ECE 536 Project

Semantic Segmentation in Urban Environments

Camila Chiu (V01047617), Vicky Zhou (V01032392), Paulina Gonzalez (V01045753)

Abstract

In our project, we address the relevant topic of semantic segmentation in urban environments by implementing the deep learning-based semantic segmentation model: DeepLab-v3+, specifically tailored for urban scene understanding. We implement the model in PyTorch, leveraging the Mapillary Vistas Dataset and integrating synthetic data from Virtual KITTI, our work focuses on classifying each pixel in an image into a predefined category such as roads, vehicles, pedestrians, buildings, and vegetation.

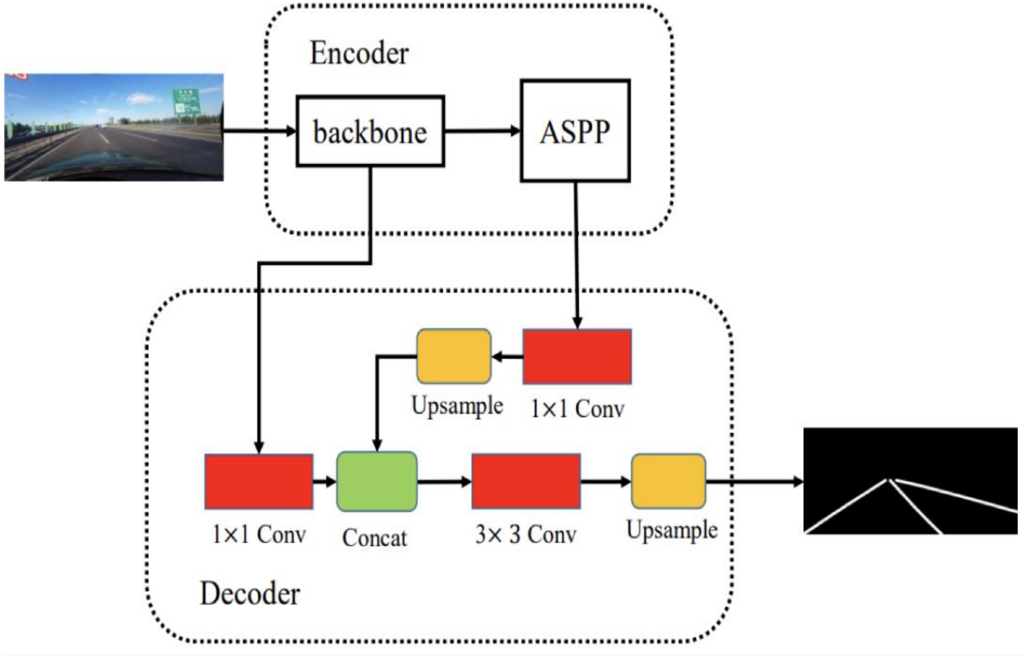
Introduction

Semantic segmentation in urban environments is a crucial task in computer vision which enables machines to understand urban scenes at a pixel level. In urban environments, this task poses challenges due to the diverse range of objects and complex scenes like cars, pedestrians, buildings, roads, and vegetation, each with distinct visual characteristics. By accurately classifying urban objects, our project explores the techniques for applications such as autonomous driving an durban planning.

Framework

DeepLabV3+

- State-of-the-art for semantic segmentation.
- Uses powerful CNN architecture.
- Components: backbone, ASPP, decoder.
- Pixel-wise classification.



Methods

Datasets:

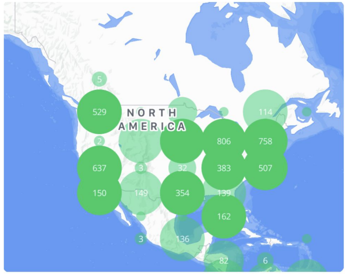
We used a combination of real world and synthetic datasets.

Real-World Dataset (Mapillary Vistas):

- A large-scale street-level image dataset encompassing diverse urban scenes from various cities around the world.
- Contains 25,000 high-resolution images annotated with 124 semantic objects, including roads, vehicles, pedestrians, and more.
- Used for fine-tuning the model and evaluating its performance in real-world scenarios.

Synthetic Dataset (Virtual KITTI):

- A synthetic dataset designed to mimic the KITTI driving dataset, containing rendered images of urban driving scenes.
- Offers pixel-accurate semantic annotations for key urban elements, ideal for pre-training the semantic segmentation model.
- Enabled the model to learn rich representations of urban environments in a cost-effective manner.



