JOURNAL OF EARTH SCIENCE AND GEOMATICS INDONESIA

Volume 2 (2026), Page 1-16, E-ISSN 3163-9763

Road Analytics In Multiresolution Aerial Images

Mehdi Ravanbakhsh [1,2]*, Milad Ghorbani[2]

1. University of Western Australia, Western Australia
2. Mapizy Pty Ltd, Cloverdale, Western Australia

*corresponding author: mehdi.r@mapizy.com

Abstract

Accurate, detailed and up-to-date road networks are critical in many applications, autonomous driving being one of them. Currently, the existing road network data lacks sufficient detail and currency necessary to support various use cases such as asset management and monitoring as well as smart transport networks. This project has two main objectives. First, it has to be determined if ML/AI models can be developed to effectively and efficiently create derivative analytics about road network from high resolution aerial images. Second, these models could be applied to lower resolution satellite images to create similar quality road derivative products. Our quantitative analysis shows that over 95% accuracy of feature detection in aerial images across all features was achieved except for road surfaces where 92% accuracy was obtained. Transferring deep learning models from aerial to satellite images, around 15%-25% drop in accuracy was reported. However, technologies such as augmentation, super-resolution and transfer learning significantly improved the detection accuracy in satellite images at 7%-15%.

Keywords: Road Network Extraction; Deep Learning; Aerial Imagery; Satellite Imagery; Transfer Learning

Received: May 29, 2026
Revised: August 31, 2026
Accepted: September 18, 2026
Published: November 10, 2026

Citation: Ravanbakhsh , M.. (2026). Road Analytics In Multiresolution Aerial Images. Journal of Earth Science and Geomatics Indonesia, 2, 1-16.

Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license

1. Introduction

Accurate, detailed and up-to-date road information is fundamental to a wide range of geospatial applications, including infrastructure management, transportation planning, asset monitoring and autonomous driving. Deep learning-based semantic segmentation has become an important approach for extracting roads and other built and natural features from remotely sensed imagery. In particular, U-Net and its variants have demonstrated strong performance in pixel-level extraction of roads, buildings and other geospatial objects because of their ability to combine contextual information with fine spatial details (Diakogiannis et al., 2020; Wang et al., 2022). However, most existing studies focus on improving segmentation performance

within a particular imagery source or dataset, while the generalisation of a model across imagery acquired from different sensors, spatial resolutions and spectral characteristics remains a significant challenge. Reviews of deep-learning-based remote-sensing segmentation have similarly identified domain variability and model generalisation as important open challenges in operational remote-sensing applications.

This limitation is particularly important for road mapping because imagery acquisition conditions can vary substantially between aerial and satellite platforms. Aerial imagery can provide very high spatial resolution and detailed representation of road surfaces, lane markings and other road-corridor features, but acquisition can be relatively costly and spatially limited. Satellite imagery provides broader geographic coverage and more regular acquisition opportunities, but generally has lower spatial resolution and different spectral and radiometric characteristics. Consequently, a model trained using high-resolution aerial imagery may experience a substantial performance degradation when directly applied to satellite imagery. Recent studies on domain adaptation in remote sensing have shown that differences in ground sampling distance, sensor characteristics and geographic environment can produce significant domain shifts and reduce the generalisation of semantic-segmentation models.

Road extraction presents an additional challenge because the appearance, width and surrounding context of roads vary considerably between geographic regions and imagery resolutions. Recent research has therefore moved beyond conventional segmentation towards domain-adaptation approaches that explicitly address differences between source and target imagery. For example, domain-adaptive road segmentation methods have incorporated adversarial learning, self-training and topology-aware constraints to improve transfer between satellite and aerial domains. These studies demonstrate that simply applying a segmentation model trained in one domain to another is often insufficient, particularly when road width, background characteristics and spatial detail change substantially between domains.

Nevertheless, an important practical gap remains between these specialised domain-adaptation approaches and the operational use of existing segmentation models for multi-source geospatial analytics. In many real-world mapping applications, the objective is not necessarily to develop a new segmentation architecture, but to determine whether an existing high-performing model can be transferred efficiently to imagery with different spatial and spectral characteristics, while requiring only limited additional target-domain data. In this context, relatively simple and deployable techniques such as image augmentation, super-resolution and transfer learning may provide a practical alternative or complement to more complex domain-adaptation architectures. The comparative effectiveness of these techniques for transferring models trained on very-high-resolution aerial imagery to lower-resolution satellite imagery, particularly for multiple road-corridor features, remains insufficiently explored.

