

UNDERGRADUATE PROJECT PROPOSAL

|  |  |
| --- | --- |
| **Project Title:** | **Semantic Segmentation on Deep Learning Applications** |
| **Surname:** | **Feng** |
| **First Name:** | **Xiang** |
| **Student Number:** | **202018010119** |
| **Supervisor Name:** | **Dr Grace Ugochi Nneji** |
| **Module Code:** | **CHC 6096** |
| **Module Name:** | **Project** |
| **Date Submitted:** |  |

**Table of Contents**

[1 Introduction 3](#_Toc117467119)

[1.1 Background 3](#_Toc117467120)

[1.2 Aim 3](#_Toc117467121)

[1.3 Objectives 3](#_Toc117467122)

[1.4 Project Overview 4](#_Toc117467123)

[1.4.1 Scope 4](#_Toc117467124)

[1.4.2 Audience 4](#_Toc117467125)

[2 Summary of Related Literature](#_Toc117467126) 5

[3 Methodology 7](#_Toc117467129)

[3.1 Approach 7](#_Toc117467130)

3.1[.1 Semantic Segmentation...................................................................................... 7](#_Toc117467124)

[3.1.2 Dataset](#_Toc117467124)................................................................................................................ 7

[3.1.3 MIoU](#_Toc117467124) ..................................................................................................................... 7

[3.1.4 LinkNet](#_Toc117467124) .................................................................................................................. 7

[3.1.5 U-Net](#_Toc117467124) ..................................................................................................................... 8

[3.1.6 SegNet..................................................................................................................](#_Toc117467124) 9

[3.1.7 Model use in this project. .................................................................................](#_Toc117467124) 9

[3.2 Technology 10](#_Toc117467131)

[3.3 Version management plan](#_Toc117467132) 10

[4 Project Management 11](#_Toc117467133)

[4.1 Activities 11](#_Toc117467134)

[4.2 Schedule 11](#_Toc117467135)

[4.3 Data management plan](#_Toc117467136) 12

[4.4 Deliverables](#_Toc117467137) 12

[5 References](#_Toc117467138) **13**

# **Introduction**

## **Background**

Today's cities are challenged by traffic congestion and frequent accidents, and autonomous vehicles are considered as a potential solution to reduce accident rates and improve traffic flow [1]. However, in urban environments, we need to deal with complex backgrounds, variable weather and various traffic scenarios, which are difficult for traditional computer vision techniques to handle [2].

Therefore, deep learning-based semantic segmentation techniques become especially critical, which can associate each pixel in scenes to semantic categories such as roads, buildings, vehicles, pedestrians, etc [3]. This provides a powerful environment perception tool for autonomous driving [4], helping vehicles better understand and adapt the complex road environment.

This project is devoted to the in-depth study and application of semantic segmentation techniques to create high-precision real-time models to promote safer and more efficient autonomous driving technologies for urban transportation systems, contributing to the sustainable development of cities.

## **Aim**

Semantic segmentation can accurately classify each pixel in an image to improve the accuracy of environment perception, which has important applications in areas such as autonomous driving. This project aims to improve the model performance and generalization ability through integrated learning to achieve high-precision real-time road environment detection. This helps to achieve highly reliable autonomous driving and is expected to contribute to the development of urban transportation.

## **Objectives**

* When the project begins, the first task will be to study the relevant literature in-depth to understand the latest advances and breakthroughs in the area of semantic segmentation. This will help to understand the performance and efficiency of various advanced neural network models in the task.
* Next, choice some suitable model for semantic segmentation. Since different models have different performance in different scenes, three network models such as LinkNet, U-Net and SegNet will be selected for integrated learning to improve their generalization ability and performance.
* During the project period, model performance will be monitored, including steps such as weight assignment, integration strategy adjustment, and hyper-parameter fine-tuning, to ensure that the synergy between models is maximized.
* Finally, model performance will be evaluated using appropriate test datasets and multiple evaluation metrics. To ensure that the integrated semantic segmentation model meets the desired results and standards and that the model has high accuracy and real-time performance.

