

UNDERGRADUATE PROJECT REPORT

|  |  |
| --- | --- |
| **Project Title:** | **Semantic Segmentation on Deep Learning Application**  **--Segmentation of Urban Landscapes** |
| **Surname:** | **Feng** |
| **First Name:** | **Xiang** |
| **Student Number:** | **202018010119** |
| **Supervisor Name:** | **Dr Grace Ugochi Nneji** |
| **Module Code:** | **CHC 6096** |
| **Module Name:** | **Project** |
| **Date Submitted:** | **May 5, 2023** |

# **Declaration**

**Student** **Conduct** **Regulations:**

Please ensure you are familiar with the regulations in relation to Academic Integrity. The University takes this issue very seriously and students have been expelled or had their degrees withheld for cheating in assessment. It is important that students having difficulties with their work should seek help from their tutors rather than be tempted to use unfair means to gain marks. Students should not risk losing their degree and undermining all the work they have done towards it. You are expected to have familiarised yourself with these regulations.

<https://www.brookes.ac.uk/regulations/current/appeals-complaints-and-conduct/c1-1/>

Guidance on the correct use of references can be found on www.brookes.ac.uk/services/library, and also in a handout in the Library.

The full regulations may be accessed on-line at <https://www.brookes.ac.uk/students/sirt/student-conduct/> . If you do not understand what any of these terms mean, you should ask your Project Supervisor to clarify them for you.

**I** **declare** **that** **I** **have** **read** **and** **understood** **Regulations** **C1.1.4** **of** **the** **Regulations** **governing** **Academic** **Misconduct,** **and** **that** **the** **work** **I** **submit** **is** **fully** **in** **accordance** **with** **them**.

Signature: Data: 22/04/2024

REGULATIONS GOVERNING THE DEPOSIT AND USE OF OXFORD BROOKES UNIVERSITY MODULAR PROGRAMME PROJECTS AND DISSERTATIONS

Copies of projects/dissertations, submitted in fulfilment of Modular Programme requirements and achieving marks of 60% or above, shall normally be kept by the Library.

**|** **agree** **that** **this** **dissertation** **may** **be** **available** **for** **reading** **and** **photocopying** **in** **accordance** **with** **the** **Regulations** **governing** **use** **of** **the** **Library.**

Signature: Data: 22/04/2024

# **Acknowledgment**

I want to express my heartfelt gratitude to my Supervisor, Dr. Grace Ugochi Nneji, for her guidance and support throughout my final undergraduate project. Her expertise, patience, and encouragement have been invaluable in shaping this project. Also, I would like to thank the module leader Joojo Walker and other teachers throughout my undergraduate study time for their teaching and advice.

More so, I would like to acknowledge the resources and facilities provided by Oxford Brookes University in collaboration with the Chengdu University of Technology for an outstanding opportunity.

And to my family and friends, your unending love and encouragement have been with me. I will always be grateful.

# **Table of Contents**

[**Declaration** i](#_Toc129605831)

[**Acknowledgment** ii](#_Toc129605832)

[**Table of Contents** iii](#_Toc129605833)

[**Abstract** iv](#_Toc129605834)

[**Abbreviations** v](#_Toc129605835)

[**Glossary** vi](#_Toc129605836)

[**Chapter 1 Introduction** 1](#_Toc129605837)

[**1.1** **Background** 1](#_Toc129605838)

[**1.2** **Aim** 1](#_Toc129605839)

[**1.3** **Objectives** 1](#_Toc129605840)

[**1.4** **Project Overview** 1](#_Toc129605841)

[**1.4.1** **Scope** 1](#_Toc129605842)

[**1.4.2** **Audience** 1](#_Toc129605843)

[**Chapter 2 Background Review** 2](#_Toc129605844)

[**Chapter 3 Methodology** 3](#_Toc129605845)

[**3.1** **Approach** 3](#_Toc129605846)

[**3.2** **Technology** 3](#_Toc129605847)

[**3.3** **Project Version Management** 3](#_Toc129605848)

[**Chapter 4 Results** 4](#_Toc129605849)

