AUTOMATED LANDCOVER SEGMENTATION

Sakshi Sinha
Department of Computer Science
University of Adelaide
sakshi.sinha@student.adelaide.edu.au

ABSTRACT

This proposal details a project to develop and evaluate deep learning models for accurate Australian multispectral satellite image segmentation, addressing a gap in open-source research. The project will compare CNNs, transformers, and VLMs, providing optimised segmentation strategies for Aurizn.

1 INTRODUCTION

The increasing availability of detailed satellite imagery has significantly advanced remote sensing applications. Precise segmentation of such imagery is essential for various domains, including urban planning, environmental monitoring, disaster response, and resource management [1]. For example, accurate identification of land cover types (vegetation, buildings, roads, water) facilitates improved urban development strategies [2], monitoring deforestation and ecosystem health [3], assessing post-disaster damage [4], and optimizing agricultural practices [5]. Inaccurate segmentation can lead to flawed analyses and ineffective decision-making in these crucial areas. Moreover, the rich information provided by accurate segmentation enables more refined insights and a deeper understanding of the observed environment. Hence, accurate satellite image segmentation is crucial, yet open-source research tailored to Australia's unique landscapes is lacking. While private studies may exist, Aurizn requires this project to develop accessible, context-specific models.

While deep learning (DL) advancements in image segmentation have been substantial, unique challenges arise with specialized datasets. Aurizn's high-resolution multispectral dataset, acquired over Australia, differs from much of the existing research, which primarily focuses on RGB imagery [6]. Multispectral data offers richer information beyond the visible spectrum, allowing for better differentiation of features often indistinguishable in RGB images [7]. However, to effectively utilize this data, specialized techniques are necessary to capture the complex relationships between spectral bands [8].

This project will focus on addressing Aurizn's need for precise and efficient land cover segmentation in its high-resolution multispectral imagery. By developing and implementing tailored deep learning solutions, the project aims to empower Aurizn to gain actionable insights, facilitating smarter, data-driven decisions throughout its operations. This will promote more sustainable and efficient practices, fully leveraging the rich information contained in its multispectral data.

2 PROJECT AIMS AND OBJECTIVES

2.1 AIM

To investigate and develop advanced DL models for accurate semantic segmentation of Aurizn's high-resolution multispectral satellite imagery, focusing on transformers/Convolutional Neural Networks (CNNs) architectures and pre-trained models, with the goal of improving segmentation accuracy and broader image understanding for enhanced decision-making within Aurizn's operations.

2.2 OBJECTIVES

- Adapt and apply CNN and transformer-based DL models, including SAM and remote sensing-specific foundational models, to Aurizn's dataset.

Figure 1: Example of land cover segmentation in satellite imagery. Image sourced from [9].

- Investigate the effectiveness of fine-tuning pre-trained models for improved segmentation performance on Aurizn's data.
- Compare the performance of fine-tuned pre-trained models as well as models trained from scratch with an HRNet model developed from the ground up by Aurizn to determine the most effective approach for Aurizn's data and operational requirements.
- As a stretch goal, explore the potential of Vision-Language Models (VLMs) like GeoChat for tasks such as Temporal Understanding, Referring Segmentation, Scene Understanding, Counting, and Detailed Image Captioning, and assess their ability to extract complex information from satellite imagery. This will also include demonstrating the potential of VLMs to Aurizn's clients.

2.3 Expected Outcomes

- A comprehensive evaluation of different DL architectures and training strategies for multispectral satellite image segmentation, including an assessment of their strengths and weaknesses.
- Investigation and identification of promising DL models, VLMs and pre-trained models suitable for application to Aurizn's multispectral satellite imagery.
- A detailed report summarizing the findings of the investigation, including recommendations for future development and implementation of segmentation models within Aurizn's workflows.

