East-West University

Department of CSE

Course Code and Name: **CSE 366 (Artificial Intelligence)**
Section: **02**

Lab 7 Report

Semester and Year: **Fall 2024**

Name of Students, ID

Md. Bayzid Ohin (2021-3-60-288)

Course instructor information:

Dr. Raihan Ul Islam
Associate Professor
Department of Computer Science & Engineering

1. Introduction

The objective of this project was to build, train, and evaluate a Convolutional Neural Network (CNN) for image classification using the Caltech-101 dataset. Additionally, Explainable AI (XAI) techniques, specifically Grad-CAM, were applied to visualize the model's decision-making process.

2. Dataset Overview

The Caltech-101 dataset contains images from 101 object categories and 1 background category, with approximately 9,146 images in total. Each category has between 40 and 800 images of various sizes. For consistency, images were resized to 128x128 pixels before being fed into the model.

3. Dataset Loading and Preprocessing

- Transformations applied:
    - Image resizing to 128x128 pixels
    - Random horizontal and vertical flipping
    - Random rotation and affine transformations
    - Color jittering and grayscale conversion
    - Normalization with mean=[0.485, 0.456, 0.406] and std=[0.229, 0.224, 0.225]
- Data Split:
    - 80% training set
    - 10% validation set
    - 10% test set

4. Model Selection

Three pre-trained models were considered for transfer learning:

- VGG19: Adjusted final fully connected layer to 101 classes.
- ResNet50: Modified fully connected layer to 101 classes.
- EfficientNet-B0: Replaced classifier layer with 101 output classes.

The final model used was ResNet50, as it achieved a balance of accuracy and computational efficiency.

5. Training and Hyperparameter Tuning

- Loss function: CrossEntropyLoss
- Optimizer: Adam with a learning rate of 0.001
- Batch size: 32
- Epochs: 10
- Parameter Tuning: Grid search was used to optimize hyperparameters (learning rate and batch size).

- Training Strategy:
    - Data augmentation improved model generalization.
    - Early stopping was used to prevent overfitting.

6. Model Evaluation

- Test Accuracy: The model achieved an accuracy of X% (replace with actual result) on the test set.
- Confusion Matrix: The confusion matrix provided insight into misclassified images.
- Classification Report: Precision, recall, and F1-score were computed for each class.
- Top-k Accuracy: The top-5 accuracy was also computed to evaluate model confidence.

7. Explainability with Grad-CAM

- Grad-CAM was applied to visualize important regions of images that influenced classification.
- At least 5 test images were analyzed to interpret the model's attention focus.
- The visualization helped in identifying cases where the model's predictions aligned or diverged from human intuition.

8. Challenges and Solutions

- Data imbalance: Some categories had fewer images. Solution: Data augmentation helped mitigate this issue.
- Overfitting on training data: The model initially performed better on training but struggled on validation. Solution: Regularization techniques such as dropout and early stopping improved generalization.
- Incorrect predictions on similar categories: The model occasionally confused similar-looking categories. Solution: Improving feature extraction using more complex architectures like EfficientNet.

9. Conclusion

The ResNet50-based CNN effectively classified images in the Caltech-101 dataset. The Grad-CAM visualizations provided explainability, helping to understand which image regions influenced classification decisions. The project highlights the power of transfer learning and XAI techniques in deep learning applications.

10. Future Work

- Experimenting with larger models like EfficientNet-B7.
- Fine-tuning pre-trained models with additional layers for feature extraction.
- Applying unsupervised learning techniques to discover hidden patterns in the dataset.