High-Quality Road Detection Using U-Net-Based Semantic Segmentation with High-Resolution Orthophotos and DSM Data in Urban Environments

Mohamed Fawzy[1,2], Attila Juhász[1], Arpad Barsi[1]

[1] Department of Photogrammetry and Geoinformatics, Faculty of Civil Engineering, Budapest University of Technology and Economics, Műegyetem rkp. 3, H-1111 Budapest, Hungary,
{mohamed.fawzy, juhasz.attila, barsi.arpad}@emk.bme.hu
[2] Civil Engineering Department, Faculty of Engineering, South Valley University, 83523 Qena, Egypt,
mohamedfawzy@eng.svu.edu.eg

Keywords: Road Detection, U-net, Orthophotos, DSM, Urban Environments.

Abstract

Road detection and recognition from high-resolution geospatial data in urban environments is critical for numerous applications, including urban planning, navigation systems, and automated driving technologies. This study explores the potential of deep learning methodologies, specifically U-Net-based semantic segmentation, for high-quality road detection using open-access datasets. The input data consists of high-resolution digital orthophotos of a city region and corresponding Digital Surface Models (DSMs), allowing for a comprehensive analysis of two scenarios: (A) semantic segmentation using only imagery data and (B) segmentation utilizing both imagery and DSM data. Building on prior works by the authors, which include digital surface modelling and satellite image classification using U-Net and other neural network architectures, this research applies state-of-the-art techniques to leverage the spatial richness of orthophotos and the vertical information embedded in DSMs. Preliminary results indicate a significant improvement in road detection accuracy when integrating DSM data, highlighting the synergistic value of multi-source data in geospatial analysis. To validate the segmentation outputs, independent satellite imagery data are employed as a benchmark, enabling quantitative assessments of positional accuracy. The integrated orthophoto-DSM strategy achieved completeness, correctness, quality, and overall accuracy of 99.76%, 77.35%, 77.21%, and 95.95%, respectively, surpassing the solely orthophoto-based model, which achieved 73.76%, 51.10%, 42.52%, and 86.29% in these metrics. Experiments ensure that the proposed methodology achieves a robust delineation of road networks in urban environments, with scenario B outperforming scenario A. This research contributes to the growing body of literature on deep learning applications in photogrammetry and remote sensing, aligning with related studies on the semantic segmentation of urban features. The results underscore the importance of multimodal data fusion in geospatial analysis and its implications for enhancing road detection frameworks. Extracted road maps provide an accurate baseline for urban development planning and transportation management applications.

1. Introduction

Road networks are fundamental to urban infrastructure, facilitating transportation, connectivity, and economic development. Accurate and up-to-date road mapping is essential for various applications, including urban planning, traffic management, and disaster response. Consequently, a wide spectrum of road extraction algorithms has been developed. However, automated road extraction remains challenging, particularly in complex urban environments where roads exhibit varying widths, textures, and occlusions caused by trees, buildings, and other structures. To solve the complex and challenging tasks, two major categories, traditional and deep learning-based algorithms, are distinguished. The efficacy of both methods in producing satisfactory results has been well-documented. Traditional methods for road detection, such as manual digitization or classical image processing techniques, are time-consuming and prone to inconsistencies, necessitating the development of more efficient and robust approaches (Fawzy, Szabó et al. 2023). The requirement for manual interactions in traditional methods, such as the design of road feature extraction algorithms and the adjustment of threshold parameters, renders this solution less suitable for large-scale and complex remote sensing data. Conversely, numerous machine learning algorithms have been developed to address these challenges, facilitating the extraction of road networks and other urban spatial features from remote sensing images and digital surface models. In recent years, deep learning models, particularly Convolutional Neural Networks (CNNs), have demonstrated remarkable success in image classification and feature extraction. Among these, U-Net, a fully convolutional neural network initially designed for biomedical image segmentation (Ronneberger, Fischer et al. 2015), has gained significant attention for its ability to perform precise pixel-wise classification while preserving spatial information through skip connections. The effectiveness of U-Net in semantic segmentation tasks, including urban feature extraction, makes it a suitable candidate for automated road detection in high-resolution imagery (Fawzy and Barsi 2024). While high-resolution RGBI orthophotos provide valuable spectral information, road classification solely based on image data faces limitations due to spectral similarities between roads and other urban surfaces, such as rooftops and parking lots. To address these challenges, Digital Surface Models (DSMs), which provide elevation data, can be integrated into the classification process to enhance feature differentiation (Jahan, Zhou et al. 2018). Including DSM data allows for a more reliable distinction between roads and other urban

