Enhancing Aerial Data Semantic Segmentation with a Colour Range Mask Layer: A Deep Learning Approach

Ali Ahmadi[a,*], Mir Abolfazel Mostafavi[a], Nouri Sabo[a,b]

[a] *Research centre in geospatial data and intelligence, University Laval, ali.ahmadi.2@ulaval.ca, mir-abolfazl.mostafavi@scg.ulaval.ca*
[b] *Canada Centre for Mapping and Earth Observation, nouri.sabo@nrcan-rncan.gc.ca*
* Corresponding author

Keywords: U-Net, Ultra-High-Resolution, Semantic Segmentation

Abstract:

Recent advancements in airborne platforms equipped with ultra-high-resolution imaging sensors have significantly improved our capability to acquire detailed urban optical imagery. These systems offer exceptional capabilities for capturing highly precise and detailed urban data, paving the way for the generation of high-definition maps (HD maps) for innovative urban applications. However, manually extracting information from this data is a generally slow and labour-intensive process. Thus, employing deep learning algorithms for data extraction in such a context might be an alternative solution. Deep learning has revolutionised and transformed remote sensing and image analysis, especially in semantic segmentation, which divides images into meaningful regions. This transformative power of deep learning is particularly significant in urban analysis (e.g., urban planning, navigation, disaster management, and monitoring infrastructure), where detailed spatial information is crucial. Even though deep learning offers excellent potential, applying deep learning for semantic segmentation of images from urban environments presents several challenges. First, supervised deep learning algorithms require many training data to work effectively. Second, training and analysing ultra-high-resolution (less than 5 cm) images with deep learning algorithms need large storage capacity, are computationally intensive and often require advanced data augmentation, pre-processing, and model optimisation techniques to achieve optimal results Zhu et al., (2017).

Most deep learning algorithms in computer vision employ CNN algorithms to analyse and extract features from 2D input data. However, in specific scenarios, such as when defining a small number of features for the initial CNN layer, the model may struggle to correctly identify patterns or gain a proper understanding of the data. This can lead to difficulties in object recognition, resulting in errors and misclassifications. Therefore, numerous researchers propose to use multi-sensory data and fusion of diverse datasets to improve the accuracy of their results Zhong et al., (2021). However, using multi-sensory data requires more developed algorithms, memory, and computation resources Zhao et al., (2021). Therefore, some studies tried to solve this issue by adding more features, such as edge maps, Lyu et al., (2018), and Hue Saturation Intensity (HIS) Idrees et al., (2015) which is extracted from the original image. However, these data are not very useful because edge maps can be produced by deep learning through edge filters, and the HIS images are merely transformations of the same image into another visualisation standard.

To tackle these challenges, we propose a new approach called the Colour Range Mask (CRM) layer, which facilitates in-depth understanding and assessment of input images. This method enhances the effectiveness of U-Ne models, which perform well with limited training data. The CRM layer serves as a specialised pre-processing tool for input images, focusing on segmenting the red, green, and blue colour channels into a set of distinct sub-range values. This innovative approach creates a unique mask for each defined sub-range, effectively isolating pixel values within that specific range while rendering all other pixel values to zero. Doing so enhances the ability to distinguish between different land cover types, such as vegetation, water bodies, roads, and sidewalks, based on their unique spectral signatures. This method improves the accuracy of land cover classification and facilitates a deeper analysis of the underlying environmental features depicted in the input images.

To evaluate the effectiveness of the CRM layer, we undertake a structured process consisting of three key steps. The first step involves data preparation, during which we create an annotation mask for all our RGB images. The second step involves model development, which includes creating a CRM layer to enhance the U-Net algorithm. In the final step, we train and evaluate all the models developed during the previous stages. To demonstrate that our proposed CRM layer can enhance the performance of image CNN-based deep learning algorithms, we modified an encoder-decoder-based algorithm (U-Net algorithm) demonstrated in Figure 1. The modified U-Net (U-Net + CRM layer) architecture processes input images in two main blocks. The first block, shown in Figure 1 in purple, includes a 5×5 convolution layer followed by a ReLU activation, then a 3×3 convolution layer, Batch Normalization, and another ReLU activation. This block takes an RGB image as input and produces 16 feature maps. The second block commences with the CRM layer, which analyses

Abstracts of the International Cartographic Association, 10, 2, 2025
32nd International Cartographic Conference (ICC 2025), 17–22 August 2025, Vancouver, Canada.
https://doi.org/10.5194/ica-abs-10-2-2025

RGB images to generate 12 colour range mask features for each image. These features are subsequently passed to the following block, which mirrors the first block. However, this block takes the 12 mask features and produces 16 feature maps in this instance. In the next step, all 32 produced feature maps are concatenated and continue to the next block as per the standard U-Net process.

