COMP9517 Group Project – Semantic Segmentation

An innovative approach to 2D WildScene Natural Environments

Liam Chen   
email: z3278107@ad.unsw.edu.au  
student id: z3278107

Jennifer Wang   
email: z5311246@ad.unsw.edu.austudent id: z5311246Shree Baskar  
email: z5444481@ad.unsw.edu.austudent id: z5444481

Amulya Lokeshwari Sri Raja Kalidindi  
email: z5520321@ad.unsw.edu.austudent id: z5520321The Duy Nguyen   
email: z5532839@ad.unsw.edu.austudent id: z5532839

*Abstract*—TODO. The output module is generated using a computer with GPU memory NVIDIA GeForce RTX 3060 6G, programming with python 3.8.19. A python requirements.txt can be found in the submission

Keywords—semantic segmentation, deep learning, image classification, convolutional networks, Jaccard similarity, Intersection over Union (IoU), U-Net, SegNet, VGG, Deep CNN, ensemble learning, semantic scene understanding

# Introduction

Semantic segmentation (per-pixel class labelling) is a critical task in computer vision (Szeliski, 2021, p.413) and it has gained a lot of attention in the research field recently. This task is significant in applications such as autonomous driving, environmental monitoring, search and rescue, and automation, where understanding of surrounding environments is required to navigate and perform tasks effectively. (Vidanapathirana et al. 2023). While there are many notable structured 2D and 3D urban scene datasets to study the urban environment semantic segmentation (Vidanapathirana et al. 2023), natural scene semantic segmentation is less common and therefore warrants increased focus.

To address this gap, the WildScenes dataset was created (Vidanapathirana et al., 2023). It contains RGB images of traversals over six months across two locations: Venman National Park and Karawatha Forest Park in Brisbane, Australia. A total of 21.28 km was travelled and 9306 images were captured, each with a resolution of 2016 x 1512. Each image was annotated manually using a coarse-to-fine approach with multiple rounds of auditing to overcome the challenges in labelling natural environments. These challenges include ambiguity of object boundaries for elements with similar features (i.e., leaves and shrubs, or dirt and grass), and differentiating similar regions.

This project aims to implement natural scene semantic segmentation using state-of-the-art deep learning models such as U-Net, SegNet and VGG. We will leverage powerful architectures to achieve high accuracy in segmenting various natural elements that are above the benchmarks published in the original research paper. Additionally, techniques like superpixels and Conditional Random Fields (CRFs) will be incorporated to refine the segmentation boundaries, enhancing the overall performance of the model.

# Literature Review

Semantic segmentation is the process of classifying each pixel in an image as belonging to one of a set of predefined classes. Many techniques have been developed over the years, ranging from early image-processing techniques to convolutional neural networks (CNNs).

The work of Shelhamer et al. (2015) introduced Fully Convolutional Networks for semantic segmentation. These replaced the fully connected layers in traditional CNNs with a decoding, or upsampling, layer to enable pixel-wise classification on the original image scale. This paved the way for further developments in deep learning such as U-Net (Ronneberger et al., 2015), DeepLab (Chen et al., 2018), and SegNet (Badrinarayanan et al., 2017).

Despite these recent advancements, natural scene segmentation has received little attention in research. It poses additional challenges due to its less structured and highly variable elements (Vidanapathirana et al., 2023).

## Fully Convolutional Neural Networks for Semantic Segmentation (Shelhamer et al., 2015)

## SegNet

## Superpixels and Conditional Random Fields

## Hierarchical Features for Scene Labelling

## TODO

# Methods

The WildScenes dataset was used for this task. It contains 9306 RGB images with a resolution of 2016 x 1512, annotated manually using a coarse-to-fine approach.

The dataset was split using the methods from the original paper (Vidanapathirana et al., 2023). The dataset was split into training, validation, and tests sets to ensure a good class distribution and avoid geographical overlap. A minimum distance of 45 metres was kept between samples in different sets, grouped based on their (x, y) coordinates using k-means clustering. One thousand candidate splits were generated, evaluated, and the optimal split was chosen. We used the entire dataset to train our models, and manually verified the class distribution in the training, validation, and testing datasets.

Table Type Styles

| Table Head | Table Column Head | | |
| --- | --- | --- | --- |
| Table column subhead | Subhead | Subhead |
| copy | More table copya |  |  |

1. Sample of a Table footnote. (*Table footnote*)
2. Example of a figure caption. (*figure caption*)

Figure Labels: Use 8 point Times New Roman for Figure labels. Use words rather than symbols or abbreviations when writing Figure axis labels to avoid confusing the reader. As an example, write the quantity “Magnetization”, or “Magnetization, M”, not just “M”. If including units in the label, present them within parentheses. Do not label axes only with units. In the example, write “Magnetization (A/m)” or “Magnetization {A[m(1)]}”, not just “A/m”. Do not label axes with a ratio of quantities and units. For example, write “Temperature (K)”, not “Temperature/K”.

# Experimental Results

# Discussion

# Conclusion

##### Acknowledgment *(Heading 5)*

The preferred spelling of the word “acknowledgment” in America is without an “e” after the “g”. Avoid the stilted expression “one of us (R. B. G.) thanks ...”. Instead, try “R. B. G. thanks...”. Put sponsor acknowledgments in the unnumbered footnote on the first page.

# References

For papers published in translation journals, please give the English citation first, followed by the original foreign-language citation [6].

1. G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955. *(references)*
2. J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.
3. I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.