## **Project Overview**

### **Scope**

CNN are good at extracting features in semantic segmentation tasks, but generalization ability and performance of single CNN model are relatively weak. To solve it, an integrated learning method is used to improve the performance for semantic segmentation of the dataset by taking full advantage of the strengths of different neural networks as well as the complementary among them. Meanwhile, the goal of the semantic segmentation model is to accurately label various feature categories in images, such as "roads", "buildings", "vehicles", "pedestrians" etc. Due to the complexity of the task, integrating multiple models is important to improve the accuracy and stability of the prediction.

The significances of this study are as follows:

Improving Autonomous Driving Safety

Improves driving efficiency

Promote the wide application of autonomous driving technology

Solving urban transportation problems

Promote the advancement of automated driving technology

Contribute to urban sustainable development

### **Audience**

For government and city planners, this project enhances traffic management efficiency. Real-time road condition sensing in autonomous driving can optimize traffic flow and reduce congestion [1]. Autonomous driving also mitigates accidents caused by human factors, enhancing road safety. For drivers, it ensures safe, fatigue-free driving, and selects optimal routes based on real-time conditions, saving time, reducing stress, and improving travel efficiency [5]. For the urban environment, improved traffic management and reduced congestion cut emissions, enhancing city air quality.

# **Summary of Related Literature**

Mohammed et al. [6] employed a double-branching strategy and double-attention block to improve accuracy and focus on image information, meanwhile reducing the number of parameters using the FDSS-nbt block, which achieved 74.9% mIoU. Rudra PK et al. [7] proposed new learned downsampling module that preserves spatial details and reduces computational burden to maintain 68% mIoU. Zhou et al. [8] model based on CSP encoder-decoder architecture reduces the computational cost, enhances feature extraction, and achieves 81% mIoU. Sun et al.[9] 's ResNet18-based multi-feature fusion network achieved 75% mIoU on the CityScapes dataset, proving the efficiency of multi-feature fusion.

Yu et al. [10] proposed BiSeNet model which uses a parallel structure as well as global and local information perception to improve accuracy, achieving 68.4% mIoU. Chaurasia et al. [11]: LinkNet model aims for fast semantic segmentation and uses an efficient decoder structure to reduce the computational burden, achieving 68.3% mIoU, proving that efficient semantic segmentation is feasible. Chen et al. [12] utilized DeepLab model to extend the sensory field by deep void convolution and achieves 65.88% mIoU, showing the strong performance of deep convolutional networks in urban scenarios on the urban dataset. Badrinarayanan et al. [13] suggested SegNet model which achieves semantic segmentation by using the encoding-decoding structure and achieves 65.2% mIoU, which is key to the information transfer and reconstruction mechanism. The summary of the different models proposed by different researchers can be seen in Table 1.

|  |  |  |  |
| --- | --- | --- | --- |
| **Researchers** | **Data Set** | **Techniques** | **Performance** |
| Mohammed et al.[6] | Cityscapes,CamVid | DSANet,FDSS-nbt,SIEN | mIoU=74.9% |
| Rudra et al. [7] | Cityscapes | Fast-SCNN,Coarse | mIoU=68% |
| Zhou et al. [8] | Cityscapes,CamVid | CSP en–decode -network | mIoU=81% |
| Sun et al. [9] | Cityscapes,CamVid | ResNet18,CMCA | mIoU=75.0% |
| Yu et al. [10] | Cityscapes,CamVid, COCO-Stuff | BiSeNet | mIoU=68.4% |
| Chaurasia et al. [11] | Cityscapes,CamVid | LinkNet | mIoU=68.3% |
| Chen et al. [12] | Cityscapes | Deep-Lab CRF | mIoU=65.88% |
| Badrinarayanan et al.[ [13] | CamVid | SegNet | mIoU=65.2% |

Table 1: Performance on different Semantic Segmentation models

# **Methodology**

## **Approach**

The aim of this project is to integrate multiple semantic segmentation models to create a better performing model. Below are the key steps of the project:

First, data enhancement operations are performed on the dataset, including random cropping, random flipping, and adding noise.

Second, train LinkNet,U-Net and SegNet individually, record and compare the training results of single models.

Then, combine the single models for integrated training and record the training results. This result is compared and analyzed with the single model training results in order to obtain the direction of improvement.

Finally, fine-tune the integrated model to improve model performance and avoid overfitting problems.