3 LITERATURE REVIEW

3.1 Introduction to Semantic Segmentation In Remote Sensing

Semantic segmentation is a fundamental task in computer vision, particularly relevant in remote sensing applications such as urban planning, land cover mapping, disaster response, and resource management [1]. The goal of semantic segmentation is to classify each pixel in an image into predefined categories, such as vegetation, buildings, roads, and water bodies, as illustrated in Figure 1. Recent advancements in DL have significantly improved segmentation accuracy compared to traditional machine learning and rule-based methods [10].

While DL-based approaches offer substantial improvements, their effectiveness depends on overcoming key challenges like data scarcity and class imbalance, which Aurizn faces, must be addressed to enhance image analysis and decision-making [6]. This project aims to tackle these challenges by exploring and adapting advanced DL models to Aurizn's specific multispectral dataset.

3.2 Deep Learning for Semantic Segmentation

3.2.1 Convolutional Neural Networks

CNN architectures like U-Net [11], DeepLabv3+ [12], and HRNet [13, 14] have shown promise in remote sensing, offering efficient feature extraction and fine-grained spatial detail capture. However, their limitations in capturing long-range dependencies, especially in high-resolution imagery, necessitate further investigation [15]. A core objective of this project is to determine the effectiveness of CNNs, particularly pre-trained CNNs, on Aurizn's datset and compare their performance against transformer-based models. This comparison will directly inform the selection of the most suitable architecture for Aurizn's operational needs.

3.2.2 Transformer-based Architectures

Transformers, including ViTs and Swin Transformers, offer improved long-range dependency modeling through self-attention [16]. The emergence of models like SAM, with its zero-shot and few-shot capabilities, is particularly relevant for Aurizn's dataset, where labeled data may be limited [17]. This project will explore and adapt SAM and other transformer-based models to Aurizn's multispectral imagery, assessing their ability to generalise and perform effectively in this specialised domain.

3.3 Multispectral Image Segmentation in Remote sensing

Multispectral satellite imagery captures reflected energy beyond the visible spectrum, as illustrated in Figure 2, provides crucial information for detailed land cover analysis [7]. While offering richer data for improved classification compared to RGB, it also presents unique challenges. The high dimensionality of multispectral data increases computational cost and can hinder model performance. Varying spectral resolutions and bandwidths across sensors require careful preprocessing, including atmospheric correction and radiometric calibration [8, 6].

This project will focus on effectively utilizing spectral information, addressing the complexities of multispectral data through techniques like spectral attention, multi-branch networks, and pan-sharpening [18].

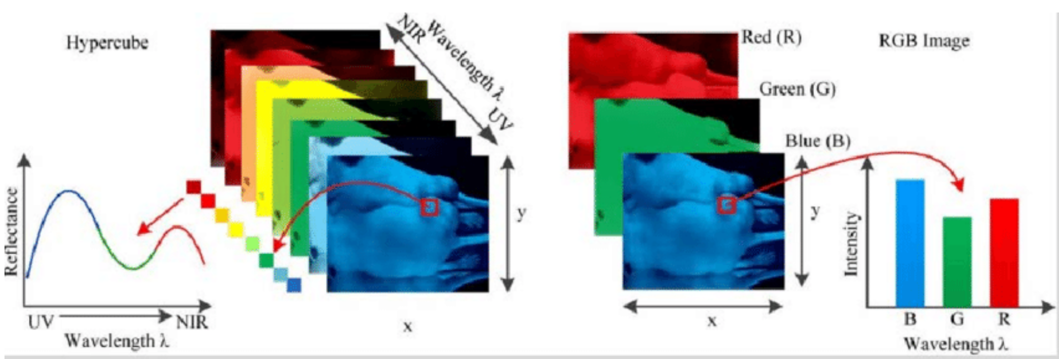

Figure 2: Difference between RGB and multispectral imaging. Image sourced from [19].

3.4 Transfer Learning and Fine-Tuning

3.4.1 Benefits of Transfer Learning

Transfer learning has been instrumental in remote sensing applications, allowing models pre-trained on large datasets to be adapted to specific tasks with limited labeled data. This is particularly beneficial for segmentation, where labeled data acquisition is expensive and time-consuming [20].

