ECE 579 Intelligent Systems, Fall 2024

Final Project Report

Project Title: Image Segmentation for Autonomous Driving in Diverse Weather and Low-Light Conditions

Student names: Jaskirat Singh Sudan, Shubham Sunil Jagtap, Sreeshma Inti
Department name: Department of Electrical and Computer Engineering, University of Michigan-Dearborn

Responsibilities of each student:

a) **Jaskirat Singh Sudan:** Model development, transfer learning, fine-tuning models, model validation, and data pipeline creation.
b) **Shubham Jagtap:** Dataset preprocessing, testing alternate models, data collection, evaluation metrics, and documentation.
c) **Sreeshma Inti:** hardware compatibility checks, testing, and documentation.

1. Introduction:

- The initiative is to make the autonomous vehicle more capable of distinguishing and recognizing objects, such as roads, other vehicles, and pedestrians, in challenging situations: nighttime operation and poor weather conditions. This competence is fundamental for ensuring safety and dependability in self-driving systems.
- Semantic segmentation is a part of the core program in autonomous driving where a significant morphology of images is presented to vehicles for understanding what it should take action on behalf of. This is very poor as a model in scenarios like rain, fog, snow, and low light. The usual way around is pre-trained deep learning models. In pretraining, one trains models on enormous datasets such as ImageNet and COCO for some generic feature representations, while fine-tuning is again training them with task-specific datasets. Here, we fine-tune models meant for segmenting objects in complicated scenes using the BDD100K dataset, which provides rich driving scenarios in day and night conditions.
- This paper investigates the application and evaluation of three pre-trained models, namely Xception U-Net, DeepLabV3+, and MobileNetV2 U-Net, for semantic segmentation tasks. First, they were all trained on ImageNet to learn feature extraction. After that, fine-tuning was done on the BDD100k dataset, which provides quality annotations both for daytime and nighttime driving scenarios. The approach consisted of thorough preprocessing, cleaning the dataset, augmenting, and normalizing the data, followed by training using targeted loss functions like Dice Loss and Binary Cross-Entropy for better segmentation performance. The paper elaborates on the structural design for every model, the fine-tuning approach employed on each one, and an evaluative comparison evaluation as well based on metrics, such as accuracy, IoU, Dice Coefficient, and Precision-Recall. The results indicate the optimal model for improving the performance of autonomous vehicles in real-world, highly challenging cases.

2. Description of Technologies Related to Project

- Our project uses a combination of some well-established techniques and current advances in computer vision and deep learning for image segmentation in autonomous driving under adverse weather conditions and low-light conditions. The next sections explain the main technologies and methods of such work.
- **Convolutional Neural Networks (CNN):** Among the innovations that have transformed modern paradigms in image segmentation, CNNs stand out. U-Net, DeepLab, and Fully Convolutional Networks (FCNs) are few models that are extraordinary at modeling spatial features and comparatively trending in segmentation tasks. They have encoder-decoder architectures to ensure integrated feature representation as well as superior resolution outputs.
- **Pretrained Models and Transfer Learning:** Pretrained models like MobileNetV2, ResNet, and Xception, trained on large-scale datasets such as ImageNet, provide a strong starting point for segmentation tasks. Transfer learning uses these pre-trained models by allowing fine-tuning on task-specific datasets, such as BDD100k, which enables adaptation to challenging applications, including low-light and weather-affected environments.

- **Loss Functions:** The efficacy of the loss function is very significant for segmenting training models, though Binary Cross-Entropy and Categorical Cross-Entropy are the most traditionally used loss functions. Many advanced alternatives, like Dice Loss and Focal Loss, have also been created from the development of modern to cope with class imbalance and improve segmentation performance further.
- **Data Augmentation:** Training data are made richer and therefore more varied for easy generalization of models through certain data-augmentation methods, such as flipping, rotation, brightening, and zooming. These techniques become more important when training the model to get the best results in varying conditions.
- **Evaluation Metrics:** Performance evaluation is an indispensable part of all semantic segmentation approaches. Some of the metrics used for its measurement include: Accuracy, Intersection over Union (IoU), Dice Coefficient, Precision, and Recall.