This study addresses this gap by investigating the transferability of deep-learning models developed from 10–12 cm aerial imagery to 30–50 cm satellite imagery, rather than treating aerial and satellite imagery as independent segmentation problems. The study uses U-Net-based models as a consistent baseline and systematically evaluates three strategies for reducing the cross-domain performance gap: (i) image augmentation to increase robustness to spatial and spectral variation, (ii) super-resolution to compensate for the loss of spatial detail in satellite imagery, and (iii) transfer learning using a limited amount of target-domain data. The study further examines the combined

use of super-resolution and augmentation. The objective is therefore not to introduce another U-Net architecture, but to quantify how existing segmentation models can be made more transferable across substantial changes in sensor and spatial resolution.

A further distinction of this work is the breadth of the geospatial features considered. Rather than evaluating road segmentation alone, the models are used to extract road surfaces, lane markings, traffic islands, dividers, trees, turf and buildings within and around road corridors. Lane markings are additionally classified into broken, solid and double-solid categories. This provides an assessment of cross-domain transfer at the level of a broader road-network analytics workflow rather than a single binary road-extraction task. The datasets used in this study cover Queensland and Western Australia, providing variation in both geographic environment and imagery source.

The main contribution of this study is consequently an experimental framework for assessing and improving the cross-resolution and cross-sensor transferability of geospatial deep-learning models. The results quantify the performance degradation that occurs when aerial-trained models are applied directly to satellite imagery and evaluate the extent to which augmentation, super-resolution and transfer learning can recover this loss. This provides practical evidence for developing scalable road-network analytics in situations where high-resolution aerial training data are available but operational mapping must be performed using more widely available satellite imagery.

2. Materials and Methods

The processing workflow proposed is shown in Figure 1. It includes four steps as follows. First, training samples representing various targeted road features, namely road surface, lane markings, traffic islands, divider, tree cover and turf, are collected. Mapizy has already created a database of training data of features around the road and in the road corridor which can be used with minimal preparation. Next, AI models will be created and target features will be detected. A range of features such as roads, buildings, vegetation, road width, traffic islands, dividers and marking lines will be extracted. This step requires intensive fine-tuning of deep learning optimisation parameters.

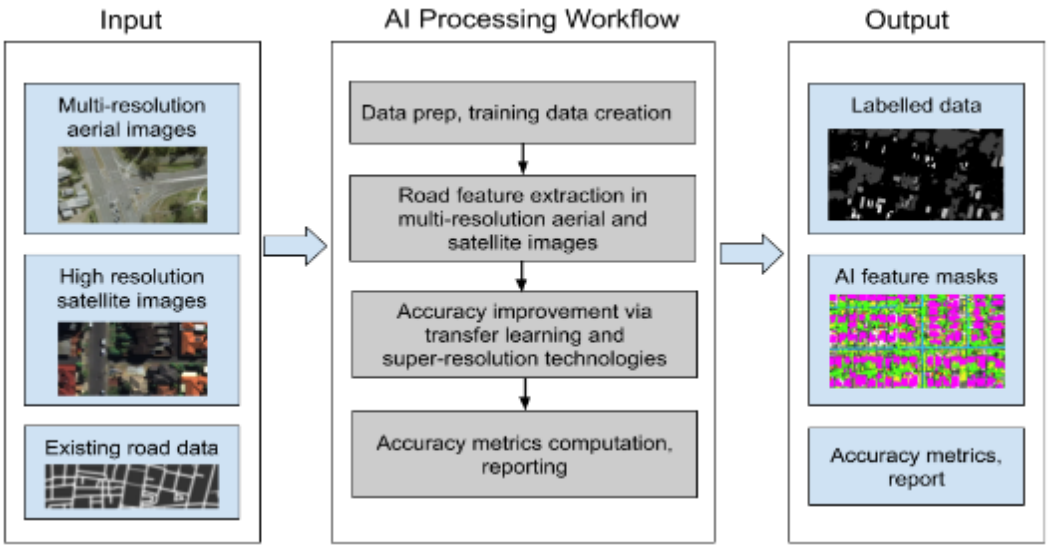


Figure 1. Project workflow

After that, the quality of deep learning results will be improved through utilisation of transfer learning, augmentation and super-resolution technologies. The result obtained through deep learning will be evaluated against reference data and quality metrics will be estimated for the selected areas. Further test areas will be visually assessed by human inspection. The outcome of this section will be quality metrics estimated to assess the performance of our approach.