3.4.2 Fine-Tuning Pretrained RGB Models on Multispectral Images

Adapting RGB-trained models to multispectral data requires careful adjustments, including band-wise feature fusion and spectral attention [21, 22, 23]. A key objective is to investigate the effectiveness of fine-tuning pre-trained RGB models on Aurizn's multispectral dataset, comparing their performance with baseline HRNet model to find the most efficient approach.

3.5 Vision-Language Models (VLMs) for Remote Sensing

VLMs like GeoChat offer new possibilities for remote sensing through natural language interaction [24, 25]. As an exploratory objective, this project will explore the application of VLMs to Aurizn's data for tasks like temporal understanding, object counting, and detailed captioning. This will demonstrate the potential of VLMs to Aurizn and its clients, showcasing their ability to extract complex information from satellite imagery.

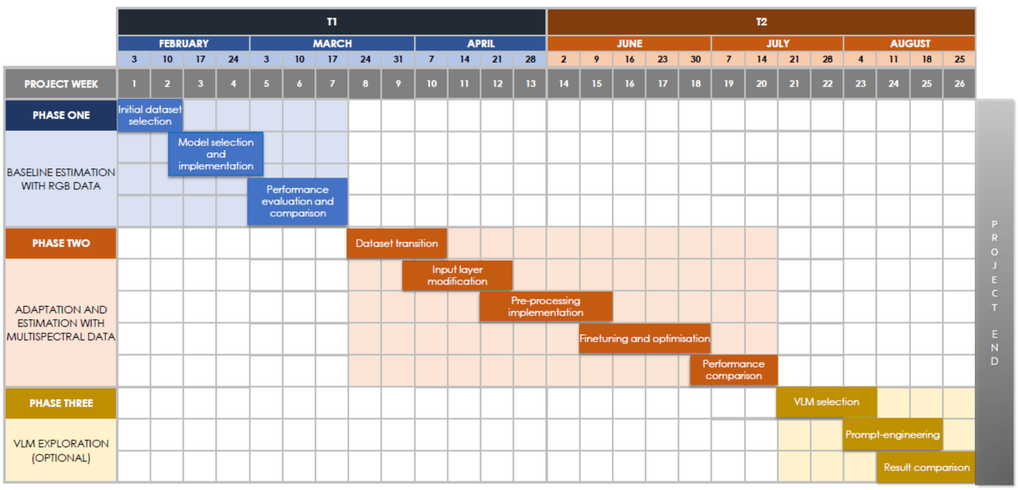

Figure 3: Project Timeline.

4 METHODOLOGY

The project design includes the following steps, with the planned timeline illustrated in Figure 3:

4.1 Phase 1: Baseline Establishment with RGB Data

- **Initial Dataset Selection:** Experiments will begin using a smaller, readily available RGB image dataset to establish a baseline and gain initial insights into model performance.
- **Model Selection and Implementation:**
  - This phase will evaluate the efficiency of various pre-trained models, such as DeepLabV3 with ResNet backbones, alongside models initialized from scratch to establish a comprehensive performance baseline.
  - The range of tested models will be expanded to include additional transformer and CNN architectures relevant to remote sensing, with both pre-trained weights and training from scratch will be applicable.
- **Performance Evaluation and Comparison:** The performance of each model will be rigorously evaluated using standard segmentation metrics [26], including:
  - Dice coefficient,
  - Intersection over Union (IoU),
  - Overall Accuracy, and
  - Per-class Precision and Recall.

- **Model Selection for Multispectral Adaptation:** Based on the comparative analysis, promising model architectures will be selected for adaptation to multispectral data.

4.2 Phase 2: Adaptation and Optimisation for Multispectral Data

- **Dataset Transition:** The established methodologies and selected models will be transitioned to Aurizn's high-resolution multispectral dataset.
- **Input Layer Modification:** Input layers of the selected models will be modified to accommodate the multiple spectral bands of the multispectral data.
- **Preprocessing Implementation:** Appropriate preprocessing techniques will be implemented to optimize the multispectral data for model input. This may include radiometric calibration, normalization, and pan-sharpening to enhance resolution when adapting RGB pre-trained models.
- **Fine-Tuning and Optimisation:** The fine-tuning process will be optimised for multispectral data, exploring techniques such as spectral attention mechanisms and band-wise feature fusion.
- **Comparative Performance Analysis:** The performance of the fine-tuned pre-trained models will be directly compared to a baseline HRNet model using the same evaluation metrics as in Phase 1.