3. **Methods used in project:**

- Work Description: To face the challenge of segmenting objects under different weather and light conditions, three highly effective semantic segmentation models were employed.
- Xception U-Net: Employed Xception as the encoder, with a custom decoder for segmentation. Pre-trained on ImageNet and fine-tuned on BDD100k daytime images, and then again on nighttime images.
- DeepLabv3: It was built on a ResNet-101 backbone pre-initialized with COCO data, with atrous convolution and spatial pyramid pooling techniques incorporated to extract multi-scale features. The model utilized Focal Loss in remedying class imbalance and intensification of small object detection.
- MobileNetV2 U-Net MobileNetV2 encoder pretrained on ImageNet was utilized for lightweight real-time segmentation tasks, mainly with BDD100k.

Preprocessing and Data Augmentation:

- Image preprocessing involved scaling all images and masks to a size of 256x256 and normalizing pixel values.
- To further enhance generalization and robustness, we used various augmentation techniques, which consisted of horizontal flipping, brightness adjustment, rotation, and zooming.
- The techniques were used to cope with the basic variability of environmental conditions that influence the performance of the model in real-life scenarios, such as low light and inclement weather.

Training and Fine-Tuning:

- To establish a good baseline, initial training of all model was done using daytime images from the BDD100k dataset.
- All model fine-tuning was done over nighttime imagery to specialize the models in low-light conditions.
- Dice loss, binary cross-entropy, and focal loss have previously been used for accuracy improvements in segmentation and class imbalance.

a. **Evaluating Results:**

i) Metrics used:

- **Accuracy:** Percentage of correctly classified pixels.
- **Intersection over Union (IoU):** Measures overlap between predicted and ground truth masks.
- **Dice Coefficient**: Evaluates segmentation similarity, particularly for imbalanced datasets.
- **Precision and Recall:** Assess false positives and false negatives, respectively.

ii) Model Comparison:

- **Xception U-Net**: Achieved balanced accuracy, but the model required heavy adjustments for nighttime and fine-tuning improved nighttime performance, although it still faced difficulties in detecting small objects.
- **DeepLabV3+**: Achieved highest IOU and Dice Coefficient among all compared methods; it outperformed others in multi-scale segmentation. However, its heavy computational intensity made it less practical in real-time applications.
- **MobileNetV2 U-Net**: Showed competitive accuracy at the same time having a considerably lowered demand for computational resources, making it viable in resource-constrained settings. However, its effectiveness was slightly lowered in finding smaller objects.

iii) Limitations:

- Small object segmentation was difficult, particularly under adverse conditions, such as nighttime scenes with clutter.
- Low-light conditions required the models to be redefined considerably, showing how complex these environments can be.

iv) Results: Insights Gained:

- **DeepLabV3+** model exhibited enhanced performance attributed to its sophisticated architecture; however it underscored the inherent trade-off between accuracy and computational efficiency.
- **MobileNetV2 U-Net** was the most feasible in real-time deployment, which signifies the importance of lightweight models in the application of autonomous driving.
- **Xception U-Net** offered a middle ground, balancing accuracy and complexity, with strong adaptability through fine-tuning.

4. Experiments:

a) **Data:** Training and evaluating the models is done with the 100,000-video set from BDD100K, one of the largest publicly available driving video datasets in the world. It contains 100,000 videos at 720p with 30 frames per second, depicting diverse real-world driving scenarios corresponding to different geographic locations, including New York and the San Francisco Bay Area; illumination: daytime, nighttime, dusk; and weather conditions like rain or fog. Our data set also includes dense pixel-level annotations for 19 object classes, which are combined into binary masks for vehicle segmentation (foreground vs. background).For the experiments:

- **Training Set**: 92 samples (80% of total images)
- **Validation Set**: 23 samples (20% of total images)
- The data underwent extensive **preprocessing** steps, including resizing the images to 256x256, normalization, and data augmentation techniques (horizontal and vertical flipping, random rotation) to increase model robustness and generalize better across various conditions.

b) Experiments Conducted:

Three different models were used for image segmentation tasks: DeepLabV3+, MobileNetV2, and Xception Unet. These models were trained using transfer learning (for DeepLabV3+ and MobileNetV2) and fine-tuning (for MobileNetV2 and Xception Unet). The following experiments were conducted to evaluate their performance in segmenting vehicles in different environmental conditions:

