Deep Machine Learning - Road Segmentation

Dion Lim, Adithyaa Ramesh
Chalmers University of Technology
Electrical Engineering Department

CHALMERS UNIVERSITY OF TECHNOLOGY

Introduction

Road segmentation is crucial for various autonomous driving applications. This project focuses on using deep learning techniques to accurately identify and segment road regions from images, and then conducting a comparative study of the model's performance on different datasets.

Figure 1: Road Segmentation Examples

Road segmentation requires classifying every single pixel in the image! Training Convolutional Neural Networks (CNNs) for this task requires large amounts of labeled data, so we leverage a pre-trained model. For each pixel in the image, we need to know its correct label beforehand (road vs non-road). Our approach leverages Convolutional Neural Networks (CNNs) to process images and classify pixels as road or non-road.

Evaluation over datasets

- KITTI dataset: Images of Highways and rural areas. (Shadow) [2]
- Comma10k dataset: Images of Roads in Day and Nighttime conditions. [1]
- CityScapes dataset: Images of urban streets. [5]

Figure 2: KITTI input image

(a) Comma10k input image

(b) CityScape input image

- Metrics: Evaluation will focus on accuracy, Intersection over Union (IoU), and pixel-wise classification.
- Evaluated and compared the performance of the models over different datasets.

Our Model in Action

We visualized our model's prediction after performing a forward pass on the test image above, and compared it with the ground truth image:

Figure 4: Ground truth mask (CityScape)

Figure 5: Model Prediction (VGG trained on CityScape)

Results

Models trained:
- Modified VGG-16 Model trained on CityScape data
- Modified VGG-16 Model trained on comma10k data
- UNET Model trained on Comma10k data

Modified VGG-16 Model

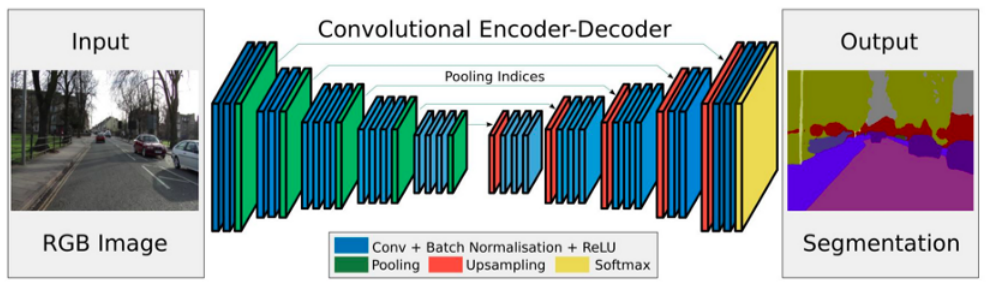

Figure 6: Modified VGG-16 Framework

[3]

The model consists of an encoder-decoder architecture, where the pre-trained VGG-16 model serves as the encoder, extracting features from the input image. At the same time, the decoder is responsible for upsampling these features back to the original image size for segmentation.

UNet

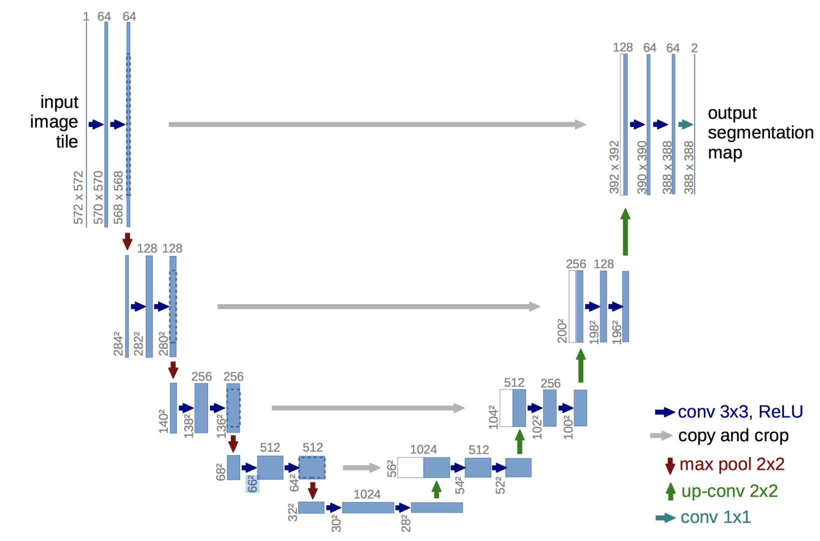

Figure 7: UNET used for Road Segmentation

[4]

This UNet model uses an encoder-decoder setup for image segmentation, with the encoder extracting features and the decoder upsampling them back to the original resolution. Skip connections link encoder and decoder layers to preserve spatial details, which is ideal for tasks like road segmentation.

Comparative Study

Dataset	VGG16 Model Accuracy (CityScape) (%)	VGG16 Model Accuracy (Comma10k) (%)	U-Net Model Accuracy (CityScape) (%)
Cityscapes	95.6	77.463.6	92.5
KITTI	92.0	84.7	87.5
Comma10k	81.0	95.8	91.5

Table 1: Accuracy of VGG and U-Net Models across Datasets

References

[1] C. AI.
Comma10k: A large dataset for autonomous driving, 2020.
Accessed: October 24, 2024.

[2] A. Geiger, P. Lenz, and R. Urtasun.
Are we ready for autonomous driving? the kitti vision benchmark suite.
In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3354–3361, 2012.

[3] S. Overflow.
How to get the encoder from a trained vgg16 network, 2021.
Accessed: October 24, 2024.

[4] A. Schneuwly.
Road segmentation using u-net.
`https://github.com/aschneuw/road-segmentation-unet/tree/master`, 2018.
Accessed: 2024-10-24.

[5] C. Team.
The cityscapes dataset for semantic urban scene understanding.
`https://www.cityscapes-dataset.com/dataset-overview/`, 2016.
Accessed: October 24, 2024.

adithyaa@student.chalmers.se — dionl@student.chalmers.se