Imagery distribution in North America

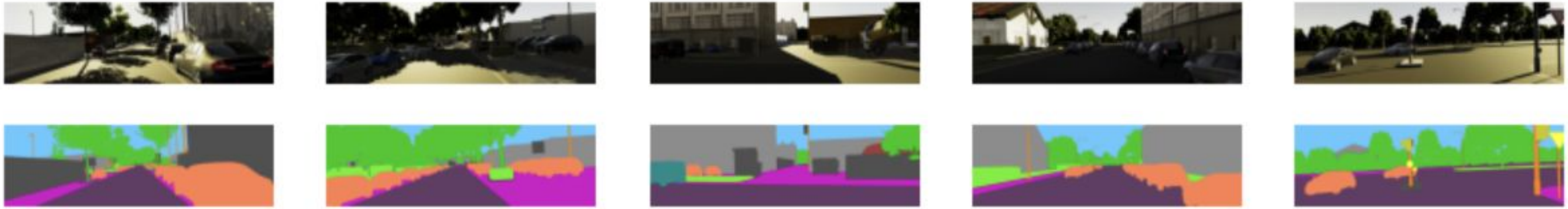
Seattle, United States

Virtual KITTI 2 Dataset

Methods

Labeling

- Developed a strategy for handling color-encoded labels in the datasets, involving development of color to class mapping.
- Each unique color corresponds to a different semantic category in the dataset, giving rise to a need for a systematic approach to convert color-encoded labels into class indices, for calculating the loss and back propagating errors.

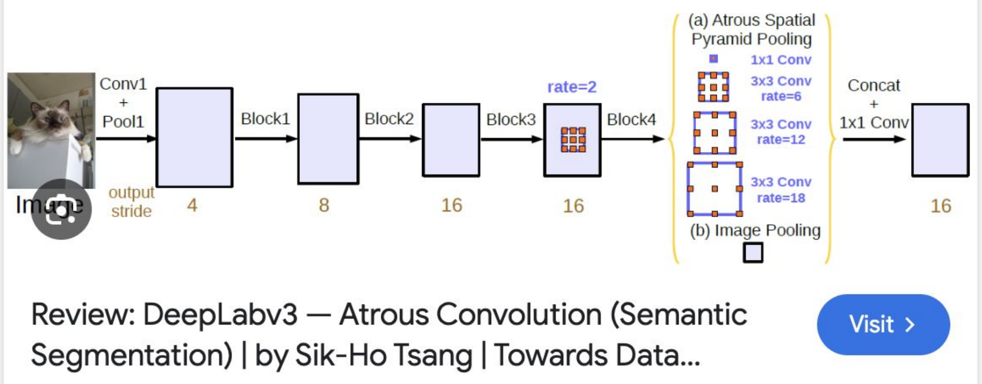


Methods

Model

Semantic Segmentation Model:

- Employed a deep learning-based approach for pixel-wise classification of urban scenes.
- Used a variant of the DeepLab-v3+ architecture, known for its effectiveness in capturing spatial hierarchies and context.
- Integrated advanced neural network techniques, such as atrous convolutions and spatial pyramid pooling, to enhance the model's ability to understand diverse urban elements.



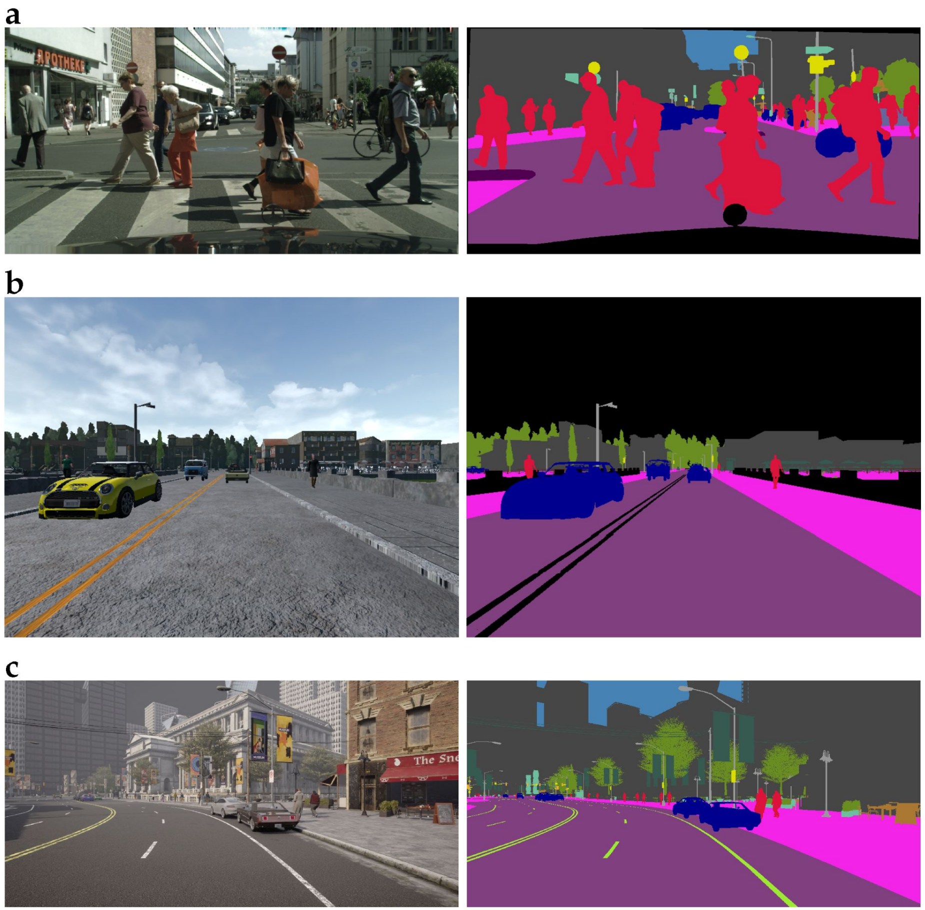
Data Preprocessing:

- Applied transformations including **resizing images** to a uniform dimension (256x256) and **normalizing pixel values** based on the ImageNet dataset's mean and standard deviation.
- Implemented custom data loaders to handle both real and synthetic datasets efficiently.

Methods

Training

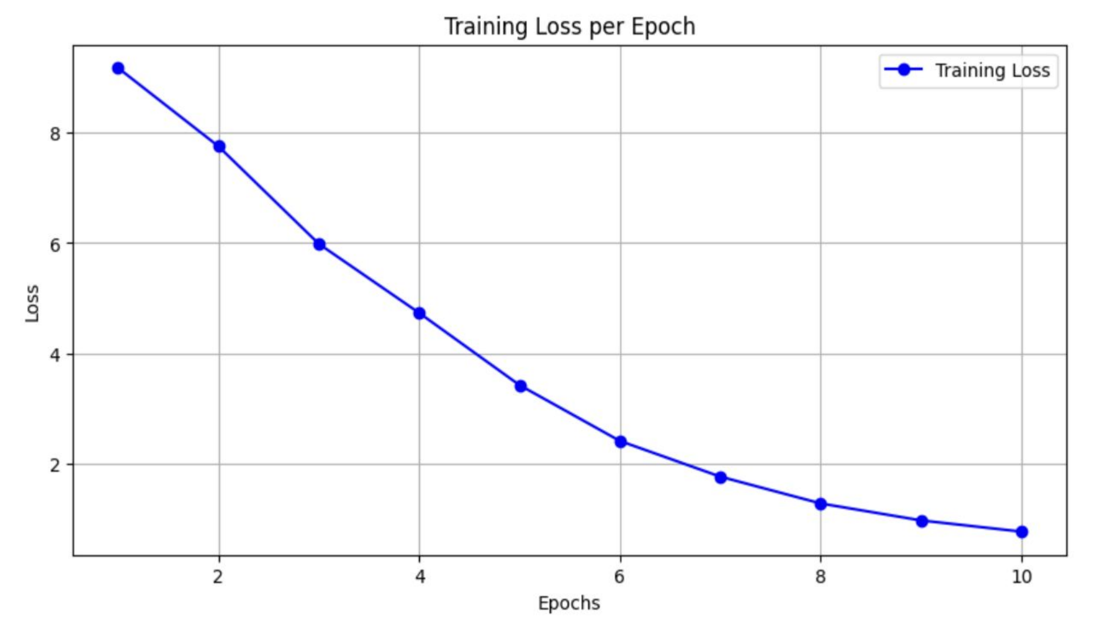
- Opted for a two-phase training approach, starting with pre-training on **synthetic data** to learn general features, followed by fine-tuning on real-world data for improved specificity to urban scenes.
- Established **training loop,** including forward propagation, loss calculation, and backpropagation steps.
- Employed the cross-entropy loss function, optimized using the Stochastic Gradient Descent (**SGD**) algorithm with momentum.



Results

Training Loss Over Epochs

- The loss consistently **decreases** from the first epoch to the last, demonstrating the model's learning process. The significant **drop in loss**, particularly between the first few epochs, suggests rapid initial learning.
- This trend is expected and desired in a well-functioning training loop, where the model progressively **minimizes** the error between its predictions and the actual data.
- We note that training was conducted on a limited subset (10%) of the available dataset due to computational constraints, which may affect the generalization potential of the model.