elements, improving overall classification accuracy (Al-Najjar, Kalantar et al. 2019).

Despite advancements in deep learning-based road detection, challenges persist in extracting road networks from high-resolution RGBI orthoimages. The state-of-the-art methods have attempted to mitigate these challenges by incorporating CNNs and additional supplementary data, such as DSMs and point cloud data, to enhance classification performance (Ayala, Sesma et al. 2021, Fawzy, Dowajy et al. 2024). However, further research is needed to optimize the integtared models for better generalization, reduced false detections, and improved classification accuracy in highly heterogeneous urban landscapes (Neupane, Horanont et al. 2021). The applications of CNNs have significantly advanced the field of remote sensing and urban feature extraction. Traditional methods for urban classification, such as pixel-based and object-based image analysis, often require extensive manual input and face challenges in distinguishing spectrally similar classes (Fawzy, Szabó et al. 2023). The emergence of CNNs has enabled automated feature learning, making them highly suitable for processing RGBI orthoimages and DSMs (Ma, Liu et al. 2019). The state-of-the-art offers comprehensive demonstrations of neural network applications to extract different aspects of the urban environment from RGBI orthophotos and digital surface models.

Various approaches have attempted to mitigate issues by incorporating DSM data improving differentiation between roads and surrounding features (Gehrke, Morin et al. 2010). (Huang, Zhao et al. 2018) used CNNs for urban land-use mapping, showcasing their ability to extract complex urban structures from satellite imagery. (Shukla and Jain 2020) investigated automated urban land information extraction from UAV data, revealing that DSM integration reduces false classifications and improves feature detection accuracy. (Osco, Junior et al. 2021) present a comprehensive review of deep learning in UAV remote sensing, detailing the classification and regression techniques used in recent image-based applications and grouping these into the environmental, urban and agricultural contexts. (Liu, Yang et al. 2019) proposed a trainable chain fully convolutional neural network (CFCN), which fuses high spatial resolution UAV images and the digital surface model for building extraction. An improved U-net was used for the coarse extraction of the buildings based on multilevel features. (Lin and Wang 2021) investigated a Robust U-Net (RU-Net) model that integrates multiple CNN architectures for material classification in urban environments. The efficient usage of the normalized difference vegetation index and DSM images is declared, and it is proved that near infrared and DSM provide more information on material features, reducing the likelihood of misclassification caused by similar features in RGB images. (Shao, Xiao et al. 2022) introduced a channel attention mechanism and a spatial attention mechanism with the objective of enhancing the use of spectral and spatial information based on the U-Net framework. This was done to address the problem of automatic extraction of road networks from a large number of remote sensing images. A novel network architecture based on U-Net and Atrous Spatial Pyramid Pooling (ASPP) is introduced by (Wulamu, Shi et al. 2019) to extract road features in different scales through different convolutional layers.

The primary goal of this recent work is to develop an accurate and reliable strategy for detecting roads in urban areas using high-resolution RGBI orthophotos and DSMs. To achieve the study objective, a U-Net-based semantic segmentation approach is applied, which effectively identifies road features by leveraging its encoder-decoder structure. Additionally, the research incorporates the DSM data into the classification process, enhancing the ability to distinguish roads from other urban elements such as buildings and bare soil. By integrating spectral and elevation information, the presented approach aims to improve classification accuracy and reduce the road detection errors. By incorporating elevation data into the deep learning framework, the study seeks to improve classification reliability and reduce misclassification problems, thereby advancing the state-of-the-art automated road extraction from remote sensing data.