**3.1.1** **Semantic Segmentation:**

Semantic segmentation task is to categorize each pixel in an image into predefined categories, thus providing detailed information at the pixel level [14]. Semantic segmentation helps in accurate recognition and understanding in areas such as autonomous driving and medical imaging [14], providing in-depth analysis of objects and areas in an image.

**3.1.2 Dataset**

The Cityscapes dataset contain complex street scenes. In the dataset, the image resolution is 1024 x 512 x 3. 2975 of them were used for the training set, 500 for the test set, and 500 for the validation set.

**3.1.3 MIoU:**

(equation 1)

MIoU [15] is mainly used to measure the degree of overlap between the segmentation results predicted by the model and the true results. In formula, 'k' represents the category number, 'Pii' represents number of overlapping pixels, and 'Pji' represents the number of misassigned pixels.' 1/(k+1)' is the average weight to ensure that each category contributes equally to the mIoU. Thus, mIoU is affected by the category number, the positive sample number, and the pixel overlap between different categories.

**3.1.4 LinkNet:**

As shown in Figure 3, LinkNet is a light weight,efficient neural network structure for semantic segmentation that uses an encoder-decoder architecture. Special jump connections solve the gradient loss problem. The encoder catches image features and decoder recover image details. The encoder gradually reduces the image resolution and extracts the features and the decoder restores the resolution while fusing the features and assigning semantic categories to each pixel [17].

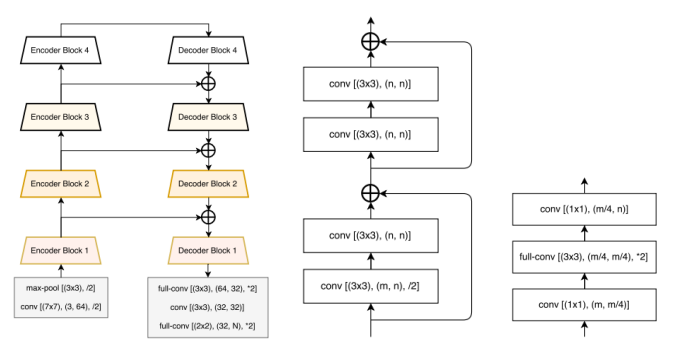


Figure 3 LinkNet structure diagram and encoder-decoder[16]

**3.1.5 U-Net:**

U-Net ass depicted in Figure 4 is a deep learning model for image segmentation. It uses an encoder-decoder architecture and jump connections to reduce images, extract features and recover to original size, which solves the gradient loss problem, improves accuracy and preserves features at all layers [18].U-Net is able to catch image features accurately, meanwhile its simple architecture is easy to modify.

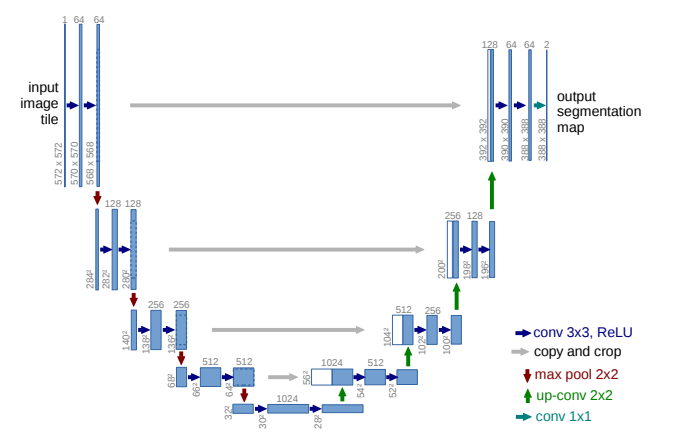


Figure 4 U-Net model [18]

**3.1.6 SegNet:**

SegNet as presented in Figure 5 is a lightweight semantic segmentation model used in areas such as autonomous driving [19]. It consists of an encoder and decoder that extracts features on the original image and improves the segmentation quality by using maximum pooling index, which is suitable for real-time semantic segmentation with limited computational resources.