Deep learning have been the technology of choice when it comes to man-made object detection (Choi et al., 2021; Kang et al., 2022; Diakogiannis et al., 2020; Zhang et al., 2022; Zhang et al., 2020; Wang et al., 2022). Among various deep learning architectures developed in the literature, UNET has shown promising performance (Diakogiannis et al., 2020; Wang et al., 2022), which was adopted in Mapizy and used in a range of projects in mapping man-made and natural objects in aerial and satellite images. The architecture of the model is shown in Fig. 2. This model uses convolutional layers to extract features of a given image and output a semantic mask. In this project, we used two versions of UNET, namely light version or UNETv3 and a deeper version or UNETv4. Deeper version has more layers so it can extracted greater number of image features and its training takes more time as larger number of model parameters need to be optimised.

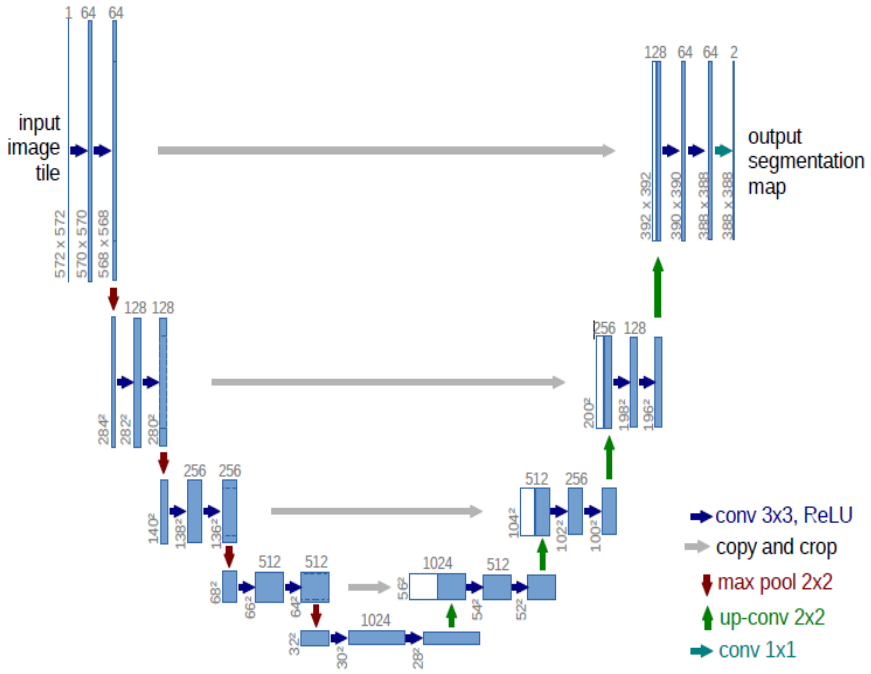


Figure 2. UNET architecture. Each blue box corresponds to a multi-channel feature map. The number of channels is denoted on top of the box. The x and y size is provided at the lower left edge of the box. White boxes represent copied feature maps and arrows denote the operation performed.

AI models developed on training data coming from aerial images cannot be directly applied to other datasets like satellite images where resolution and spectral quality differs (see Fig. 3). In the literature, there are a couple of technologies used to improve the accuracy of detection, these being augmentation, image super-resolution and transfer learning (Shorten & Khoshgoftaar, 2019; Liang et al., 2021; Hoar et al., 2021).



Figure 3. Aerial imagery (left). Satellite image of the same area (right).

Image augmentation is a technique that is used to artificially expand the training data. Sufficient training data is crucial for optimal deep learning model development. Otherwise, AI models overfit the training data resulting in poor performance when applied on unseen data. In augmentation, various geometric and spectral transformations are applied to the training data to accommodate variations often present in real world images. This helps deep learning models to learn the true underlying pattern in the data and thus create more accurate result when applied to new datasets.

In this study, we employed a range of image transformations such as resizing, spectral channel shift, HSV, etc. applied on labelled data collected in aerial images. These transformations make aerial deep learning models more transferable to satellite images (Shorten & Khoshgoftaar, 2019).

Three data sources were provided by the TMR team and sourced through Mapizy. These are aerial images at 10cm resolution, and Airbus Pleiades and Pleiades Neo images at 50cm and 30cm resolutions respectively. The TMR's data was from QLD only and Mapizy data covers Perth densely populated urban areas as well as two large areas in QLD. Fig. 4 shows satellite images sourced by Mapizy covering Redland and Ipswich LGAs in QLD.