2. Methodology

Given the research objectives, the presented methodology (Figure 1) begins with preprocessing and combining the input DSM and RGBI orthophoto. The merged data is used to generate a training dataset containing the appropriate classes and labels. Leveraging the collected training dataset, a U-Net model is trained, tested, and validated for image classification in urban environments. Following the pre-trained model development, the full scene is classified into multiple land cover classes using the RGBI photo and the combined DSM and RGBI data. Road features are detected and assessed to evaluate the RGBI orthophoto with DSM and solely orthophoto for road extraction.

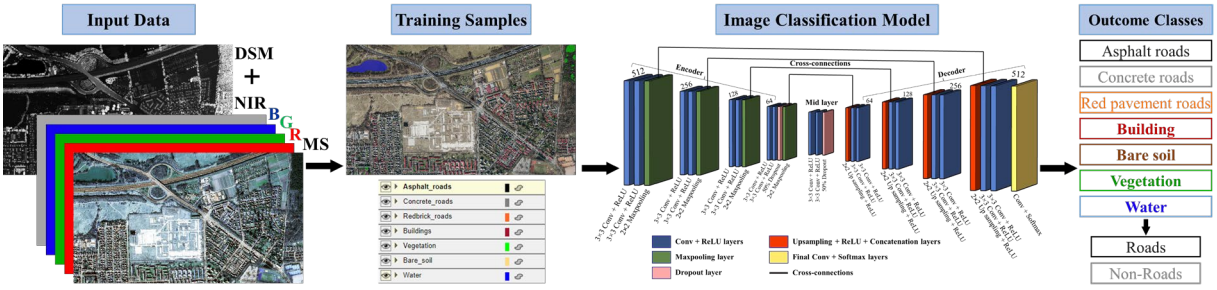

Figure 1. Procedures of the presented methodology.

During the classification stage, two strategies were applied for the road extraction process. Firstly, one class is assigned to all road samples which led to misclassification between some road patches and other urban classes. Secondly, three main types of road surfaces were considered: asphal, concrete, and red pavement roads. The red pavement can be made of red material or coated with red additions using various techniques. For example, asphalt is coloured with pigments such as red iron oxide or naturally reddish aggregates which is often used for cycle

paths and footpaths. Also, concrete is mixed with red substances or contain naturally reddish crushed stone such as red andesite or basalt. Red paving stones, clinker bricks, that are made from natural clay and fired to produce a red colour, and eosin-based and resin-based coated materials are commonly used for paved roads, sidewalks and sport courts or particular road surfaces where a durable, non-slip finish is required. Due of the significant differences in spectral characteristics, the three types of roads were treated as distinct classes during data training and classification, then merged as a combined road layer for accuracy assessment.

3. Experimental Works

3.1. Study Area and Data Used

According to the trends of the last years, more and more free accessible spatial data sources can be found worldwide, which support GIS-based spatial data analyses. In our research, freely available German data were utilized. Both the orthophotos and the DSM were shared by the State Office for Geoinformation and Land Surveying of Lower Saxony, together with other typical and valuable spatial data sources such as digital terrain model, 3D building models (LoD1 and LoD2), house data, land use, topographic, cadastral and overview maps of the German state (LGLN 2025). The selected study area is a part of the area called Bothfeld, located in the north-east of Hannover (52°25'29.34"N, 9°47'8.46"E). Roads of different grades, primary, secondary and side roads are noticed to be essential for a comprehensive study (Figure 2).

Figure 2. The location of the study area, Hanover.