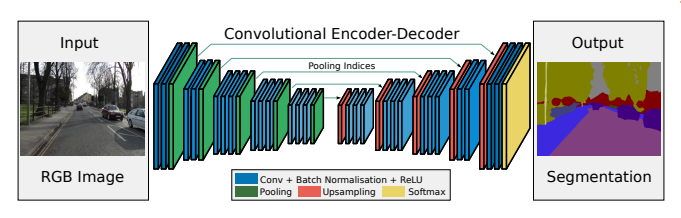


Figure 5 SegNet model [13]

**3.1.7 Proposed Model**

This Proposed model for this project will use the CityScapes dataset as seen in Figure 6. After the data enhancement, it is passed through the three models sequentially and the feature vectors of their middle layers are extracted to obtain more semantic information. To reduce the parameters, a global average pooling layer is added after each base model, then the feature vectors of the three models are connected through a voting mechanism, and finally the feature graphs are merged using void convolution to generate segmented images.

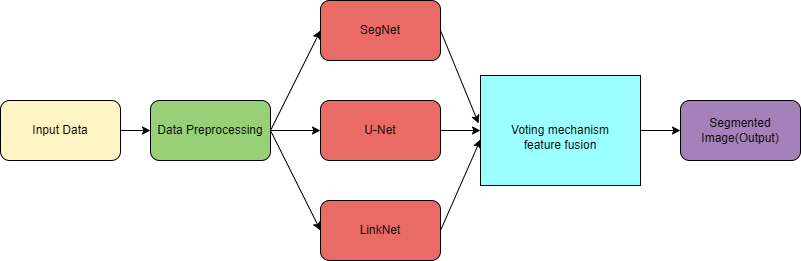


Figure 6 Proposed Model structure

## **Technology**

The part information is shown in Table 2.

|  |  |  |
| --- | --- | --- |
| Software | Framework | Tensorflow 2.10.0  Cudatoolkit 11.8.0  Cudnn 8.9.2.26 |
| Language | Python 2.9 |
| Libraries | Numpy  Matplotlib  Pandas  Keras 2.10.0  Glob3 |
| Version management plan | GitHub |
| Operation System | Windows 10 |
| Hardware | Central processing unit(CPU) | AMD Ryzen 9 4900HS with Radeon Graphics 3.00GHz |
| Graphic Processing Unit(GPU) | NVIDIA GeForce RTX 2060 Max-Q |

Table 2 The technologies of the project

## **Version Management Plan**

This project will use GitHub as the primary version control tool for regular commits and updates of project data. Also, a local and remote backup strategy is used, including code, documentation, and every version to ensure data redundancy and availability and to easily track project development.

# **Project Management**

## **Activities**

1. Research real-time semantic segmentation to grasp recent advances in the field.
2. Analyze different semantic segmentation models comparatively.
3. Study integrated learning in semantic segmentation, focusing on sub-model weight assignment and integration strategies.
4. Explore multiple datasets pertinent to the project and identify the needed ones.
5. Preprocess the selected dataset, including feature transformation, data enhancement, and segmentation.
6. Understand and construct the model based on literature's structure.
7. Choose the best-performing and structurally clear model as the backbone network for integrated learning in this project, based on implementation results.
8. Fine-tune selected model's hyper-parameters, compare performance to the original model, and analyze the differences.
9. Confirm integration strategy, weight allocation, and integrate the adjusted model.
10. Document model testing results across multiple training sessions, fine-tune hyper-parameters, and compare performance to the original model.
11. Analyze experimental results to gauge model accuracy and compare it to the original single model.
12. Complete the experimental report, encompassing project background, research methodology, results, and conclusions.
13. Create a PowerPoint presentation summarizing the course based on the report.