[**Chapter 5 Professional Issues** 5](#_Toc129605850)

[**5.1** **Project Management** 5](#_Toc129605851)

[**5.1.1** **Activities** 5](#_Toc129605852)

[**5.1.2** **Schedule** 5](#_Toc129605853)

[**5.1.3** **Project Data Management** 5](#_Toc129605854)

[**5.1.4** **Project Deliverables** 5](#_Toc129605855)

[**5.2** **Risk Analysis** 5](#_Toc129605856)

[**5.3** **Professional Issues** 5](#_Toc129605857)

[**Chapter 6 Conclusion** 6](#_Toc129605858)

[**References** 7](#_Toc129605859)

[**6.1** **Formatting Requirements** 7](#_Toc129605860)

[**6.2** **Written Presentation** 7](#_Toc129605861)

[**Appendices** 8](#_Toc129605862)

# **Abstract**

Semantic segmentation is an important research direction in the field of computer vision , which plays a key role in many daily life applications, such as self-driving, medical image analysis and video monitoring. Nevertheless, today's semantic segmentation techniques still face challenges such as complex scene structure, multi-scale object recognition and changing environmental conditions, which bring very huge disturbances to semantic segmentation. To deal with these challenges, this study designs a high-performance semantic segmentation framework. The model does this by fusing multiple advanced deep learning techniques, such as using the ASPP module with deeply separable convolution to capture multi-scale contextual information, adopting the Transformer module to enhance model's ability to capture global dependency, and applying multi-scale pooling to optimize the model's ability to process features at different scales. This module fusion method significantly improves the accuracy and robustness of semantic segmentation, and achieves excellent segmentation results on diverse datasets while increasing the segmentation speed. By evaluating the model on test datasets, the model demonstrates its excellent performance in handling complex image segmentation tasks. The core contribution of this research is to propose an efficient and accurate semantic segmentation framework integrating deep separable convolution, ASPP, Transformer and multi-scale pooling, which brings new research directions and application potentials to the field of semantic segmentation.

Keywords: Semantic Segmentation, Computer Vision, Deep Learning, Technological integration, ASPP, Transformer Module, Multi-scale Pooling

# **Abbreviations**

CNN: Convolutional Neural Network

ASPP: Atrous Spatial Pyramid Pooling

ResNet: Residual Network

MIOU: Mean Intersection Over Union

GAP: Global Average Pooling

DWC: Depth-wise Convolution

SGD: Stochastic Gradient Descent

PA: Pixel Accuracy

CPA: Class Pixel Accuracy

FPS: Frames Per Second

ReLU: Rectified Linear Unit

BN: Batch Normalization

TN: True Negative

TP: True Positive

FP: False Positive

FN: False Negative

GUI: Graphical User Interface

GPU: Graphics Processing Unit

# **Glossary**

**Semantic Segmentation[1]:** This is a technique that segments an image into multiple parts and each part is labeled with different categories for understanding scenes and objects in the image.

**Deep Learning [2]:** Learning from large amounts of data by using networks with multiple layers of processing units, and is widely used for tasks such as image recognition, speech recognition, and semantic segmentation.

**Convolutional Neural Network (CNN) [3]:** A convolutional neural network is a deep learning model specifically designed to process data that has a clear grid structure, such as images. By stacking convolutional, pooling, and fully connected layers, CNN are able to extract features from images efficiently and use them for tasks such as classification, detection, and segmentation.

**Image Augmentation [4]:** This is a technique where a series of transformations (e.g., rotation, scaling, cropping, etc.) are applied to the training image to increase the variety of the dataset artificially. This helps the model learn more generalized features, which improves performance on unseen data.

**Feature Extraction [5]:**In deep learning, feature extraction usually means using high-level features learned from previous layers of the model that are helpful for the task at hand.

**Muti-scale Feature Fusion [6]:** This is a technique that can combine features from different resolutions of an image to capture information ranging from coarse to detailed.

**Attention Mechanism [7]:** Attention Mechanism allows deep learning models to dynamically focus on important parts of the input data. It plays a key role in improving the explainable and performance of the model.