Test areas were chosen in WA and QLD for both accuracy metric quantification and visual assessment. In three WA test sites and one QLD test site as shown in Fig. 5, reference data was created and used for quality metrics estimation.



Figure 4. Airbus Pleiades Neo at 30cm resolution for Redland (left) and Ipswich (right) in QLD.



Figure 5. Outlines of the test areas in WA (left) and QLD, Ipswich (right) used for quantitative accuracy assessment.

3. Results

The collected labelled data was used to train and create deep learning models used for the detection of roads, line markings, traffic islands, dividers, buildings, trees and turf. Some sample results are shown in Figs. 6, 7, 8 & 9.



Figure 6. Road detection result in 10cm aerial (left) and 50cm satellite (right) images.

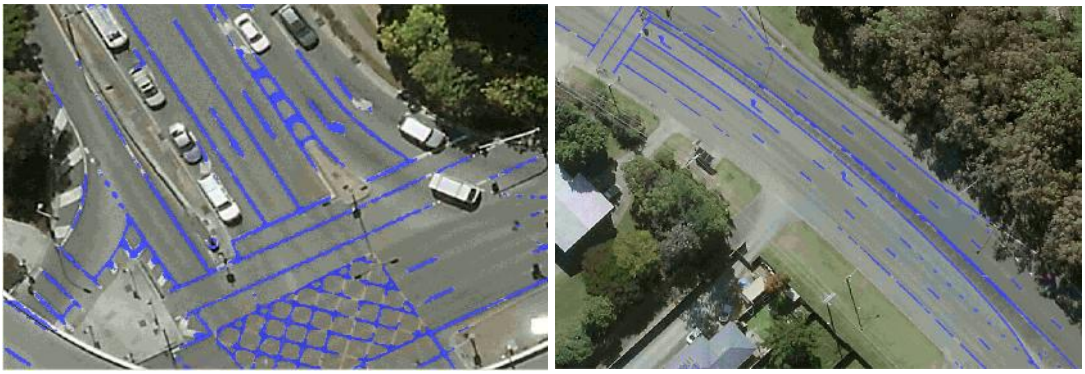


Figure 7. Line markings detection in 10cm aerial images.



Figure 8. Building, tree and turf detection in 10cm (left) and 50cm (right) aerial images.

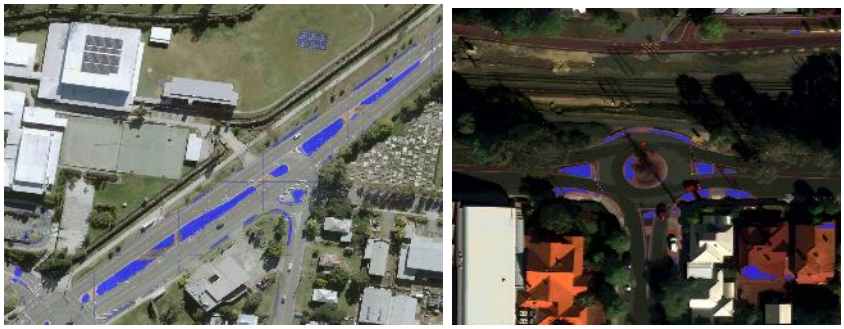


Figure 9. Traffic island detection in 10cm aerial (left) and 50cm satellite images (right).

3.1 Quality metrics

The performance of the deep-learning models was evaluated using accuracy, precision and recall. These metrics were calculated from the numbers of true-positive (TP), true-negative (TN), false-positive (FP), and false-negative (FN) pixels obtained by comparing the predicted segmentation masks with the reference data. The metrics are defined as follows: [8]:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad \text{..... (1)}$$

$$\text{Precision} = \frac{TP+TN}{TP+TN+FP} \quad \text{..... (2)}$$

$$\text{Recall} = \frac{TP+TN}{TP+TN+FN} \quad \text{..... (3)}$$

Accuracy measures the proportion of correctly classified pixels in the complete image. Precision measures the proportion of pixels predicted as the target feature that are actually part of the reference feature, thereby indicating the level of false-positive detection. Recall measures the proportion of reference feature pixels successfully detected by the model and therefore indicates the extent of missed features. These metrics are complementary and are appropriate for road-feature segmentation because both omission and commission errors can affect the quality and completeness of extracted road features.