## **Schedule**

The project schedule designed with Gantt can be seen in Figure 7

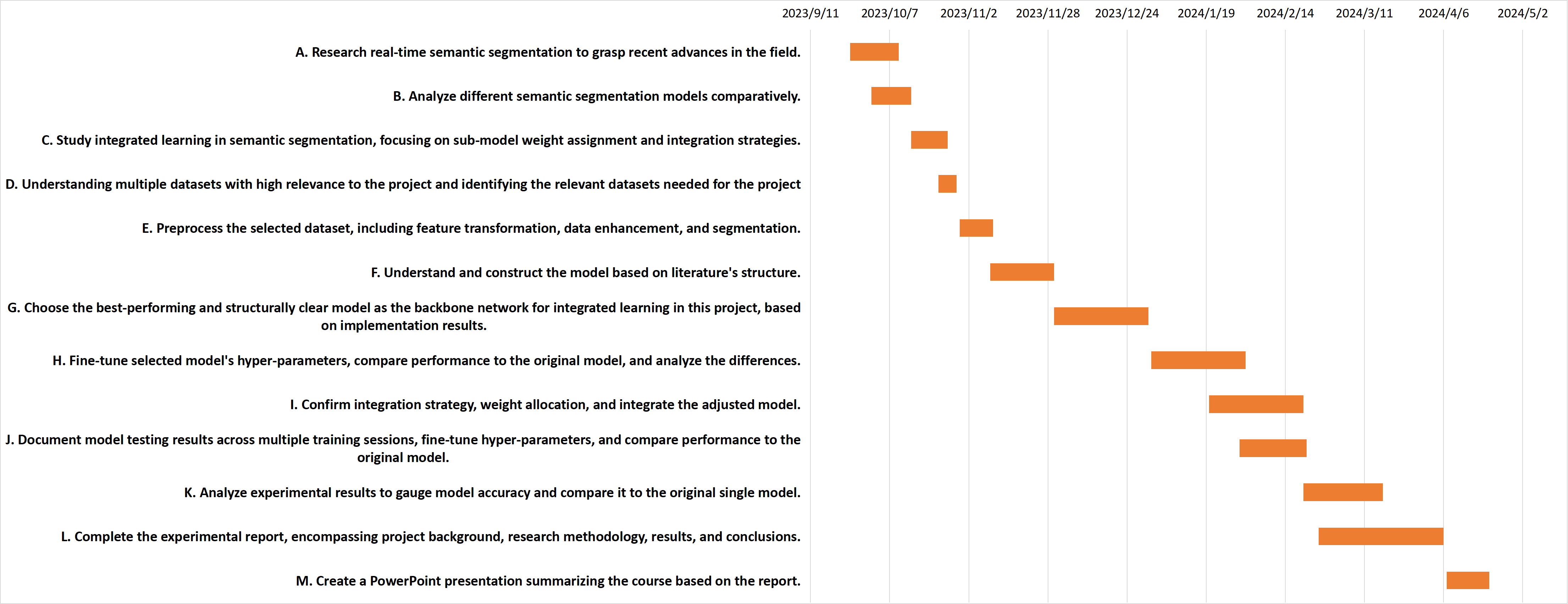


Figure 7 The project schedule.

**4.3 Data management plan**

This project will use github for data management, this includes figures, code, logs.Github will upload weekly project results as well as updating weekly project progress and also, ensuring backup of data. It will create local folders and record or manage the relevant data for easy local recall. Use Zotero to manage references for literature that is being used.

**4.4 Deliverables**

## The project proposal

1. Weekly report
2. Progress Report
3. Final Project Report
4. Project codes
5. Project presentation’s ppt
6. Github links for project management