**Stochastic Gradient Descent (SGD) [8]:** It is an optimization algorithm used to minimize the loss function of the model during the training process.The SGD optimizer can effectively reduce the consumption of computational resources and speed up the training process.

**Atrous Spatial Pyramid Pooling (ASPP) [9]:** ASPP is a technique to capture multi-scale information by using null convolution with different sampling rates.ASPP can effectively improve the feature resolution and model performance in semantic segmentation tasks.

**Cross-entropy Loss [10]:** It is a loss function commonly used in classification tasks to quantify the difference between the probability distribution predicted by the model and the true label. It is essential for training high-performance classification models.

**Dice Loss [11]:** It is a loss function used in image segmentation tasks. It is particularly suitable for dealing with the problem of category imbalance, where the accuracy of segmentation is improved by optimizing the model to the overlap of the segmented regions of the Jung family.

# **Introduction**

## **Background**

With urban traffic congestion and frequent accidents becoming an increasing problem, self-driving vehicles are widely recognized as a potential solution to reduce accident rates and improve traffic flow **[12]**. However, the application of self-driving in urban environments faces a number of challenges, which include dealing with complex backgrounds, variable weather conditions, and diverse traffic scenarios **[13]**. These factors make it difficult for traditional computer vision techniques to be adapted, increasing the difficulty of realizing self-driving technology **[13]**.

In this case, semantic segmentation is especially crucial as a machine vision technique that allows a fine understanding of the surrounding environment **[14]**. It can relate each pixel in an image to a semantic category of roads, buildings, vehicles, and pedestrians **[15]**. This can help self-driving systems to recognize and understand their surroundings more accurately, providing a powerful environment perception tool for vehicles **[16]**.

Although semantic segmentation technology shows great potential in self-driving, it still faces many challenges in practical applications, including how to effectively deal with complex scene structures, adapt multi-scale object recognition, and deal with changing environmental conditions **[17]**. These challenges require semantic segmentation models not only have better accuracy and robustness, but also need to realize real-time and fast image processing to adapt the real-time decision-making requirements for self-driving vehicles **[18]**.

Therefore, it has become an urgent problem for how to make the semantic segmentation technology adapt the application requirements of self-driving vehicles in complex urban environments. This study is devoted to in-depth research on semantic segmentation technology, aiming to develop a high-accuracy real-time semantic segmentation model to provide safer and more efficient self-driving technology for urban transportation systems to promote sustainable urban development.

* + 1. **Convolutional Layers**

The convolutional layer is the core building block of a convolutional neural network and is responsible for executing most of the computations. It requires several components, including input data, filters, and feature maps. The input is assumed to be a color image, consisting of a three-dimensional matrix of pixels. This means that the input has three dimensions: height, width, and depth, corresponding to the RGB in the image. There is also a feature detector, also known as a kernel or filter, which moves through the various sense fields of the image, checking for the presence of features. This process is called convolution **[29]**.

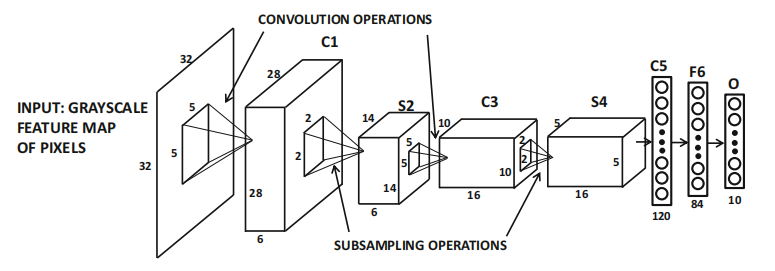


Figure 1: convolutional architecture[29]

* + 1. **Activation Functions**

Activation functions are a type of function added to an artificial neural network designed to help the network learn complex patterns in the data. The Sigmoid function compresses the input to the (0,1) interval, which is suitable for binary classification problems, but it is easy to cause gradient vanishing. Tanh function outputs in the range of (-1,1), and its output centering improves the stability of the model training, but there is the problem of gradient vanishing.The ReLU function is the most widely used. Input is greater than 0 to remain unchanged, the output is 0 when the input is less than 0. It solves the gradient disappearance problem, the training speed is fast **[29]**.

