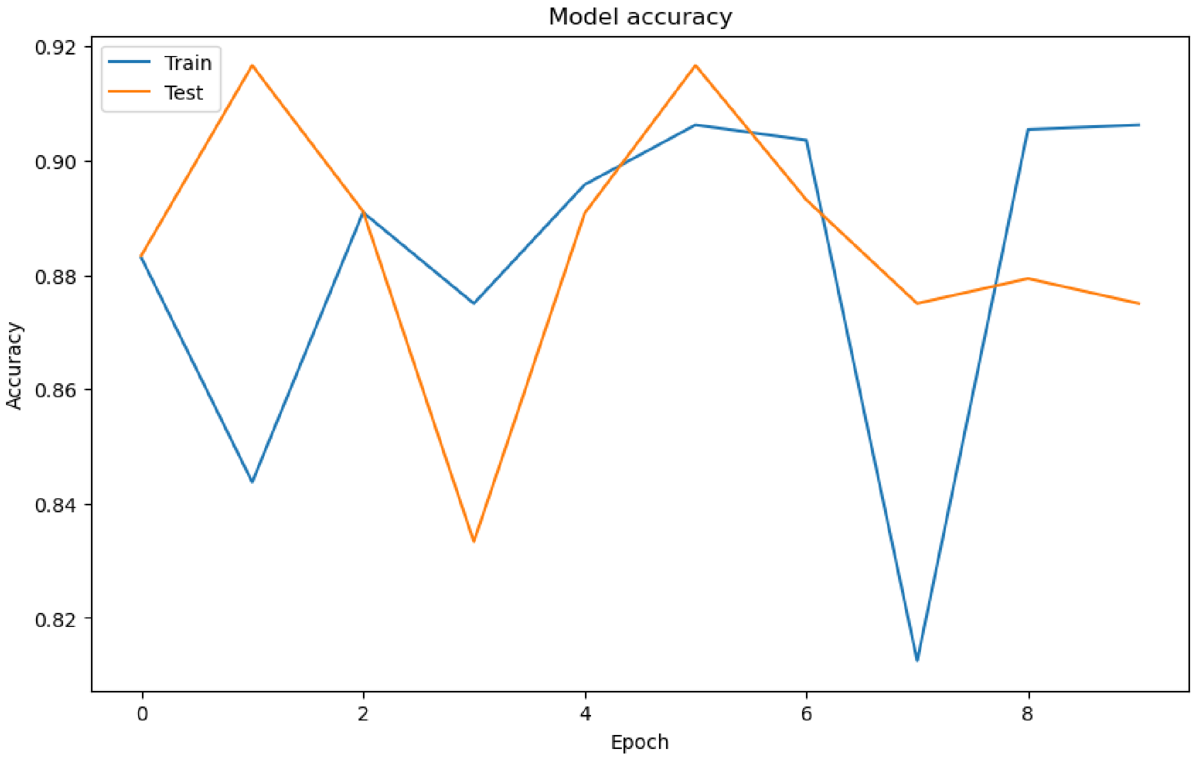
---

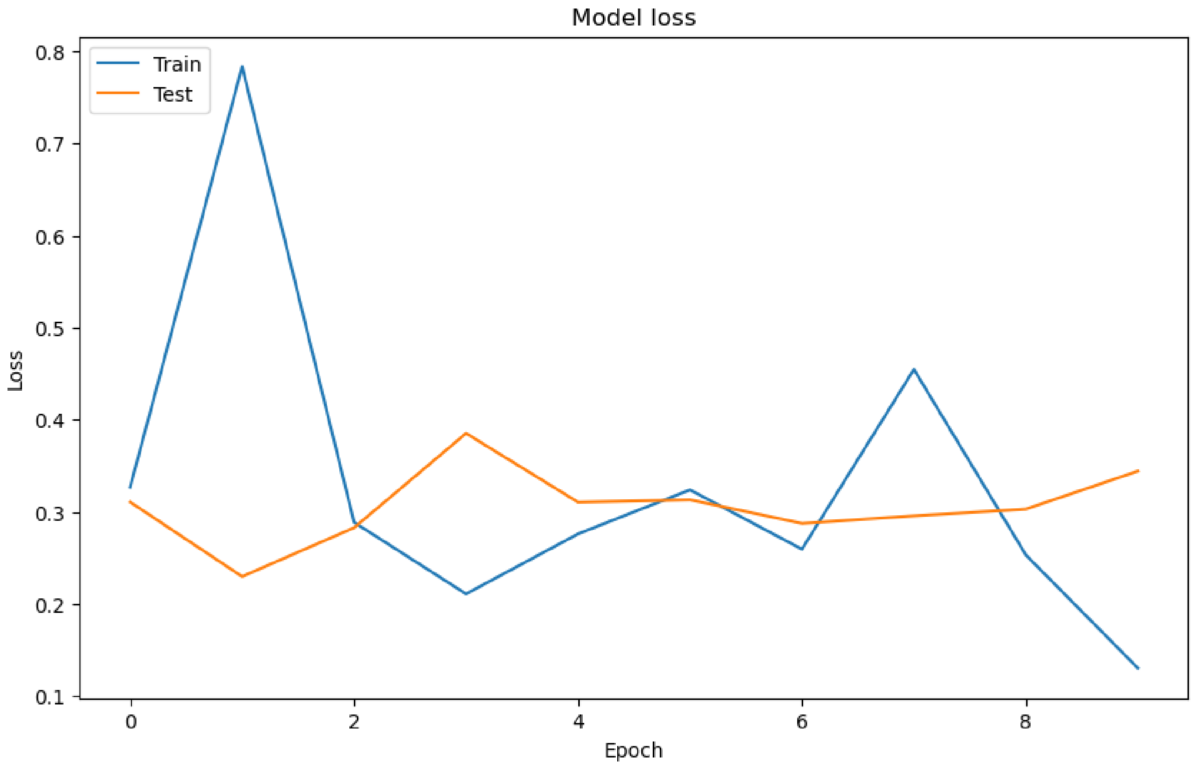
4. Results and Discussion

Model Performance: The model achieved a training accuracy of approximately 78.64% in the initial epoch and showed significant improvement over subsequent epochs, reaching a validation accuracy of 86.80% by the end of the training phase. The loss decreased steadily, indicating good convergence of the model.

Challenges:

- **Data Limitation**: The relatively small size of the dataset posed a challenge, making the model prone to overfitting. Data augmentation helped mitigate this issue to some extent.
- **Overfitting**: Regularization techniques such as dropout and fine-tuning were employed to address overfitting, which could be observed during the training process.





Discussion: The use of transfer learning with InceptionV3 allowed the model to achieve high accuracy even with limited data. This project demonstrated the effectiveness of leveraging pre-trained models for specific tasks where computational resources or data availability is limited.

---

5. Conclusion

This project successfully implemented an image classification model using InceptionV3 on the Intel Image Dataset. The model achieved a high level of accuracy, demonstrating the efficacy of transfer learning in computer vision tasks. The use of pre-trained models like InceptionV3 significantly reduced the training time and improved the model's performance on a relatively small dataset.

Key Takeaways:

- Transfer learning is a powerful tool for improving model performance, especially when working with small datasets.
- InceptionV3's architecture is well-suited for complex image classification tasks due to its ability to capture different levels of abstraction through its inception modules.

---

6. Future Work

Enhancements:

- **Data Augmentation**: Further augmenting the dataset with techniques like rotation, scaling, and color jittering could help improve the model's generalization ability.
- **Ensemble Learning**: Combining the predictions of multiple models, such as ResNet and InceptionV3, could yield better classification accuracy.
- **Deployment**: Deploying the model into a web or mobile application could make it accessible for real-world use cases.

Research Directions:

- **Exploring Other Architectures**: Future work could explore the application of more recent architectures such as EfficientNet or MobileNet for even better performance.
- **Model Explainability**: Integrating techniques to visualize and understand the decision-making process of the model could provide insights into the model's predictions, which is crucial for applications in sensitive domains like healthcare.