COSE474-2024F: Final Project Proposal
"Traffic Sign Recognition for Autonomous Driving Using CLIP"

Name: Kim jieul

1. Introduction

Accurate traffic sign recognition is a crucial component in the development of autonomous driving and intelligent transportation systems. Recently, pre-trained foundation models such as CLIP have recently shown significant potential across various visual tasks by effectively comprehending the associations between text and images.

2. Problem definition & challenges

(1) Problem definition
to develop a traffic sign recognition system using CLIP that achieves high accuracy in classifying a wide variety of traffic signs while operating in real-time and exhibits robustness against various environmental challenges and traffic sign conditions.
(2) Challenges
- Varying sizes and conditions of signs: It is difficult to accurately recognize traffic signs when they appear at different sizes, angles, and lighting conditions.
- Real-time performance: Autonomous driving requires immediate recognition and classification of traffic signs, making low-latency and efficient model performance essential.
- Data imbalance: It occurs when the distribution of classes in a dataset is uneven. In a traffic sign recognition dataset, there might be a large number of images for common traffic signs but very few images for rare signs.

3. Related Works

CLIP has shown remarkable results in image-text matching tasks. Although it has been used widely for image classification and object detection, its usage in specialized tasks such as traffic sign recognition is still limited. This project will extend the research on CLIP to evaluate and improve its effectiveness.

4. Datasets

- GTSRB (German Traffic Sign Recognition Benchmark): A benchmark dataset containing 43 different classes of German traffic signs, widely used in autonomous driving research.
- LISA Traffic Sign Dataset: A dataset containing traffic signs from the U.S., captured under varying angles and conditions, which will help evaluate the model's generalization ability.
- Korean Traffic Sign Dataset (KTS): It includes images of various traffic signs specific to South Korea. It contains annotations for each sign.
- BDD100K: One of the largest public datasets for autonomous driving, consisting of around 100,000 video clips and images.

5. Comparison with SOTA (state-of-the-art)

- SOTA Methods: Convolutional Neural Networks (CNNs) are traditional models such as ResNet and EfficientNet that achieve high accuracy in traffic sign classification, often exceeding 95% on benchmark datasets like GTSRB.
- CLIP Model: By leveraging its zero-shot learning capabilities, the CLIP model is expected to match or exceed SOTA methods in accuracy while providing robustness across various traffic sign conditions.

6. Goals to achieve throughout this project

- Develop and optimize a traffic sign recognition system using the CLIP model.
- Compare the performance of my model with SOTA.

7. Brief/tentative schedule

- Week 1: Data preparation and exploration.
- Week 2: Fine-tuning the CLIP model and establishing a baseline performance.
- Week 3: Model optimization and real-time performance testing.
- Week 4: Final evaluation and comparison with SOTA results.
- Week 5: Report writing and submission of results.