4.3 Phase 3: VLM Exploration

- **VLM Selection**: As a stretch goal, the potential of Vision-Language Models (VLMs) like GeoChat will be explored for advanced tasks such as temporal understanding, referring segmentation, scene understanding, counting, and detailed image captioning.
- **Prompt Engineering:** Prompts will be used to test VLM capabilities against the data.
- **Result Comparison:** The results of the VLMs will be compared to ground truth data.

5 CHALLENGES

5.1 Multispectral Data Handling

Managing high-dimensional data and variations in sensor resolutions can complicate feature extraction and model performance.

5.2 Class Imbalance

Class imbalance can hinder model training and affect segmentation accuracy.

5.3 Fine-Tuning Pre-Trained Models

Adapting RGB-based models to multispectral data requires special techniques like band-wise feature fusion and can lead to a loss of spatial resolution.

5.4 Model Generalisation

Ensuring models generalise well across diverse geographic regions and handle data drift over time remains a challenge.

5.5 Computational Resources

High-resolution imagery demands substantial computational power and memory, making efficient resource management crucial.

REFERENCES

[1] J. Cheng, C. Deng, Y. Su, Z. An, and Q. Wang, "Methods and datasets on semantic segmentation for unmanned aerial vehicle remote sensing images: A review," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 211, pp. 1–34, 2024.

[2] T. Arulananth, P. Kuppusamy, R. K. Ayyasamy, S. M. Alhashmi, M. Mahalakshmi, K. Vasanth, and P. Chinnasamy, "Semantic segmentation of urban environments: Leveraging u-net deep learning model for cityscape image analysis," *Plos one*, vol. 19, no. 4, p. e0300767, 2024.

[3] A. Alzu'bi and L. Alsmadi, "Monitoring deforestation in jordan using deep semantic segmentation with satellite imagery," *Ecological Informatics*, vol. 70, p. 101745, 2022.

[4] M. Rahnemoonfar, T. Chowdhury, and R. Murphy, "Rescuenet: A high resolution uav semantic segmentation dataset for natural disaster damage assessment," *Scientific data*, vol. 10, no. 1, p. 913, 2023.

[5] Z. Cai, Q. Hu, X. Zhang, J. Yang, H. Wei, Z. He, Q. Song, C. Wang, G. Yin, and B. Xu, "An adaptive image segmentation method with automatic selection of optimal scale for extracting cropland parcels in smallholder farming systems," *Remote Sensing*, vol. 14, no. 13, p. 3067, 2022.

[6] J. Lv, Q. Shen, M. Lv, Y. Li, L. Shi, and P. Zhang, "Deep learning-based semantic segmentation of remote sensing images: a review," *Frontiers in Ecology and Evolution*, vol. 11, p. 1201125, 2023.

[7] L. Ramos and A. D. Sappa, "Multispectral semantic segmentation for land cover classification: An overview," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2024.

[8] K. Zhang, F. Zhang, W. Wan, H. Yu, J. Sun, J. Del Ser, E. Elyan, and A. Hussain, "Panchromatic and multispectral image fusion for remote sensing and earth observation: Concepts, taxonomy, literature review, evaluation methodologies and challenges ahead," *Information Fusion*, vol. 93, pp. 227–242, 2023.

[9] A. J. Davies, "Semantic segmentation of aerial imagery using u-net in python," Jan 2025.

[10] Y. Mo, Y. Wu, X. Yang, F. Liu, and Y. Liao, "Review the state-of-the-art technologies of semantic segmentation based on deep learning," *Neurocomputing*, vol. 493, pp. 626–646, 2022.