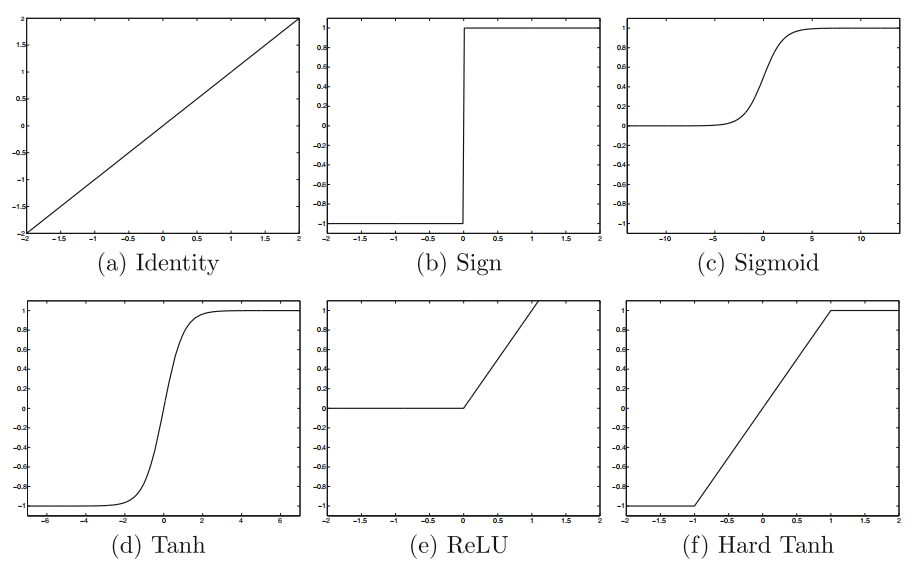


Figure 2: Activation function image[29]

* + 1. **Pooling Layers**

The pooling layer, also known as the downsampling layer, performs a dimensionality reduction operation designed to reduce the number of parameters in the input. Similar to the convolutional layer, the pooling operation lets the filter scan the entire input, but the difference is that this filter has no weights. The kernel applies an aggregation function to the values in the sensory field to populate the output array. There are two main types of pooling: 1. Maximum pooling: as the filter moves through the input, it selects the pixel with the largest value and sends it to the output array. 2. Average pooling: as the filter moves through the input, it calculates the average value in the receptive field and sends it to the output array. Although a lot of information is lost in the pooling layer, it still brings many benefits to the convolutional neural network. The layer helps reduce complexity, increase efficiency, and limit the risk of overfitting **[29]**.

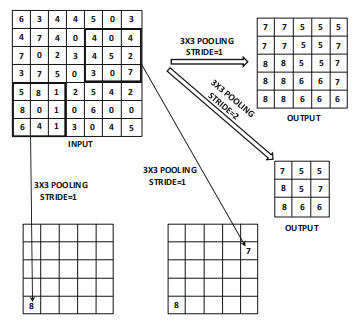


Figure 3: Pooling schematic[29]

* + 1. **Up-Sampling**

Up-sampling increases the size of the data by inserting new pixels or feature points, using methods such as transposed convolution to recover detail and support accurate prediction.

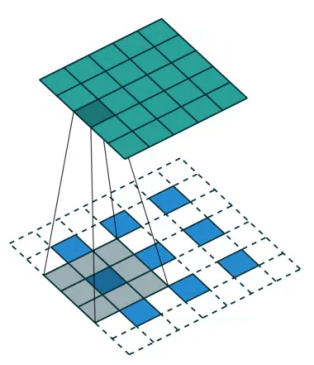


Figure 4: Up-Sample process

* + 1. **Batch Normalization**

Batch normalization is a technique used to improve the efficiency and stability of deep neural network training by normalizing the distribution of the inputs at each layer so that the mean is 0 and the standard deviation is 1. This helps to address the problem of internal covariate bias in training, allowing the network to use higher learning rates and be less sensitive to the choice of initialization weights. Batch normalization also indirectly provides a regularization effect that accelerates the convergence of the network and improves the generalization ability of the model **[29]**.

