CASE STUDY #2 — SEMANTIC SEGMENTATION IN AUTONOMOUS DRIVING

Rhichard Koh (100842848)

DECEMBER 12, 2024

FACULTY OF SCIENCE, ENGINEERING & INFORMATION TECHNOLOGY, DURHAM COLLEGE

Advanced Neural Networks: Deep Learning for Spatial Analysis (COSC 41001)

Dr. Noopa Jagadeesh

Table of Contents

Choose a specific use case for semantic segmentation in autonomous driving. Investigate the selected use case and its relevance to autonomous driving.

Semantic segmentation is a computer vision technique used to classify each pixel in an image into predefined categories, such as road, sidewalk, pedestrian, or vehicle. In autonomous driving, this is crucial for detecting drivable areas in urban environments, helping vehicles identify safe and permissible spaces to navigate. Urban settings are particularly complex due to diverse road types, dynamic and static obstacles, and varying environmental conditions. Drivable area detection aids in path planning, obstacle avoidance, and regulatory compliance, utilizing techniques like deep learning models, sensor fusion, and real-time processing. However, challenges such as variability in road appearance, occlusions, and computational demands make implementation difficult. Reliable drivable area detection enhances safety by preventing off-road excursions and handling complex scenarios, improves decision-making with better maneuvering and adaptive speed control, and enhances passenger experience with smoother rides and increased trust. Additionally, it ensures regulatory compliance and ethical navigation. As urban populations grow and demand for autonomous solutions rises, advancements in machine learning, sensors, and hardware continue to refine semantic segmentation, making it essential for the future of autonomous driving in urban environments.

Discuss how semantic segmentation helps solve challenges associated with your chosen use case.

Drivable area detection in urban environments is challenging due to the dynamic and cluttered nature of city landscapes. Semantic segmentation plays a crucial role by classifying each pixel of an image, enabling autonomous vehicles to accurately interpret their surroundings. Urban roads vary widely in surface texture, markings, and environmental conditions, while dynamic obstacles like pedestrians and occlusions from large vehicles further complicate navigation. Semantic segmentation addresses these issues through robust feature extraction, temporal consistency, and multi-sensor data integration. It aids in understanding complex urban layouts, adapting to real-time changes, and efficiently processing high-resolution data using lightweight architectures and hardware acceleration. Additionally, it enhances safety by minimizing errors, identifying hazards early, and supporting adaptability through continuous learning. Semantic segmentation also facilitates advanced features like semantic map building and interaction prediction. By tackling challenges like construction zones and rare edge cases, it improves the reliability, efficiency, and safety of autonomous vehicles, making it an indispensable tool for navigating urban environments.

Include examples of existing solutions or research in this area.

Semantic segmentation for drivable area detection is essential in autonomous driving, enabling vehicles to navigate complex urban environments safely. Significant advancements in this field come from both academia and industry. Research efforts like Google's DeepLab series, PSPNet, and AdapNet have introduced powerful models that improve segmentation accuracy and adaptability to varied road conditions. Industry leaders such as Tesla, Waymo, and NVIDIA DRIVE leverage these advancements to develop robust systems for real-time drivable area detection, combining vision-based approaches with multi-sensor fusion and high-definition mapping. Datasets like Cityscapes, KITTI, and ApolloScape provide critical benchmarks for training and evaluating models, while tools like CARLA simulator enable virtual testing of algorithms in realistic environments. Emerging techniques, including self-supervised learning, transformer-based models, and domain adaptation, promise further enhancements in handling rare scenarios and new domains. Collaborative efforts between academia and industry ensure ongoing progress, making semantic segmentation a cornerstone of autonomous vehicle technology.

Identify challenges specific to the use case (e.g., occlusion, class imbalance, environmental variations).

Semantic segmentation for drivable area detection is essential for autonomous vehicles to safely and efficiently navigate urban environments, yet it poses significant challenges due to the complexity and variability of city settings. Occlusions, class imbalances, environmental variations, and dynamic objects like pedestrians and cyclists are major obstacles that require advanced modeling techniques, such as temporal inference and multi-sensor fusion. Variations in road conditions, lighting, weather, and clutter further complicate segmentation accuracy. Real-time processing constraints, high-resolution data requirements, and the need for robust integration with other systems add to the technical difficulties. Moreover, labeling and dataset diversity present significant hurdles due to the labor-intensive nature of pixel-level annotations and the need to address edge cases, rare events, and adversarial conditions. Ensuring safety, regulatory compliance, and model explainability is critical, especially in high-stakes environments. Addressing these challenges involves leveraging data augmentation, efficient algorithms, continuous learning, and rigorous testing under diverse conditions. Collaborative efforts among engineers, researchers, policymakers, and urban planners, alongside the development of robust models and ethical frameworks, are essential to advancing the reliability and safety of autonomous driving systems.