Results

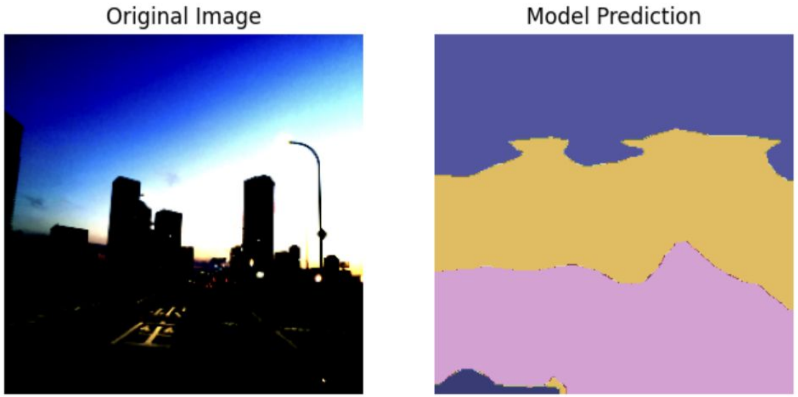
Model Evaluation and Prediction Analysis

Model Evaluation with Real Images:

- We assessed the model's performance quantitatively using a subset of **real-world images.**
- The model was run in evaluation mode (model.eval()) to ensure that batch normalization and dropout layers functioned in inference mode, providing a stable prediction output.

Prediction Generation:

- In the subsequent phase, the model was used to **generate predictions** on a separate set of images that had not been used during training or the initial evaluation phase.
- This step was crucial for understanding how the model generalizes to new, unseen data and provides insight into the model's **real-world applicability.**



Discussion & Conclusions

Results Summary:

- The DeepLab-v3+ model showed promising results in segmenting urban scenes, with a notable learning curve evidenced by decreasing loss values.
- Achieved reasonable pixel accuracy and mIoU on a limited dataset, demonstrating the model's capability in understanding diverse urban elements.
- The results underscore the potential of deep learning in complex urban scene understanding, with room for further improvement.

Future work:

- Expand computational resources to utilize the entire dataset for training and evaluation.
- Explore additional data augmentation strategies to mitigate the effects of image quality variability.

Discussion & Conclusions

Model Challenges

Limited Dataset Usage:

- Due to constraints on storage and computational resources, the model was trained on only 10% of the available datasets.
- This limitation reduced the model's exposure to the full variability and complexity of urban scenes, likely impacting its generalization capabilities.

Image Quality and Lighting Variability:

- The diversity in lighting conditions and occasional low image quality within the dataset posed additional challenges for the model, affecting its ability to consistently recognize and classify objects across different scenes.

Resource Constraints:

- Encountered significant limitations related to Google Colab's computational capacity and runtime restrictions, constraining both the scale of training data and the complexity of models that could be explored.

References & Acknowledgements

[1] A. Das, Y. Xian, Y. He, Z. Akata, and B. Schiele, “Urban Scene Semantic Segmentation with Low-Cost Coarse Annotation.” Accessed: Feb. 25, 2024. [Online]. Available: https://openaccess.thecvf.com/content/WACV2023/papers/Das_Urban_Scene_Semantic_Segmentation_With_Low-Cost_Coarse_Annotation_WACV_2023_paper.pdf

[2] “Mapillary,” www.mapillary.com. https://www.mapillary.com/dataset/vistas

[3] K. O. A. D. R. K. Maksims Ivanovs, "Improving Semantic Segmentation of Urban Scenes for Self-Driving Cars with Synthetic Images," 14 March 2022. [Online]. Available: https://www.mdpi.com/1424-8220/22/6/2252.

[4] N. Klingler, "Image Segmentation with Deep Learning," [Online]. Available: https://viso.ai/deep-learning/image-segmentation-using-deep-learning/.

[5] M. Angella, "Mastering the Art of Image Segmentation: A Comprehensive Guide to Image Segmentation Techniques," 1 August 2023. [Online]. Available: https://medium.com/@1marcusangella/mastering-the-art-of-image-segmentation-a-comprehensive-guide-to-image-segmentation-techniques-545a75010a35.

[6] A. A. F. Harahap, "My First Project in Computer Vision: Cityscape Semantic Segmentation Using UNet," 23 April 2022. [Online]. Available:https://medium.com/@adliaulia16/my-first-project-in-computer-vision-cityscape-semantic-segmentation-using-unet-9d90cb9716b3.

[7] S. Kim, "Segmentation," 13 August 2022. [Online]. Available: https://github.com/Sangh0/Segmentation.

[8] M. Walia, "Semantic Segmentation vs. Instance Segmentation: Explained," 13 October 2022. [Online]. Available: https://blog.roboflow.com/difference-semantic-segmentation-instance-segmentation/.