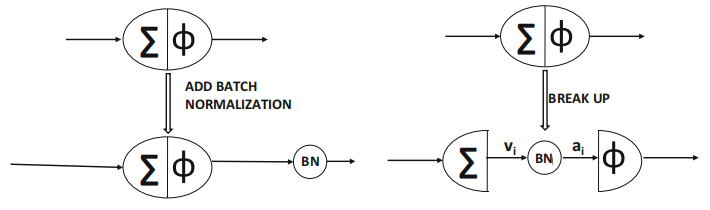


Figure 4: Batch Normalization

* + 1. **ASPP**

ASPP module improves the accuracy of semantic segmentation by processing images in parallel using null convolution with different sampling rates, effectively capturing information at different scales. This method expands the sensory field of the model, allowing it to simultaneously understand both details and broader contextual information in the image, and is particularly suitable for dealing with size variations in images, thus improving segmentation performance. In short, ASPP allows the model to better adapt to targets of different sizes and improve semantic segmentation **[30]**.

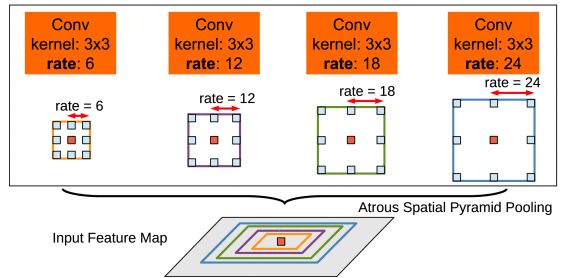


Figure 5: ASPP Module[30]

* + 1. **Loss Functions**

Cross-entropy Loss **[10]**: Cross-entropy loss measures the discrepancy between model predictions and actual labels, and is a key tool for optimizing model accuracy in tasks dealing with classification and semantic segmentation.

Dice Loss **[11]**: Dice Loss is a loss function based on Dice coefficients for evaluating the similarity of two samples, which is particularly suitable for the case of category imbalance in image segmentation, and significantly improves the accuracy and performance of semantic segmentation by optimizing the overlap rate between predicted and real labels.

* + 1. **Optimization Algorithms**

SGD Optimizer:SGD is an optimization algorithm commonly used for deep learning model training, which progressively reduces the value of the loss function by randomly selecting a sample of data to compute the gradient and updating the model parameters in each iteration.SGD aims to efficiently find the global optimal solution, and although the update at each step may not be as accurate as gradient descent based on the entire data, it is more efficient and practical when dealing with large datasets **[8]**.

## **Aim**

The primary aim of this project is to create a high-performance semantic segmentation model using deep learning techniques. The model integrates a variety of advanced deep learning techniques, including the ASPP module based on DWC, the Transformer module, and multi-scale pooling. It aims to significantly improve the accuracy and robustness of the semantic segmentation model in complex scene structures, multi-scale object recognition and diverse environmental conditions. In addition, to enhance the applicability and interactivity of the project, I plan to use Flask build a website to implement GUI, which will display the segmentation results of the semantic segmentation model, aiming able to intuitively see the results of the model processing various images. In conclusion, this project aims to provide a novel semantic segmentation framework and promote the development and popularization of semantic segmentation technology in practical applications by developing a user-friendly GUI.