The detailed metadata of the downloaded spatial data sources are as follows: Orthophoto, Update tile: 2022.07.01, Camera sensor: Ultracam Eagle, Spectral channels: RGBI, Colour depth: 8, EPSG: 25832, Ground pixel size: 20 cm. Digital surface model, Update tile: 2024.01.30, Sensor: ALS (4 points/m2), EPSG: 25832, Resolution: 1 m.

3.2. Reference Data

Objective data is produced as a benchmark for evaluating model outcomes, allowing for quantitative assessments of positional accuracy. Manually digitised polygons of asphalt, concrete, and red pavement roads, were extracted using an independent satellite imagery data as a reference to validate the classification results (Figure 3). The extracted polygons include not only road surfaces, but also car parks, significant squares and courtyards which are covered by road materials.

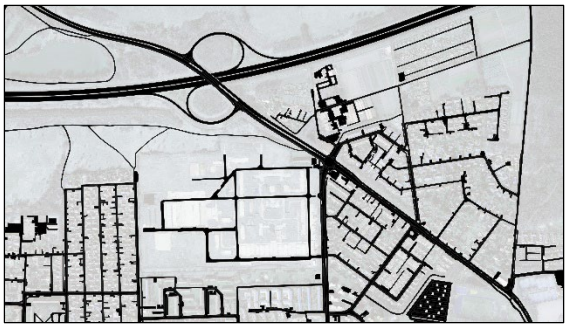

Figure 3. Manually extracted reference roads.

3.3. Sample Selection and Data Labelling

The image classification performance improves with more useful training data (Xu, Su et al. 2021). MATLAB Image Labeler offers a simple window for interactively labelling data for instance semantic segmentation networks (MathWorks 2020). Training samples and their related classes, for the training, testing, and validation phases, were prepared by manually labelling regions of interest ensuring a representative distribution of all classes.

3.4. U-Net for Image Classification

3.4.1. U-Net Model Architecture

The U-Net model, employed in this study, follows the standard encoder-decoder architecture with skip connections to facilitate feature preservation. The encoder consists of four convolutional blocks, each containing two 3×3 convolutional layers followed by Rectified Linear Unit (ReLU) activations and a 2×2 max-pooling layer. A dropout layer (50% probability) is incorporated before the max-pooling operation in the final encoder block to prevent overfitting. The mid layer positioned at the center of the U-Net, comprises two 3×3 convolutional layers with ReLU activations and an additional dropout layer to enhance generalization. The decoder mirrors the encoder but replaces max-pooling with 2×2 upsampling layers to restore spatial resolution. Each decoder block includes two 3×3 convolutional layers with ReLU activations to refine the extracted features. The final output is produced using a 1×1 convolutional layer with a softmax activation function, which assigns class probabilities to each pixel in the input image. Connection links between corresponding encoder and decoder blocks allow the network to retain spatial information and enhance segmentation accuracy. The model processes an input composed of five channels (red, green, blue, near infrared from the orthoimage, for scenario A, in addition to one elevation band from the DSM for scenario B) and classifies each pixel into one of seven urban land cover classes: asphalt road, concrete road, red pavement road, building, bare soil, vegetation and water.

3.4.2. U-Net Model Training

The training of the U-Net model was conducted using a supervised learning approach with labelled image patches based on the pre-defined samples and labelled data. A dataset of 16000 image patches (128×128 pixels) was used for training, while an additional 2000 patches were allocated for validation. The initial weights and biases for the convolutional layers were assigned using a randomly generated distribution. The model optimization utilized the

Stochastic Gradient Descent with Momentum (SGDM) algorithm, which accelerates convergence by incorporating momentum in weight updates. The learning rate was set to 0.05, and a momentum factor of 0.9 was applied to stabilize training. Batch normalization was performed after each convolutional layer to mitigate internal covariate shifts and improve gradient flow. The network was trained using a mini-batch size of 16 and maximum epochs of 100, which allowed efficient processing while maintaining a stable optimization process. The cross-entropy loss function was employed to quantify the difference between predicted and ground truth labels, guiding parameter updates throughout the training iterations. To monitor training progress, the loss and accuracy curves were continuously evaluated for both the training and validation sets. The final models achieved a validation accuracy of 98.16% for the model of the RGBI orthophoto (scenario A), and 99.15 for the model of RGBI orthophotos integtated with the DSM (scenario A), demonstrating strong learning capability and robustness in semantic segmentation tasks.