3.2 Augmentation

Augmentation techniques were used to expand the training dataset, which included samples of varying resolutions and spectral characteristics. This increased the model's accuracy but required more resources and time. The models were tested on high-resolution (10cm) and low-resolution (50cm) aerial images, and low-resolution (50cm) satellite images.

The result presented in Tabs. 1,2 and 3 shows that augmentation has been effective in improving the detection accuracy across image resolutions. However, in the spectral space as can be seen in satellite images, the improvement wasn't observed.

Table 1. Accuracy assessment result estimated for road detection in test areas in Perth. Accuracy is shown in percentage.

AI model	**UNETv3**			**UNETv3 + augmentations**		
Image data	**Aerial High-resolution**	**Aerial Low-resolution**	**Satellite Low-resolution**	**Aerial High-resolution**	**Aerial Low-resolution**	**Satellite Low-resolution**
Test Area 1	95.59	90.77	68.97	95.21	93.10	69.30
Test Area 2	95.14	90.54	70.89	94.52	92.36	69.99
Test Area 3	95.71	88.58	78.19	94.98	92.10	77.90
Average	95.48	89.96	72.68	94.9	92.52	72.4

Table 2. Accuracy assessment result estimated for road detection in test areas in Perth. Precision is shown in percentage.

AI model	**UNETv3**			**UNETv3 + augmentations**		
Image data	**Aerial High-resolution**	**Aerial Low-resolution**	**Satellite Low-resolution**	**Aerial High-resolution**	**Aerial Low-resolution**	**Satellite Low-resolution**
Test Area 1	98.85	97.99	88.94	98.03	97.37	89.76
Test Area 2	98.51	97.79	94.49	98.29	97.3	93.2
Test Area 3	98.5	96.11	92.17	97.96	96.47	91.71
Average	98.62	97.3	91.87	98.09	97.05	91.56

Table 3. Accuracy assessment result estimated for road detection in test areas in Perth. Recall is shown in percentage .

AI model	**UNETv3**			**UNETv3 + augmentations**		
Image data	**Aerial High-resolution**	**Aerial Low-resolution**	**Satellite Low-resolution**	**Aerial High-resolution**	**Aerial Low-resolution**	**Satellite Low-resolution**
Test Area 1	96.74	92.78	80.03	97.18	95.73	79.54
Test Area 2	96.63	92.75	76.4	96.23	95.06	76.79
Test Area 3	97.21	92.47	86.02	97.02	95.63	86.19
Average	96.86	92.66	80.81	96.81	95.47	80.84

Two example results showing the effect of augmentation on satellite images are shown in Figs. 10 & 11.



Figure 10. Detection of buildings in satellite images with and without augmentation shown in red and green colours respectively.



Figure 11. Detection of road network with augmentation in satellite images. Without augmentation no road was detected.

3.3 Super-resolution

Super resolution was applied on satellite images to improve the spatial resolution creating more detailed information about the features. The quantitative result as presented in Tab. 4 shows that super-resolution has been effective in improving the detection accuracy across image resolutions. In addition to this, training time is shorter and less computational resources are required. In Fig. 12, the effect of super-resolution for road surface detection is shown.

Table 4. Accuracy assessment result estimated for road detection in test areas in Perth. Accuracy is shown in percentage.

AI model	**UNETv3**			**UNETv3 + Super-resolution**		
Image data	**Aerial High-resolution**	**Aerial Low-resolution**	**Satellite Low-resolution**	**Aerial High-resolution**	**Aerial Low-resolution**	**Satellite Low-resolution**
Test Area 1	95.59	90.77	68.97	95.59	93.48	82.21

AI model	UNETv3			UNETv3 + Super-resolution		
Image data	Aerial High-resolution	Aerial Low-resolution	Satellite Low-resolution	Aerial High-resolution	Aerial Low-resolution	Satellite Low-resolution
Test Area 2	95.14	90.54	70.89	95.14	94.27	82.59
Test Area 3	95.71	88.58	78.19	95.71	93.75	82.88
Average	95.48	89.96	72.68	95.48	93.83	82.56