**5. References**

1. Duarte, F. & Ratti, C., 2018. The Impact of Autonomous Vehicles on Cities: A Review. Journal of Urban Technology, 25(4), pp.3-18. Available at: https://doi.org/10.1080/10630732.2018.1493883
2. Janai, J. *et al.* (2020) ‘Computer Vision for Autonomous Vehicles: Problems, Datasets and State of the Art’, *Foundations and Trends® in Computer Graphics and Vision*, 12(1–3), pp. 1–308. Available at: https://doi.org/10.1561/0600000079.
3. Hao, S., Zhou, Y. and Guo, Y. (2020) ‘A Brief Survey on Semantic Segmentation with Deep Learning’, *Neurocomputing*, 406, pp. 302–321. Available at: https://doi.org/10.1016/j.neucom.2019.11.118.
4. Li, J. *et al.* (2021) ‘Lane-DeepLab: Lane semantic segmentation in automatic driving scenarios for high-definition maps’, *Neurocomputing*, 465, pp. 15–25. Available at: https://doi.org/10.1016/j.neucom.2021.08.105.
5. Martínez-Díaz, M. and Soriguera, F. (2018) ‘Autonomous vehicles: theoretical and practical challenges’, XIII Conference on Transport Engineering, CIT2018, 33, pp. 275–282. Available at: https://doi.org/10.1016/j.trpro.2018.10.103.
6. Elhassan, M.A.M. et al. (2021) ‘DSANet: Dilated spatial attention for real-time semantic segmentation in urban street scenes’, Expert Systems with Applications, 183, p. 115090. Available at: https://doi.org/10.1016/j.eswa.2021.115090.
7. Poudel, R.P.K., Liwicki, S. and Cipolla, R. (2019) ‘Fast-SCNN: Fast Semantic Segmentation Network’. arXiv. Available at: http://arxiv.org/abs/1902.04502 (Accessed: 23 October 2023).
8. Zhou, L. et al. (2023) ‘Real-time semantic segmentation in traffic scene using Cross Stage Partial-based encoder–decoder network’, Engineering Applications of Artificial Intelligence, 126, p. 106901. Available at: https://doi.org/10.1016/j.engappai.2023.106901.
9. Sun, J. and Li, Y. (2021) ‘Multi-feature fusion network for road scene semantic segmentation’, Computers & Electrical Engineering, 92, p. 107155. Available at: https://doi.org/10.1016/j.compeleceng.2021.107155.
10. Yu, C. et al. (2018) ‘BiSeNet: Bilateral Segmentation Network for Real-Time Semantic Segmentation’, in V. Ferrari et al. (eds) Computer Vision – ECCV 2018. Cham: Springer International Publishing (Lecture Notes in Computer Science), pp. 334–349. Available at: https://doi.org/10.1007/978-3-030-01261-8\_20.
11. Chaurasia, A. and Culurciello, E. (2017) ‘LinkNet: Exploiting encoder representations for efficient semantic segmentation’, in 2017 IEEE Visual Communications and Image Processing (VCIP). 2017 IEEE Visual Communications and Image Processing (VCIP), St. Petersburg, FL: IEEE, pp. 1–4. Available at: https://doi.org/10.1109/VCIP.2017.8305148.
12. Chen, L.-C. et al. (2018) ‘DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs’, IEEE Transactions on Pattern Analysis and Machine Intelligence, 40(4), pp. 834–848. Available at: https://doi.org/10.1109/TPAMI.2017.2699184.
13. Badrinarayanan, V., Kendall, A. and Cipolla, R. (2017) ‘SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation’, IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(12), pp. 2481–2495. Available at: https://doi.org/10.1109/TPAMI.2016.2644615.
14. Lateef, F. and Ruichek, Y. (2019) ‘Survey on semantic segmentation using deep learning techniques’, Neurocomputing, 338, pp. 321–348. Available at: https://doi.org/10.1016/j.neucom.2019.02.003.
15. Garcia-Garcia, A. et al. (2017) ‘A Review on Deep Learning Techniques Applied to Semantic Segmentation’. arXiv. Available at: http://arxiv.org/abs/1704.06857 (Accessed: 24 October 2023).
16. Chaurasia, A. and Culurciello, E. (2017) ‘LinkNet: Exploiting Encoder Representations for Efficient Semantic Segmentation’, in 2017 IEEE Visual Communications and Image Processing (VCIP), pp. 1–4. Available at: https://doi.org/10.1109/VCIP.2017.8305148.
17. K. Wu and F. Cai (2022) ‘Dual Attention D-LinkNet for Road Segmentation in Remote Sensing Images’, 2022 IEEE 14th International Conference on Advanced Infocomm Technology (ICAIT), pp. 304–307. Available at: https://doi.org/10.1109/ICAIT56197.2022.9862683.
18. Ronneberger, O., Fischer, P. and Brox, T. (2015) ‘U-Net: Convolutional Networks for Biomedical Image Segmentation’. arXiv. Available at: http://arxiv.org/abs/1505.04597 (Accessed: 25 October 2023).
19. Badrinarayanan, V., Kendall, A. and Cipolla, R. (2017) ‘SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation’, IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(12), pp. 2481–2495. Available at: https://doi.org/10.1109/TPAMI.2016.2644615.
20. G. M. Gad et al. (2020) ‘Real-Time Lane Instance Segmentation Using SegNet and Image Processing’, in 2020 2nd Novel Intelligent and Leading Emerging Sciences Conference (NILES). 2020 2nd Novel Intelligent and Leading Emerging Sciences Conference (NILES), pp. 253–258. Available at: https://doi.org/10.1109/NILES50944.2020.9257977.