4. Results and Discussions

4.1. Image Classification procedures

Classification outcomes of two pre-trained U-Net models are produced using RGBI orthophoto alone (Figure 4), and RGBI orthophoto combined with DSM (Fig. 5). The entire scenes are classified into three types of roads, building, bare soil, vegetation, and water classes.

Figure 4. Classified image using only the orthophoto.

Figure 5. Classified image using orthophoto and DSM.

The classified image obtained solely from high-resolution RGBI orthophotos exhibits a clear distinction between vegetation (green), water bodies (blue), and bare soil (brown). The classification of road surfaces, however, presents several challenges. Asphalt (black) and concrete (grey) roads are often confused with other artificial surfaces, such as rooftops and parking lots, leading to misclassifications. The red pavement roads (orange) are relatively better delineated due to their distinct spectral signature, but they sometimes blend with bare soil areas, causing ambiguity. Additionally, occlusions caused by vegetation overhanging roads introduce errors, particularly in narrow street segments where shadows obscure details. Despite these limitations, the classification provides a reasonable segmentation of urban features with notable errors in distinguishing roads from other impervious surfaces.

The integration of DSM data substantially improves the classification results, particularly in differentiating road surfaces from other built-up elements. The elevation information helps to resolve the spectral confusion between asphalt roads and dark-coloured rooftops, reducing misclassification. Similarly, concrete roads are more accurately distinguished from lighter rooftops, enhancing segmentation precision. Red pavement roads benefit from the elevation-based differentiation, although some confusion remains in areas where bare soil shares similar spectral properties. Vegetation and water are robustly classified with minimal errors. The DSM-enhanced classification refines road extraction by reducing misclassifications and improving the continuity of the detected road network, particularly in complex urban environments where spectral similarities previously hindered accuracy.

4.2. Road Extraction Outcomes

The road classes, asphalt, concrete and red pavement roads, are merged as one road class, while the other classes are combined as non-road class in a binary image for accuracy assessment (Figure 6, 7).

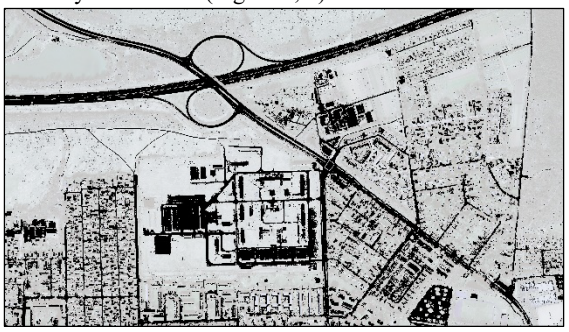

Figure 6. Extracted roads using only orthophoto.

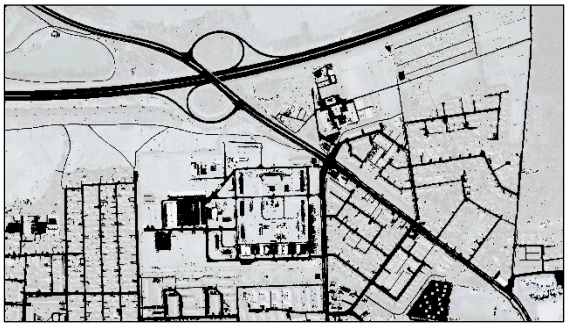

Figure 7. Extracted roads using orthophoto and DSM.