- **DeepLabV3+**:
  a) **Pretrained on ImageNet and COCO dataset**: The model was trained with the original COCO dataset for feature extraction, followed by transfer learning on the BDD100K dataset. The classification layer was modified to output binary segmentation (foreground vs background).
  b) Metrics Collected: Accuracy, IoU, Dice Coefficient, Precision, Recall.
- **MobileNetV2**:
  a) **Pretrained on ImageNet**: The MobileNetV2 model used the same procedure, applying transfer learning on BDD100K for vehicle segmentation tasks. Fine-tuning was performed using nighttime images to adapt the model to low-light conditions.
  a) **Metrics Collected**: Accuracy, Dice Loss, Binary Cross-Entropy
- **Xception Unet**:
  a) **Custom Model**: The Xception Unet architecture, with a pre-trained Xception backbone, was used for segmentation. Fine-tuning was performed to optimize the model for accurate vehicle segmentation.
  b) **Metrics Collected**: Accuracy, Dice Loss, Binary Cross-Entropy

c) Present and Discuss Your Results:

- The results from the experiments show significant improvements in the performance of the models after applying transfer learning and fine-tuning.

DeepLabV3+:

- The model performed decently in terms of **accuracy (97.34%)** and **Dice Coefficient (0.7378)**. However, the **precision** and **recall** values were lower, suggesting that there were still issues in correctly classifying foreground pixels, especially under difficult lighting conditions.2

MobileNetV2:

- Fine-tuning on nighttime images improved the model's performance, achieving **90.4% accuracy** with a **Binary Cross-Entropy of 0.197**. Despite a **Dice Loss of 0.92**, indicating challenges in distinguishing foreground from background in complex scenes, the significant accuracy boost demonstrates the model's effectiveness in low-light conditions.

Xception Unet:

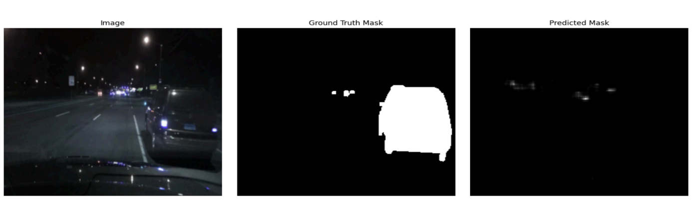
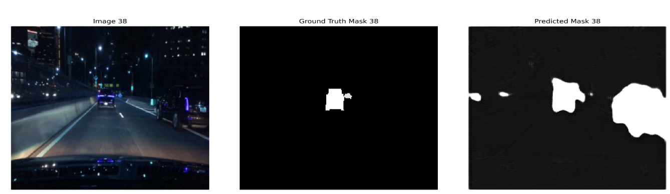
- **The Xception Unet** delivered outstanding results with **99.5% accuracy** and a **Dice Coefficient of 0.8999**, highlighting its excellent segmentation quality. Its low **Binary Cross-Entropy (0.0129)** and **Dice Loss (0.1001)** confirmed its superior ability to distinguish vehicles from the background, making it the most reliable model across diverse conditions.

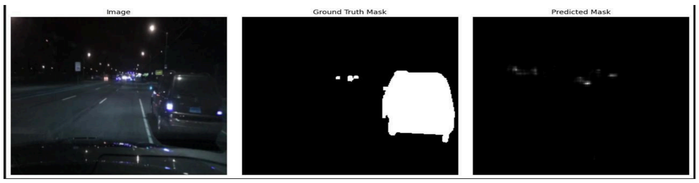
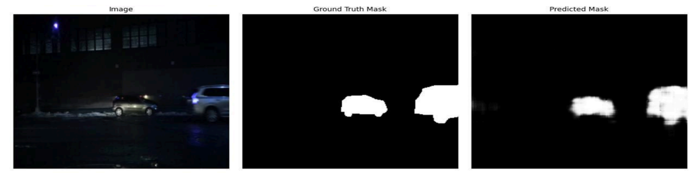
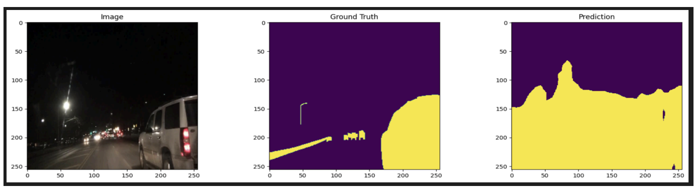
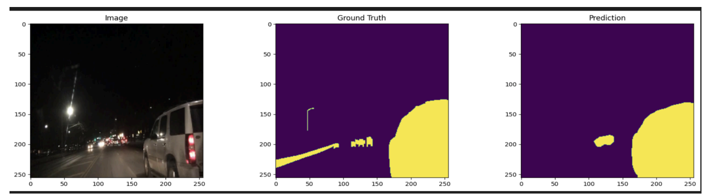
d) Discussion:

- **Model Comparison**: The **Xception Unet** model showed a significant advantage in both **accuracy** and **segmentation quality**, outperforming both **DeepLabV3+** and **MobileNetV2**. This suggests that the architecture's deeper and more sophisticated feature extraction, combined with skip connections, makes it better suited for the complex task of vehicle segmentation under varying conditions.
- **Challenges**: Despite the good performance of **MobileNetV2** and **DeepLabV3+**, challenges such as **vehicle occlusions**, **partial visibility**, and **background interference** still exist. Fine-tuning further and incorporating more diverse data might improve model performance in edge cases.
- **Transfer Learning Impact**: Transfer learning helped in adapting pre-trained models like **DeepLabV3+** and **MobileNetV2** to specific conditions present in the BDD100K dataset. However, **Xception Unet** showed that a custom architecture, designed for the task, could outperform pre-trained models in terms of segmentation quality.

e) Present and Discuss Your Results

Metric	DeepLabV3+ (Post-FT)	MobileNetV2 (Post-FT)	Xception Unet (Post-FT)
Accuracy (%)	97.34	90.40	99.50
Dice Loss	-	0.92	0.1001
Binary Cross-Entropy	-	0.197	0.0129
IoU	0.5852	-	-
Dice Coefficient	0.7378	0.08	0.8999

Results (MobileNetV2) - Daytime Image	Results (MobileNetV2) - Night Time Images (After Fine-Tuning)
	
Results (Xception Unet) - Nighttime Images (Before Fine-Tuning)	Results (Xception Unet) - Nighttime Images (After Fine-Tuning)

	
Result (DeepLabV3+) : Nighttime Images (Before Fine Tuning)	Result (DeepLabV3+) : Nighttime Images (After Fine Tuning)
	

5. Conclusion:

This study focused on improving image segmentation for autonomous vehicles in challenging environments, including low-light and adverse weather conditions. The goal was to develop robust models capable of accurately distinguishing vehicles from their surroundings, enhancing perception and safety in autonomous systems. Three architectures were evaluated: **DeepLabV3+**, **MobileNetV2**, and **Xception Unet**. Using the BDD100K dataset—characterized by diverse lighting, weather conditions, and varying vehicle sizes and occlusions—we applied transfer learning with fine-tuning techniques and distilled the models. Among the tested architectures, **Xception Unet** outperformed the others, achieving a 99.50% accuracy and the highest Dice Coefficient, demonstrating superior pixel-level segmentation performance in complex environments. While **DeepLabV3+** and **MobileNetV2** delivered reasonable results, **Xception Unet** proved to be the most reliable due to its strong feature extraction capabilities, effective segmentation, and low loss values. Key insights highlight the importance of model architecture, pre-trained models, and fine-tuning on relevant datasets in enhancing segmentation performance. Future research could explore additional environmental variables and novel architectures to further improve accuracy. This work underscores the significance of advanced segmentation techniques in autonomous driving and contributes to optimizing both datasets and models, increasing system reliability under challenging conditions.

6. References:

[1] J. Long, E. Shelhamer, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 4, pp. 640–651, April 2017.

[2] A. Howard, M. Zhu, B. Chen, D. K. Li, and H. G. Adam, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," *arXiv preprint arXiv:1704.04861*, 2017.

[3] C. Chou, Y. Xu, and X. Yang, "Xception: Deep Learning with Depthwise Separable Convolutions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 8, pp. 1874–1888, August 2018.

[4] J. Sun, L. Zhang, and S. Liu, "Vehicle Detection and Recognition in Complex Traffic Scenarios Using Deep Learning Models," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 8, pp. 3124–3135, August 2020.

[5] B. Liu, L. Wei, and P. Hu, "Vehicle Segmentation in Adverse Weather Using Deep Learning Models," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 5, pp. 4928–4939, May 2020.

[6] Berkeley DeepDrive, "BDD100K: A Large-Scale Driving Video Dataset," *https://bdd-data.berkeley.edu/*, Accessed: Dec. 2024.

[7] G. Jégou, M. Drozdzal, D. Vazquez, D. Romero, and P. V. Gool, "The One Hundred Layers Tiramisu: Fully Convolutional DenseNets for Semantic Segmentation," *arXiv preprint arXiv:1611.09326*, 2016.