RGB-D Semantic Segmentation based on CNN with Attention Module

Class of 2023 - Haoming(Hammond) Liu
Major: Computer Science
Project Mentor: Professor Li Guo

Abstract

In the computer vision field, semantic segmentation is a fundamental task that assigns each pixel in the image a category label. Related studies can be applied to artificial intelligence industries (e.g. self-driving cars) but with higher security requirements. Most existing semantic segmentation models targeting both RGB and Depth features have not yet exploited the potential of fusing such multimodal features. This project aims to find a more robust way to parse the semantic information from the RGB-D image data by inserting properly designed attention modules to the mainstream CNN models, which could balance the focus between different channels and thus improve the prediction accuracy. The proposed model will be evaluated on published datasets and compared with other state-of-the-art methods by pixel accuracy and mIoU measures.

Research Question & Significance

Semantic segmentation is an interesting and challenging problem in the computer vision field. On the whole, it segments and parses an image into different image regions associated with semantic categories, including background (e.g. wall, floor, ceiling) and objects (e.g. person, picture, television). As shown by previous studies, incorporating Depth features into RGB features is helpful to improve segmentation accuracy. However, most existing methods tend to fuse RGB and Depth features by simply switching backbones, averaging the feature maps, or concatenating channels, which have not yet exploited the potential of fusing such multimodal features. This project aims to find a more robust way to parse the semantic information from the RGB-D image data.

The studies regarding this task can be applied to multiple fields and industries, such as auto-driving, remote sensing, and image processing. To avoid the safety concerns for automatic vehicles or robotics, such models are usually expected to provide rigorous predictions on the semantic understanding of the scene. This undoubtedly requires a higher prediction accuracy for real-life applications and thus shows the significance of related studies. Since the accuracy of the existing methods still has enough room for improvement, this topic is worthy of further research and exploration.

Project Design and Feasibility

This project will follow the mainstream method (i.e. Convolutional Neural Networks) to perform the semantic segmentation task. In particular, the proposed model is expected to capture the contextual dependencies from different channels with different scales. The main difficulty of RGB-D semantic segmentation is how to make use of and associate RGB and Depth features. Most recent researches in natural language processing and computer vision have shown the effectiveness of applying attention mechanisms in balancing the focus. Hence, inserting properly designed attention modules into CNN models could be a feasible way to improve the accuracy of predictions. Overall, the project can be split into three stages:

Stage 1: Paper Reading & Design Brainstorming (Mar. 15th - June 14th)

The first stage aims to build up the foundation for semantic segmentation research. I'll start by collecting and reading papers in this field, which is crucial for understanding the why and how of model designs. Reading materials will include models for both RGB and RGB-D images as well as papers that illustrate attention mechanisms. Hopefully, I'll get more inspiration to design my own model by the end of this stage. I'll share my reading and designing notes with my project mentor as feedback for this stage.

Stage 2: Model Implementation & Adaptation (June 15th - July 14th)

At the second stage, I'll start by going through the coding details of other mainstream models, which is undoubtedly helpful to implement my own models. A proposed model will be evaluated by its performance on published datasets with commonly recognized measures. I'll be continuously adjusting or redesigning the models to get a higher prediction accuracy. The code of these proposed models will be handed to the mentor as feedback for this stage.

Stage 3: Project Report and/or Final Paper (July 15th - Aug. 15th)

When the project is approaching an end, I'll collect all my notes, code, and statistics and form a project report to capture my work. If the performance of the proposed models reaches certain expectations, I'll expand my report into a final paper and seek publications with the help of my project mentor.

Background

I've taken all the required major courses of Computer Science with an A-level grade for each, which reflects my comprehensive level in a sense. Besides, I also took Machine Learning as an elective course last semester. These courses built up the cornerstone for me as an explorer in the world of computer science and made me properly capable of performing this project.

Researching computer vision is my long-term interest. I completed the DURF project "Large-parallax Image Stitching Algorithm based on Sub-plane Segmentation" with the help of Professor Xianbin Gu last year, which enriched my knowledge and skills in this field. Currently, I've already started to read papers about semantic segmentation, such as PSP Net and DANet.

Feedback and Evaluation

The final product of this project will be a semantic segmentation model, which can be evaluated by performing such tasks on some published RGB-D datasets (e.g. NYU Depth V2, and SUN RGB-D). Both pixel accuracy and mIoU (mean Intersection-over-Union) will be used as the evaluation measures while comparing the proposed model with other mainstream methods.

The project receives full supervision from the NYU Shanghai DURF program committee and the project mentor. I'll schedule meetings with the mentor regularly to update the progress and seek professional guidance and suggestions. The final report of this project will include all the useful notes, code, and statistics collected at different stages.

Dissemination of Knowledge

- The whole project will be captured by a project report (and a final paper if possible).
- The code of models will be continuously updated to Github with detailed instructions.
- The work of this project will be presented at the Research Symposium of Fall 2021.