Proposal for a Comparative Study of Scene Classification Using Decision Trees and Convolutional Neural Networks

Akhil	Mohammad Soroush Ghorbanimehr	Fahimeh Nezhadalinaei
Concordia	Concordia	Concordia
40306047	40282337	40222652
akhilxd1@gmail.com	sorush.gm@gmail.com	fahime.nezhadali@gmail.com

1. Introduction

In this project, we aim to classify scenes into five distinct categories: Airport, School, Hospital, Bakery, and Bowling. We will utilize the Places dataset provided by the MIT Computer Science and Artificial Intelligence Laboratory (CSAIL), available at [1]. The dataset comprises 365 scene categories and includes more than 10 million images categorized into various scene types, making it ideal due to its comprehensiveness, high quality, and wide use in the research community, ensuring robust and generalizable model training.

1. Dataset

We will curate a subset of the Places dataset consisting of 2500 images distributed equally among the five classes (Airport, School, Hospital, Bakery, and Bowling). The images will be divided into three sets: training, validation, and test. Notably, the images within the Places dataset currently exhibit varying dimensions, ranging from 1024x768 pixels to 5184x3456 pixels. To standardize our dataset and optimize model training, we will uniformly resize all images to dimensions such as 224x224 pixels or 256x256 pixels. The distribution of images for each set, post-resizing, will be as follows:

- Training set: 350 images per class (total 1750 images)
- Validation set: 75 images per class (total 375 images)
- Test set: 75 images per class (total 375 images)

3. Models

We will implement and evaluate three different models for this classification task:

- Supervised Decision Tree
- Semi-supervised Decision Tree
- Supervised Convolutional Neural Network (CNN)

4. Evaluation Metrics

To assess the performance of our models, we will use the following metrics based on the confusion matrix:

4.1. Accuracy:

The ratio of correctly predicted instances to the total instances.

4.2. Precision:

The ratio of true positive predictions to the total positive predictions made by the model.

4.3. Recall:

The ratio of true positive predictions to the total actual positives.

4.4. F1 Score:

The harmonic mean of precision and recall, providing a single measure of the model's performance.

5. Implementation Plan

The implementation plan outlines the subsequent actions required for execution.

5.1. Data Collection and Preprocessing:

- Download and curate the Places dataset to obtain 2500 images.
- Split the data into training, validation, and test sets as specified.
- Perform necessary preprocessing steps such as resizing and normalization.

5.2. Model Implementation:

- Implement the supervised decision tree model.
- Develop the semi-supervised decision tree approach.
- Construct and fine-tune the CNN model using transfer learning.

5.3. Training and Validation:

- Train each model on the training set.
- Validate the models using the validation set to tune hyperparameters and avoid overfitting.

5.4. Evaluation:

- Evaluate the models on the test set using the specified metrics.
- Compare the models based on their accuracy, precision, recall, F1 score, and confusion matrix.

5.5. Analysis and Reporting:

- Analyze the performance of each model.
- Generate a detailed report summarizing the findings and providing insights into the strengths and weaknesses of each approach.

6. Conclusion

This project will provide a comparison of different machine learning models for scene classification using the Places dataset. By evaluating the models using multiple metrics, we aim to identify the most effective approach for classifying scenes into the specified categories. The insights gained from this study will contribute to the broader field of image classification and machine learning.

References

[1] http://places.csail.mit.edu/browser.html