## **Objectives**

In order to achieve the main goal of this project, the following specific objectives have been set:

* In-depth research and experimentation of advanced deep learning techniques: research and implement multiple deep learning modules, such as DWC-based ASPP module, Transformer module, edge detection module, and multi-scale pooling module, and try to integrate them. Aims to improve the performance of semantic segmentation models through these advanced techniques.
* Improve the accuracy and robustness of the model in complex environments: through the above techniques, the accuracy and stability of the model in complex scene structures, multi-scale object recognition, and diverse environmental conditions are improved to meet the common challenges in semantic segmentation.
* Evaluating model performance: evaluate the accuracy and robustness of the model using evaluation metrics (e.g., accuracy, recall, MIOU, PA, Dice coefficients, Kappa scores, FPS) by running the model on a test dataset. This will provide an objective benchmark for comparing different model architectures and configurations.
* Developing a Graphical User Interface (GUI): Build a website using the Flask framework to implement a GUI for presenting semantic segmentation results so that users can easily upload images, start the segmentation process and view the model outputs intuitively, enhancing the applicability and interactivity of the project.
* Contribute to the development and popularization of semantic segmentation technology: Through the development of an easy-to-use GUI and a high-performance semantic segmentation model, we provide new ideas and tools for the research, development, and application of semantic segmentation technology, and promote the development and popularization of this technology in practical applications.

## **Project Overview**

### **Scope**

The core goal of this project is to develop a high-performance semantic segmentation model using deep learning techniques, aiming to significantly improve accuracy and robustness in complex scenarios, multi-scale object recognition, and variable environmental conditions, which will contribute to the realization of self-driving technologies. The scope of the project includes the design and implementation of a semantic segmentation model that integrates multiple deep learning techniques, in particular the fusion of the DWC-based ASPP module, the Transformer module, and multi-scale pooling techniques. This integration strategy works on extracting richer image features as well as reducing the training parameters of the model to improve its efficiency and accuracy. In addition to this, the project creates a website to implement a GUI through the Flask framework to visualize the results of semantic segmentation to enhance the user's application experience and interactivity. Through the GUI, users can upload images for segmentation and instantly view the processing results of the model. Finally, the project will evaluate the performance of the model on a test dataset and compare it with other existing models to show the advantages of my model. Through the above efforts, this project is not only devoted to promote the innovation of semantic segmentation technology, but also to promote the popularization and application of this technology in real-world applications through the development of an easy-to-use GUI.

### **Audience**

For government and city planners, this project enhances traffic management efficiency. Realtime road condition sensing in autonomous driving can optimize traffic flow and reduce congestion **[13]**. 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 **[19]**. For the urban environment, improved traffic management and reduced congestion cut emissions, enhancing city air quality.

# **Background Review**

**2.1 Summary of Related Literature**

The research background of semantic segmentation techniques centers around a central task in the field of computer vision - understanding the category to which each pixel in an image belongs **[15]**. This technique enables computers not only to recognize the objects present in an image, but also to accurately classify the boundary of each object, which is an important foundation in the field of image processing and analysis **[14]**.

Nowadays, the introduce of deep learning, especially the application of CNN, greatly improves the accuracy and efficiency of semantic segmentation, allowing machines to process more complex image data, recognize and segment multiple objects in an image **[1]**. The progress of this technology not only promotes the development of the computer vision field, but also brings innovative possibilities for a number of application fields, such as self-driving, medical image analysis, and environmental monitoring **[1]**.

By combing through the related literature, I will compare the MIOU of some existing semantic segmentation models, and bellow Table1 is some comparing of the performance for some semantic segmentation models in the related literature.

|  |  |  |  |
| --- | --- | --- | --- |
| **Author** | **Model** | **MIou** | **Dataset** |
| Badrinarayanan et.al. **[20]** | SegNet | 56.1% | Cityscapes |
| Abdigapporov et.al.**[21]** | BiFPN | 56.4% | Cityscapes |
| Paszke et.al. **[22]** | ENet | 58.3% | Cityscapes |
| Poudel et.al. **[23]** | Fast-SCNN | 68% | Cityscapes |
| Yu et al **[24]** | BiSeNet | 69% | Cityscapes |
| Fourure et al **[25]** | GridNet | 69.5% | Cityscapes |
| Chen et al **[26]** | Deep-Lab CRF | 70.4% | Cityscapes |
| Lin et al **[27]** | RefineNet | 73.6% | Cityscapes |
| Li et.al.**[28]** | BiAttnNet | 74.7% | Cityscapes |
| My Model | DeepSegASPP+Transformer | 75% | Cityscapes |