The road class extracted from the orthophoto classification exhibits several discrepancies when compared to the reference dataset (Fig. 3). The classification effectively captures some major roads but struggles with spectral confusion, particularly between asphalt roads and dark

rooftops, as well as between concrete roads and similar artificial surfaces. Red pavement roads are better distinguished, but there is notable confusion with bare soil areas, leading to over-segmentation in certain regions. Narrow roads are often fragmented due to spectral blending with adjacent features or occlusion by vegetation. Additionally, road sections in shadowed areas are frequently misclassified, causing gaps in the extracted network. The extracted road network, therefore, lacks completeness, as some segments are missing, and includes false detections where non-road surfaces are mistakenly classified as roads.

The inclusion of DSM data significantly refines the road extraction results by leveraging elevation differences to enhance classification accuracy. Asphalt roads and concrete roads are more distinctly separated from rooftops and parking areas, reducing spectral confusion. Red pavement roads benefit a lot from improved feature differentiation, although some ambiguity with bare soil persists in areas where elevation differences are minimal. Vegetation remains a primary source of occlusion, particularly along tree-lined roads, but the overall connectivity of the road network is notably improved compared to the orthophoto-only approach (Fig. 6). The extracted roads exhibit higher accuracy measures, with fewer interruptions and misclassifications. However, bridges and overpasses introduce some challenges, as elevation variations within the road network can lead to minor classification inconsistencies. Despite this, the DSM-enhanced extraction aligns more closely with the reference dataset, demonstrating the effectiveness of integrating elevation data for accurate road mapping.

4.3. Accuracy Assessment

Road detection is evaluated focusing on the extracted roads pixels comparing to the reference data. Four metrics, completeness, correctness, quality, and overall accuracy are applied, in conjunction with a confusion matrix, to assess the extracted road accuracy (Eq. 1-4) (Fig. 7). True Positives (TP) represents the number of road pixels correctly classified, True negative (TN) indicates the number of non-road pixels successfully detected, False Positives (FP) reflects the misclassified road pixels, and False Negatives (FN) relates to the incorrectly classified non-road pixels (Heipke, Mayer et al. 1997).

$$\text{Completeness} = TP/(TP+FN) \quad (1)$$

$$\text{Correctness} = TP/(TP+FP) \quad (2)$$

$$\text{Quality} = TP/(TP+FP+FN) \quad (3)$$

$$\text{Overall accuracy} = (TP+TN)/(TP+FP+FN+TN) \quad (4)$$

The confusion matrix for the DSM combined with RGBI orthophoto classification (Table 1) indicates a highly favourable performance. On the other hand, the confusion matrix for the RGBI orthophoto-only classification (Table 2) reveals several limitations.

Confusion matrix for orthophoto and DSM		Prediction	
		Road	Non-Road
Reference	Road	7,573,534	18,313
	Non-Road	2,217,272	45,422,673

Table 1. Confusion matrix of the extracted roads with orthophoto and DSM.

Confusion matrix for orthophoto		Prediction	
		Road	Non-Road
Reference	Road	5,599,614	1,992,233
	Non-Road	5,577,938	42,062,007

Table 2. Confusion matrix of the extracted roads with only orthophoto.

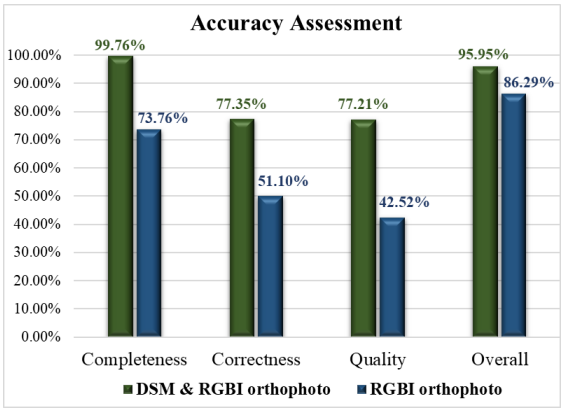

Figure 7. Accuracy metrics of the extracted roads with the DSM & RGBI orthophoto and solely RGBI orthophoto.