Objective and Use Case Description

The use case chosen focuses on semantic segmentation of driving scenes to identify drivable areas within urban environments, a critical task for autonomous vehicles (AVs) and advanced driver-assistance systems (ADAS). The goal is to accurately differentiate between regions where a vehicle can safely travel, such as the main roadway and alternative drivable areas, and non-drivable regions like sidewalks, curbs, or obstacles. The primary objective is to develop a deep learning model that segments driving scene images into direct drivable areas, alternative drivable areas (e.g., shoulders or bicycle lanes), and non-drivable backgrounds. Secondary objectives include addressing class imbalance to ensure effective learning of minority classes, evaluating the performance of a segmentation architecture (LinkNet with a ResNet34 encoder) on a dataset subset, and identifying strategies for future improvements. Semantic segmentation is essential for autonomous systems to understand free-space and lane structure, enabling safe navigation. However, challenges such as class imbalance, visual complexity due to weather, lighting, and diverse textures, and the need for real-time, low-latency inference must be overcome to enhance vehicle path planning, hazard avoidance, and overall safety.

Methodology

The project utilizes a subset of the BDD100K dataset, focusing on drivable area segmentation labels where each pixel is classified as “direct drivable area” (class 0), “alternative drivable area” (class 1), or “background” (class 2). A key limitation of the dataset is class imbalance, as background areas dominate while drivable areas and some alternative areas are underrepresented. Additionally, the dataset, though large, may not cover all global road conditions, such as extreme weather or rare terrains. To address these challenges, inverse class frequency weighting was applied in the loss function to reduce bias, and data augmentation techniques are recommended for future iterations. The chosen model, LinkNet with a ResNet34 encoder, offers a balance between performance and speed, making it suitable for real-time tasks like autonomous driving. The ResNet34 encoder, pre-trained on ImageNet, ensures robust feature extraction, while LinkNet’s efficient decoder reconstructs spatial resolution effectively. Preprocessing involved resizing images and masks to a consistent dimension (e.g., 736x1280), normalizing images to a [0,1] range, and applying class weighting in the loss function. Training employed the Adam optimizer with a learning rate scheduler and early stopping to prevent overfitting, using weighted cross-entropy loss to emphasize minority classes. Future iterations aim to enhance regularization through advanced augmentation techniques.

Results and Discussion

The evaluation metrics for the project included Cross-Entropy Loss, used during training and validation to guide optimization and early stopping, and Intersection over Union (IoU), which measured the overlap between predicted segments and ground truth areas. IoU, a critical benchmark in segmentation tasks, was computed for each class and averaged for a mean IoU score. The best model achieved a Train Loss of 0.2503, Validation Loss of 0.2985, and a mean IoU of 0.6285 on the test dataset, demonstrating reasonable segmentation performance. The model effectively segmented direct drivable areas due to their dominance in the training set but struggled with alternative drivable areas, which had fewer examples and greater variability. Early stopping was triggered at the 14th epoch after a patience of 5 epochs, preventing overfitting.

The model's strengths included robust feature extraction from the ResNet34 encoder and a lightweight LinkNet architecture, making it efficient for real-time applications. However, limitations such as class imbalance and limited data variability reduced its performance on minority classes and rare scenarios. To address these, data augmentation techniques to simulate diverse conditions, advanced sampling strategies, and exploring more sophisticated architectures like Deeplab v3+ or HRNet could improve results.

Ethically, the model must be evaluated for bias and fairness, ensuring it generalizes well across diverse regions and populations. Safety remains paramount, as segmentation errors in critical contexts could lead to accidents, necessitating thorough testing and regulatory approval. Privacy concerns related to street-level imagery also require adherence to data anonymization and privacy regulations.

References

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation.

Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2017). DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs.

He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask R-CNN.

Zhao, H., Shi, J., Qi, X., Wang, X., & Jia, J. (2017). Pyramid Scene Parsing Network.

Xu, Y., Gao, T., Chen, K., & Zhou, W. (2020). SqueezeSegV3: Spatially-Adaptive Convolution for Efficient Point-Cloud Segmentation.

Mohan, R., & Chen, Y. (2020). EfficientPS: Efficient Panoptic Segmentation.

Mendes, A., & de Carvalho, P. (2018). "Vision-Based Road Detection Using Deep Learning." International Journal of Image and Data Fusion.

Mobileye's Technology Overview and Publications.

Apollo Platform Documentation and GitHub Repository.

Dosovitskiy, A., Ros, G., Codevilla, F., Lopez, A., & Koltun, V. (2017). "CARLA: An Open Urban Driving Simulator."

Yu, F., et al. (2018). "BDD100K: A Diverse Driving Video Database with Scalable Annotation Tooling."

Zhai, S., et al. (2019). "S4Net: Self-Supervised Semantic Segmentation."

Zheng, S., et al. (2021). "Rethinking Semantic Segmentation from a Sequence-to-Sequence Perspective with Transformers."

Hoffman, J., et al. (2018). "CyCADA: Cycle-Consistent Adversarial Domain Adaptation."

Audi's Research and Development Publications.

Argo AI's Technical Blogs and Whitepapers.

https://doc.bdd100k.com/format.html#seg-mask