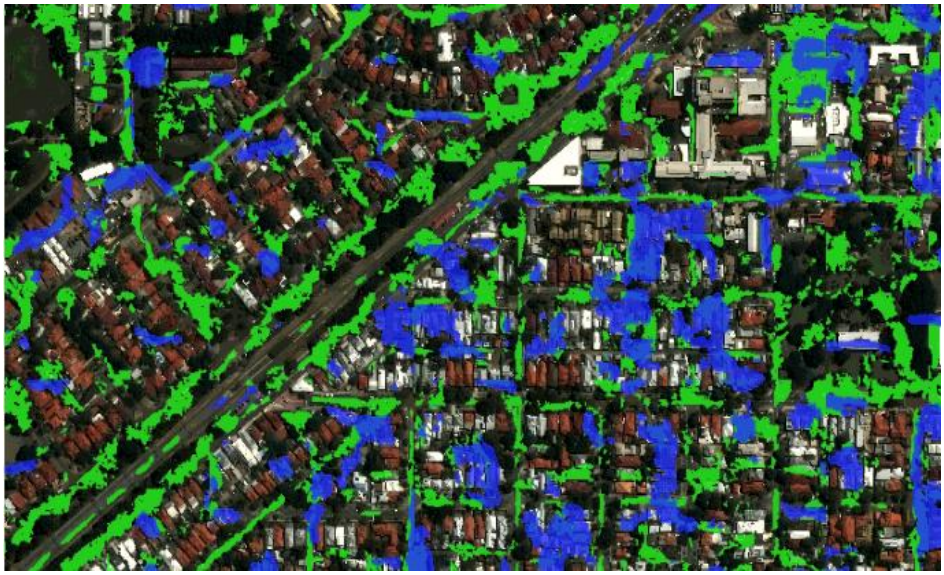


Figure 12. Detection of roads with and without super-resolution in green and blue colours respectively.

3.4 Augmentation plus super-resolution

Super-resolution transforms images into higher resolution datasets. As a result, augmentation transformations associated with the resolution such as zoom level isn't required. Instead, we extensively used spectral transformations, HSV being one of them, so that both geometric and spectral aspects are taken into account in a combined approach.

In Tabs. 5, and 6, the comparative performance of accuracy improvement technologies, these being super-resolution, augmentation and their combination.

The result of these experiments suggests that the combination of super-resolution and spectral augmentations can further improve the accuracy of deep learning models and that the performance of this combination is better than that of augmentation or super-resolution alone.

Table 5. Accuracy assessment result estimated for road surface detection in Ipswich test area using UNETv3 model. Accuracy is shown in percentage.

AI model	UNETv3			UNETv3 + Accuracy improvement		
Accuracy improvement approach	Aerial High-resolution	Aerial Low-resolution	Satellite Low-resolution	Aerial High-resolution	Aerial Low-resolution	Satellite Low-resolution
Super Resolution	92.95	88.05	67.72	92.95	91.93	74.71
Augmentation	92.95	88.05	67.72	92.95	88.46	73.24
Super Resolution + Augmentation	92.95	88.05	67.72	92.95	92.04	77.35

Table 6. Accuracy assessment result estimated for road surface detection in Ipswich test area using UNETv4 model. Accuracy is shown in percentage.

AI model	UNETv4			UNETv4 + Accuracy improvement		
Accuracy improvement approach	Aerial High-resolution	Aerial Low-resolution	Satellite Low-resolution	Aerial High-resolution	Aerial Low-resolution	Satellite Low-resolution
Super Resolution	92.24	89.98	77.46	92.24	92.51	80.76
Augmentation	92.24	89.98	77.46	92.24	88.88	81.36
Super Resolution + Augmentation	92.24	89.98	77.46	92.24	91.7	83.76

3.5 Transfer learning

Transfer learning differs compared to other accuracy improvement technologies discussed previously. While transfer learning uses a modest amount of training data from satellite images to retrain the model trained on aerial images, others either make satellite images look like aerial images or improve the accuracy of deep learning models through augmentation. Tabs. 7 and 8 show the comparative performance of transfer learning and other technologies applied on road surface detection.

Table 7. Accuracy assessment result estimated for road surface detection in Ipswich test area using UNETv3 model. Accuracy is shown in percentage.

AI model	UNETv3			UNETv3 + Accuracy improvement		
Accuracy improvement approach	Aerial High-resolution	Aerial Low-resolution	Satellite Low-resolution	Aerial High-resolution	Aerial Low-resolution	Satellite Low-resolution
Super Resolution	92.95	88.05	67.72	92.95	91.93	74.71
Augmentation	92.95	88.05	67.72	92.95	88.46	73.24

Super Resolution + Augmentation	92.95	88.05	67.72	92.95	92.04	77.35
Transfer Learning	92.95	88.05	67.72	92.95	91.57	77.94

Table 8. Accuracy assessment result estimated for road surface detection in Ipswich test area using UNETv4 model. Accuracy is shown in percentage .