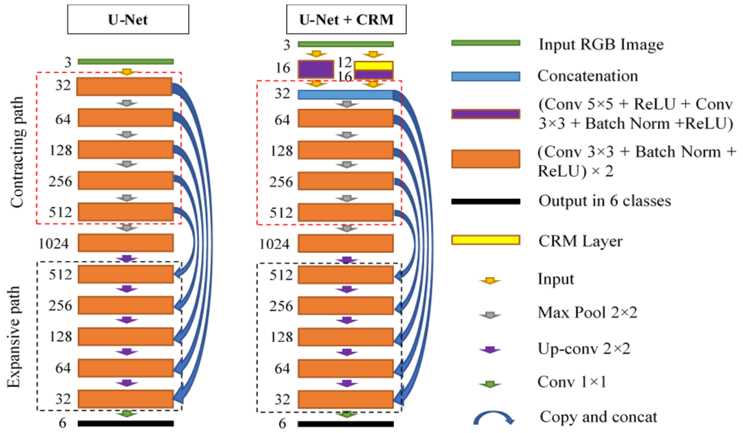

Figure 1. illustrates the U-Net algorithm and the modified U-Net model that includes the CRM layer. The U-Net algorithm and modified U-Net with the CRM layer.

We trained both algorithms on six classes—land, road, sidewalk, building, water body, and crosswalk—using consistent conditions: a multi-class cross-entropy loss function, the Adam optimiser Kingma & Ba, (2014), and the same hyperparameters and training epochs. Several metrics have been used to assess their performance, including Overall Accuracy, F1-score, and Mean Intersection over Union (mIoU). The results of our semantic segmentation models reveal significant insights into the accuracy and effectiveness of the proposed approaches for mapping urban environments. Table 1 shows that integrating the CRM layer significantly enhanced the performance of U-Net models across various evaluation metrics. Notably, the combination of U-Net with the CRM layer resulted in an impressive increase in mean Intersection over Union (mIoU), rising from 76.91% to 79.98% when compared to the standard U-Net model. This improvement highlights the effectiveness of the CRM layer in optimising the model's performance. To enhance our comparisons, we've integrated the pretrained ResNetUnet with the CRM layer results, and the outcomes are indeed promising, particularly in the mIoU metric, as Table 1 shows.

Metric	Overall Accuracy	F1 Score	mIoU
U-Net	94.33%	94.32%	76.91%
U-Net + CRM	**95.31%**	**95.30%**	**79.98%`**
ResNetUnet +CRM	**95.78%**	**95.77%**	**82.71%**

Table 1. compares the performance of U-Net and U-Net with a CRM layer using overall accuracy, F1-score, and mean Intersection over Union (mIoU) metrics.

In conclusion, the CRM layer marks a significant advancement in using CNN-based deep learning for semantic segmentation in urban analysis. By enhancing the U-Net architecture, the CRM layer effectively distinguishes different land cover types, improving classification accuracy and optimising resources. The systematic evaluation process highlights its effectiveness in analysing ultra-high-resolution imagery.

References

Idrees, M. O., Saeidi, V., Pradhan, B., Shafri, H., Oludare Idrees, M., Zulhaidi, H., & Shafri, M. (2015). Maximizing Urban Features Extraction from Multi-sensor Data with Dempster-Shafer Theory and HSI Data Fusion Techniques. In *Asian Journal of Applied Sciences*. https://www.researchgate.net/publication/275040765

Kingma, D. P., & Ba, J. (2014). *Adam: A Method for Stochastic Optimization*. http://arxiv.org/abs/1412.6980

Lyu, Y., Vosselman, G., Xia, G., Yilmaz, A., & Yang, M. Y. (2018). *UAVid: A Semantic Segmentation Dataset for UAV Imagery*. http://arxiv.org/abs/1810.10438

Zhao, J., Wang, Y., Cao, Y., Guo, M., Huang, X., Zhang, R., Dou, X., Niu, X., Cui, Y., & Wang, J. (2021). The fusion strategy of 2d and 3d information based on deep learning: A review. In *Remote Sensing* (Vol. 13, Issue 20). MDPI. https://doi.org/10.3390/rs13204029

Zhong, H., Wang, H., Wu, Z., Zhang, C., Zheng, Y., & Tang, T. (2021). A survey of LiDAR and camera fusion enhancement. *Procedia Computer Science*, *183*, 579–588. https://doi.org/10.1016/j.procs.2021.02.100

Zhu, X. X., Tuia, D., Mou, L., Xia, G.-S., Zhang, L., Xu, F., & Fraundorfer, F. (2017). *Deep learning in remote sensing: a review*. https://doi.org/10.1109/MGRS.2017.2762307