[11] X. Wang, Z. Hu, S. Shi, M. Hou, L. Xu, and X. Zhang, "A deep learning method for optimizing semantic segmentation accuracy of remote sensing images based on improved unet," *Scientific reports*, vol. 13, no. 1, p. 7600, 2023.

[12] H. Peng, C. Xue, Y. Shao, K. Chen, J. Xiong, Z. Xie, and L. Zhang, "Semantic segmentation of litchi branches using deeplabv3+ model," *Ieee Access*, vol. 8, pp. 164546–164555, 2020.

[13] S. Seong and J. Choi, "Semantic segmentation of urban buildings using a high-resolution network (hrnet) with channel and spatial attention gates," *Remote Sensing*, vol. 13, no. 16, p. 3087, 2021.

[14] J. Bai, C. Jia, S. Yu, L. Sun, L. Zhang, Z. Chang, and A. Hou, "Building extraction from high-resolution remote sensing images using improved hrnet method," in *IGARSS 2024-2024 IEEE International Geoscience and Remote Sensing Symposium*, pp. 7982–7985, IEEE, 2024.

[15] F. Fogel, Y. Perron, N. Besic, L. Saint-André, A. Pellissier-Tanon, M. Schwartz, T. Boudras, I. Fayad, A. d'Aspremont, L. Landrieu, *et al.*, "Open-canopy: A country-scale benchmark for canopy height estimation at very high resolution," *arXiv preprint arXiv:2407.09392*, 2024.

[16] Y. Zhang, M. Huang, Y. Chen, X. Xiao, and H. Li, "Land cover classification in high-resolution remote sensing: using swin transformer deep learning with texture features," *Journal of Spatial Science*, pp. 1–25, 2024.

[17] X. Ma, Q. Wu, X. Zhao, X. Zhang, M.-O. Pun, and B. Huang, "Sam-assisted remote sensing imagery semantic segmentation with object and boundary constraints," *IEEE Transactions on Geoscience and Remote Sensing*, 2024.

[18] J. Kaur, "Revolutionizing pan sharpening in remote sensing with cutting-edge deep learning optimization," in *2024 Second International Conference on Intelligent Cyber Physical Systems and Internet of Things (ICoICI)*, pp. 1357–1362, IEEE, 2024.

[19] S. Koundinya, H. Sharma, M. Sharma, A. Upadhyay, R. Manekar, R. Mukhopadhyay, A. Karmakar, and S. Chaudhury, "2d-3d cnn based architectures for spectral reconstruction from rgb images," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pp. 844–851, 2018.

[20] Y. Ma, S. Chen, S. Ermon, and D. B. Lobell, "Transfer learning in environmental remote sensing," *Remote Sensing of Environment*, vol. 301, p. 113924, 2024.

[21] J. Ouyang, P. Jin, and Q. Wang, "Multimodal feature-guided pre-training for rgb-t perception," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2024.

[22] A. Pendota and S. S. Channappayya, "Are deep learning models pre-trained on rgb data good enough for rgb-thermal image retrieval?," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4287–4296, 2024.

[23] M. Noman, M. Naseer, H. Cholakkal, R. M. Anwer, S. Khan, and F. S. Khan, "Rethinking transformers pre-training for multi-spectral satellite imagery," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 27811–27819, 2024.

[24] X. Li, C. Wen, Y. Hu, Z. Yuan, and X. X. Zhu, "Vision-language models in remote sensing: Current progress and future trends," *IEEE Geoscience and Remote Sensing Magazine*, 2024.

[25] C. Liu, J. Zhang, K. Chen, M. Wang, Z. Zou, and Z. Shi, "Remote sensing temporal vision-language models: A comprehensive survey," *arXiv preprint arXiv:2412.02573*, 2024.

[26] G. V. Vlăsceanu, N. Tarbă, M. L. Voncilă, and C. A. Boiangiu, "Selecting the right metric: A detailed study on image segmentation evaluation," *BRAIN. Broad Research in Artificial Intelligence and Neuroscience*, vol. 15, no. 4, pp. 295–318, 2024.