AI model	**UNETv4**			**UNETv4 + Accuracy improvement**		
Accuracy improvement approach	**Aerial High-resolution**	**Aerial Low-resolution**	**Satellite Low-resolution**	**Aerial High-resolution**	**Aerial Low-resolution**	**Satellite Low-resolution**
Super Resolution	92.24	89.98	77.46	92.24	92.51	80.76
Augmentation	92.24	89.98	77.46	92.24	88.88	81.36
Super Resolution + Augmentation	92.24	89.98	77.46	92.24	91.7	83.76
Transfer Learning	92.24	89.98	77.46	92.24	91.82	84.28

5. Conclusions

In this project, high-performing deep-learning models, computer vision methods and geospatial processing procedures were developed to identify and delineate features on and around roads as well as in road corridor. Seven different features, these being road surface, lane marking, traffic island, divider, tree, turf and buildings, were targeted and lane markings were further classified in three categories of broken, solid and double-solid lines.

Three different technologies were assessed to determine if these can improve the accuracy of AI models trained on aerial images to work on satellite images. The performance of these technologies in feature detection is shown in Fig. 13. A computer vision solution was also developed to create road polygons in vector format from road surface detection masks.

Our quantitative analysis shows that we achieved over 95% accuracy of feature detection in aerial images across all features except for road surfaces where 92% accuracy was obtained. Transferring deep learning models from aerial to satellite images, around 15%-25% drop in accuracy was reported. However, technologies such as augmentation, super-resolution and transfer learning significantly improved the detection accuracy in satellite images at roughly 7%-15% (see Fig. 13).

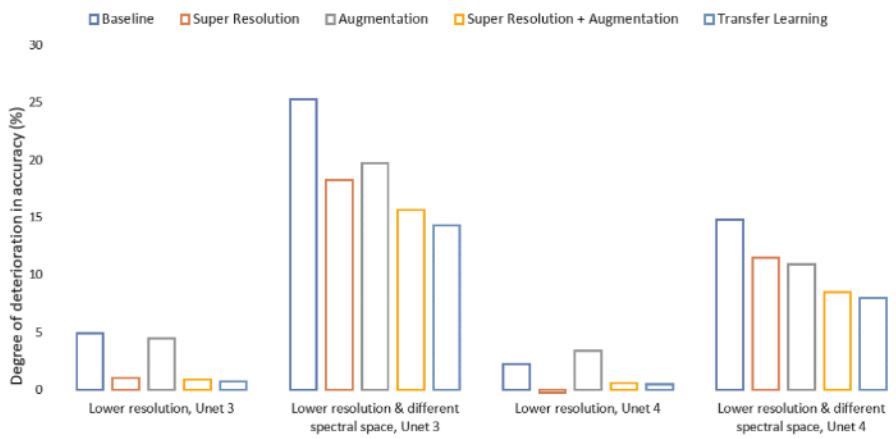


Figure 13. Comparative performance of AI accuracy improvement technologies.

6. Limitations

In this project, we assessed the capability of AI technology to detect features on and around roads as well as in the road corridor in multi-resolution aerial and satellite images. Two example results showing how our solution performed are shown in Figs. 14 and 15.



Figure 14. Sample result of all target features extracted in an aerial image for a large area.

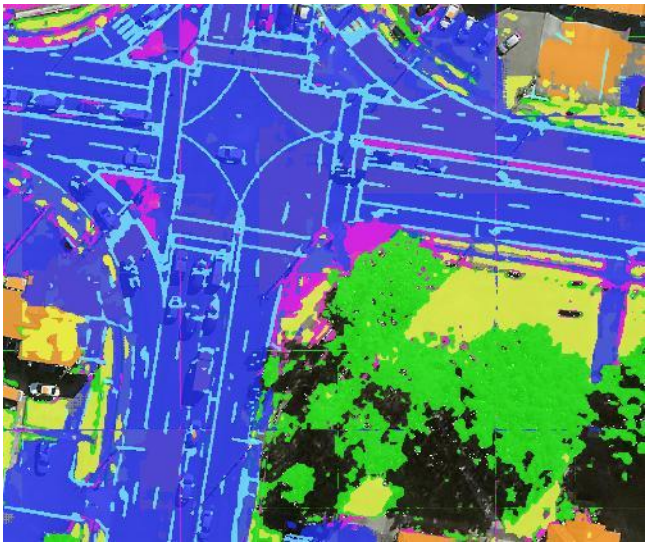


Figure 15. Close-up sample result of all target features extracted in an aerial image.