**Table 1**

First, the SegNet model proposed by Badrinarayanan et al **[20]**. is a convolutional network using a nonlinear upsampling technique that achieves a MIOU of 56.1%, which is more suitable for simple scene understanding. Later, the BiFPN network developed by Abdigapporov et al.**[21]** achieved a MIOU of 56.4% by enhancing the multi-scale fusion of features. ENet designed by Paszke et al. **[22]** is a lightweight network designed to satisfy the requirements of real-time applications with 58.3% MIOU, while Fast-SCNN proposed by Poudel et al. **[23]** achieves 68% MIOU by optimizing the computational efficiency, and both networks are suitable for semantic segmentation applications on mobile.

BiSeNet developed by Yu et al. **[24]** uses dual network structure to achieve a model that balances speed and performance with 69% MIOU. Fourure et al.**[25]**'s GridNet model enhances feature fusion through its grid structure and achieves 69.5% MIOU. Chen et al. **[26]'**s DeepLab-CRF, which is constructed using ASPP with CRF technology, achieves 69.5% MIOU on the edge processing was refined and achieved 70.4% MIOU. RefineNet proposed by Lin et al. **[27]** enhanced the information fusion by multipath refinement technique and achieved 73.6% MIOU. BiAttnNet model by Li et al. **[28]** with the introduction of bi-directional attention mechanism drastically improved the segmentation accuracy and achieved 74.7% MIOU.