In order to improve the performance of the current solution, the following are recommended:

1. Our solution was developed with limited amount of training data. Although the results obtained for some features are satisfactory but features such as traffic islands were affected by lack of training data. More training data is needed to improve the accuracy across all features.
2. We adopted UNET architecture as industry standard model for feature extraction as it has shown excellent performance in various applications. However, the deep learning architecture could be further customised to create a better result. The impact of model architecture in feature extraction is an active research area and more work is required to assess variants of UNET and other similar architectures.
3. We used three technologies to improve the performance of models when transferring them from aerial image to satellite images. However, except for one case which is super resolution plus augmentation, we didn't assess the combination of these technologies for feature extraction. Various combinations and their impact on quality metrics of feature extraction in multi-source datasets are warranted.
4. Although it was not in the scope, we managed to develop a solution for the generation of road polygons in vector format. The initial result looks promising and by adding more diverse training data for road surface detection, high quality road vectors can be created for the QLD's extensive road network.

Accuracy assessment of our solution is limited to test areas in WA and QLD. This is because reference data creation used for accuracy assessment is time-consuming. Comprehensive accuracy assessment for various geographical areas such as NT and TAS states is recommended to rigorously quantify the model performance and identify improvement areas. Having a solution that performs equally well across different urban and rural landscapes increases its scalability and enhance the chance of commercialisation.

7. Declaration of Interest

The authors declare that I have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Choi, W., Heo, J., & Ahn, C. (2021). Development of road surface detection algorithm using CycleGAN-augmented dataset. Sensors, 21(22), 7769. https://doi.org/10.3390/s21227769

Diakogiannis, F. I., Waldner, F., Caccetta, P., & Wu, C. (2020). ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data. ISPRS Journal of Photogrammetry and Remote Sensing, 162, 94–114. https://doi.org/10.1016/j.isprsjprs.2020.01.013

Hoar, D., Lee, P. Q., Guida, A., Patterson, S., Bowen, C. V., Merrimen, J., Wang, C., Rendon, R., Beyea, S. D., & Clarke, S. E. (2021). Combined transfer learning and test-time augmentation improves convolutional neural network-based semantic segmentation of prostate cancer from multi-parametric MR images. Computer Methods and Programs in Biomedicine, 210, 106375. https://doi.org/10.1016/j.cmpb.2021.106375

Kang, J., Fernandez-Beltran, R., Sun, X., Ni, J., & Plaza, A. (2022). Deep learning-based building footprint extraction with missing annotations. IEEE Geoscience and Remote Sensing Letters, 19, 1–5. https://doi.org/10.1109/LGRS.2021.3072589

Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L., & Timofte, R. (2021). SwinIR: Image restoration using Swin Transformer. In Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (ICCVW) (pp. 1833–1844). IEEE. https://doi.org/10.1109/ICCVW54120.2021.00210

Padilla, R., Netto, S. L., & da Silva, E. A. B. (2020). A survey on performance metrics for object-detection algorithms. In 2020 International Conference on Systems, Signals and Image Processing (IWSSIP) (pp. 237–242). IEEE. https://doi.org/10.1109/IWSSIP48289.2020.9145130

Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. Journal of Big Data, 6, Article 60. https://doi.org/10.1186/s40537-019-0197-0

Wang, Y., Peng, Y., Li, W., Alexandropoulos, G. C., Yu, J., Ge, D., & Xiang, W. (2022). DDU-Net: Dual-decoder-U-Net for road extraction using high-resolution remote sensing images. IEEE Transactions on Geoscience and Remote Sensing, 60, 1–12. https://doi.org/10.1109/TGRS.2022.3197546

Zhang, X., Ma, W., Li, C., Wu, J., Tang, X., & Jiao, L. (2020). Fully convolutional network-based ensemble method for road extraction from aerial images. IEEE Geoscience and Remote Sensing Letters, 17(10), 1777–1781. https://doi.org/10.1109/LGRS.2019.2953523

Zhang, Y., Lu, Z., Zhang, X., Xue, J.-H., & Liao, Q. (2022). Deep learning in lane marking detection: A survey. IEEE Transactions on Intelligent Transportation Systems, 23(7), 5976–5992. https://doi.org/10.1109/TITS.2021.3070111