# **References**

1. F. Lateef and Y. Ruichek, "Survey on semantic segmentation using deep learning techniques," Neurocomputing, vol. 338, pp. 321–348, 2019. Available: https://doi.org/10.1016/j.neucom.2019.02.003.
2. A. W. Trask, Grokking Deep Learning. Manning, 2019, pp. 1–. Available: http://ieeexplore.ieee.org/document/10280208. ISBN: 9781617293702.
3. H. Lyu, "Research on Corrosion Recognition Method of Steel Based on Convolutional Neural Network," in Proc. 2023 IEEE 6th International Conference on Information Systems and Computer Aided Education (ICISCAE), Dalian, China, 2023, pp. 507-511, doi: 10.1109/ICISCAE59047.2023.10393077.
4. D. Haba, Data Augmentation with Python: Enhance Deep Learning Accuracy with Data Augmentation Methods for Image, Text, Audio, and Tabular Data. Packt Publishing, 2023.
5. A. Meyer-Baese and V. Schmid, “Chapter 2 - Feature Selection and Extraction,” in Pattern Recognition and Signal Analysis in Medical Imaging (Second Edition), Second Edition., A. Meyer-Baese and V. Schmid, Eds., Oxford: Academic Press, 2014, pp. 21–69. doi: https://doi.org/10.1016/B978-0-12-409545-8.00002-9.
6. L. Huang, C. Chen, J. Yun, Y. Sun, J. Tian, Z. Hao, H. Yu, and H. Ma, "Multi-Scale Feature Fusion Convolutional Neural Network for Indoor Small Target Detection," Front. Neurorobot., vol. 16, 881021, 2022, doi: 10.3389/fnbot.2022.881021.
7. X. Zhu, D. Cheng, Z. Zhang, S. Lin, and J. Dai, "An Empirical Study of Spatial Attention Mechanisms in Deep Networks," in Proc. 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea (South), 2019, pp. 6687-6696, doi: 10.1109/ICCV.2019.00679.
8. X. Wang, L. Yan, and Q. Zhang, "Research on the Application of Gradient Descent Algorithm in Machine Learning," in Proc. 2021 International Conference on Computer Network, Electronic and Automation (ICCNEA), Xi'an, China, 2021, pp. 11-15, doi: 10.1109/ICCNEA53019.2021.00014.
9. L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs," CoRR, vol. abs/1606.00915, 2016. [Online]. Available: http://arxiv.org/abs/1606.00915
10. A. Mao, M. Mohri, and Y. Zhong, "Cross-entropy loss functions: Theoretical analysis and applications," in Proc. ICML 2023. [Online]. Available: https://arxiv.org/pdf/2304.07288.pdf
11. S. Kato and K. Hotta, “Adaptive t-vMF dice loss: An effective expansion of dice loss for medical image segmentation,” Computers in Biology and Medicine, vol. 168, p. 107695, 2024, doi: https://doi.org/10.1016/j.compbiomed.2023.107695.
12. 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
13. 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.
14. H. Hu, H. Cai, Z. Ma, and W. Wang, "Semantic segmentation based on semantic edge optimization," in Proc. 2021 International Conference on Electronic Information Engineering and Computer Science (EIECS), Changchun, China, 2021, pp. 612-615, doi: 10.1109/EIECS53707.2021.9587939.
15. 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.
16. 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.
17. Y. Guo and B. Yang, "A Survey of Semantic Segmentation Methods in Traffic Scenarios" in Proc. 2022 International Conference on Machine Learning, Cloud Computing and Intelligent Mining (MLCCIM), Xiamen, China, 2022, pp. 452-457, doi: 10.1109/MLCCIM55934.2022.00083.
18. K. Singh, V. Varshney, and M. Rajl, "Image Segmentation Role in Self Driving Car," in Proc. 2022 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), Greater Noida, India, 2022, pp. 861-864, doi: 10.1109/ICAC3N56670.2022.10074404.
19. 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.
20. V. Badrinarayanan, A. Kendall, and R. Cipolla, “SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 12, pp. 2481–2495, 2017, doi: 10.1109/TPAMI.2016.2644615.
21. S. Abdigapporov, S. Miraliev, V. Kakani, and H. Kim, “Joint Multiclass Object Detection and Semantic Segmentation for Autonomous Driving,” IEEE Access, vol. 11, pp. 37637–37649, 2023, doi: 10.1109/ACCESS.2023.3266284.
22. A. Paszke, A. Chaurasia, S. Kim, and E. Culurciello, “ENet: A Deep Neural Network Architecture for Real-Time Semantic Segmentation,” CoRR, vol. abs/1606.02147, 2016. Available: http://arxiv.org/abs/1606.02147
23. R. P. K. Poudel, S. Liwicki, and R. Cipolla, “Fast-SCNN: Fast Semantic Segmentation Network,” CoRR, vol. abs/1902.04502, 2019. Available: http://arxiv.org/abs/1902.04502
24. C. Yu, J. Wang, C. Peng, C. Gao, G. Yu, and N. Sang, “BiSeNet: Bilateral Segmentation Network for Real-time Semantic Segmentation,” CoRR, vol. abs/1808.00897, 2018. Available: http://arxiv.org/abs/1808.00897
25. D. Fourure, R. Emonet, É. Fromont, D. Muselet, A. Trémeau, and C. Wolf, “Residual Conv-Deconv Grid Network for Semantic Segmentation,” CoRR, vol. abs/1707.07958, 2017, [Online]. Available: http://arxiv.org/abs/1707.07958
26. L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, “DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 40, no. 4, pp. 834–848, 2018, doi: 10.1109/TPAMI.2017.2699184.
27. G. Li, L. Li, and J. Zhang, “BiAttnNet: Bilateral Attention for Improving Real-Time Semantic Segmentation,” IEEE Signal Processing Letters, vol. 29, pp. 46–50, 2022, doi: 10.1109/LSP.2021.3124186.
28. G. Lin, A. Milan, C. Shen, and I. D. Reid, “RefineNet: Multi-Path Refinement Networks for High-Resolution Semantic Segmentation,” CoRR, vol. abs/1611.06612, 2016, [Online]. Available: http://arxiv.org/abs/1611.06612
29. C. C. Aggarwal, Neural Networks and Deep Learning: A Textbook. Cham, Switzerland: Springer International Publishing AG, part of Springer Nature, 2018. doi: 10.1007/978-3-319-94463-0.
30. L. -C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 40, no. 4, pp. 834-848, Apr. 1, 2018, doi: 10.1109/TPAMI.2017.2699184.