Confusion matrix for orthophoto and DSM demonstrates that nearly all road features from the reference dataset are successfully identified, as evidenced by an exceptionally low number of missed road cases (FN) compared to correctly identified ones (TP). Although some non-road areas are misclassified as roads (FP), the overall balance of correctly identified non-road areas (TN) is sufficient to yield very high overall accuracy. Regarding the standard accuracy measures, the classification shows excellent completeness (99.76%), suggesting that nearly every actual road pixel is detected, implying that most of the predicted road pixels are indeed true roads. These factors, when combined, result in a high-quality measure, underscoring the effectiveness of incorporating DSM data with the RGBI orthophoto. The DSM's contribution of elevation and structural details appears to be pivotal in reducing ambiguity, thereby enhancing the separation of road and non-road classes.

Meanwhile, in the confusion matrix for the orthophoto, the number of undetected road features (reflected in a higher rate of FN) is notably larger, which results in a lower completeness rate (73.76%). Additionally, the matrix indicates a greater number of non-road areas being erroneously classified as roads, thereby reducing the correctness. The reliance solely on spectral information from the RGBI imagery leads to increased misclassification, particularly in scenes where roads share similar spectral characteristics with adjacent non-road features. This results in a decrease in both the quality measure and the overall accuracy, pointing to the inherent challenges of distinguishing road features without the supplementary information provided by DSM data.

Moreover, the direct comparison of the two confusion matrices highlights several key differences:

- **Completeness**

The DSM with RGBI approach achieves near-perfect detection of road features, whereas the RGBI-only method misses a significant portion of these features. This indicates that the structural information from the DSM is crucial for capturing subtle or ambiguous road characteristics that might otherwise be overlooked.

- **Correctness**

The integrated method exhibits a higher correctness (77.35%), meaning it is more likely to be correct when a road is predicted. In contrast, the RGBI-only classification with 51.10% suffers from a higher rate of false alarms (non-road pixels incorrectly labelled as road), demonstrating that spectral data alone can lead to confusion in class differentiation.

- **Quality and Overall Accuracy**

The quality measure, reflecting the balance between completeness and correctness, is markedly superior in the DSM and RGBI case (77.21%) against 42.52% for the orthophoto. This directly translates into higher overall accuracy (95.95%) versus 86.29% without the DSM, affirming the benefit of combining DSM data with spectral information in complex classification tasks.

5. Conclusions and Future Work

This study presented a deep learning-based methodology for high-quality road detection using U-Net semantic segmentation, leveraging high-resolution RGBI orthophotos and Digital Surface Model (DSM) data. The findings highlight the substantial improvement in classification accuracy when integrating DSM with spectral information, demonstrating the effectiveness of multimodal data fusion in road detection applications. The results indicate that the inclusion of DSM significantly enhances road classification performance by mitigating spectral confusion, reducing false positives, and improving overall detection completeness and correctness. Specifically, the integrated orthophoto-DSM strategy achieved a completeness measure of 99.76%, correctness of 77.35%, quality of 77.21%, and overall accuracy of 95.95%, substantially outperforming the solely orthophoto-based model alone , that obtained 73.76%, 51.10%, 42.52%, and 86.29% in the same metrics.

A comparative analysis of classification results revealed that roads extracted merely from RGBI orthophotos exhibited significant misclassification due to spectral similarities between road surfaces and other urban elements such as rooftops and parking areas. In contrast, the DSM-integrated approach effectively differentiated road features based on elevation variations, leading to a more accurate and continuous delineation of the road network. The current approach effectively distinguished different road categories; as a result, the classification accuracy for road types with similar spectral and elevation characteristics have been improved. These findings underscore the importance of incorporating elevation data in deep learning frameworks for geospatial applications, particularly in complex urban environments where traditional image-based segmentation methods face limitations. Despite the promising results, certain challenges remain including roadside trees that interfere with the road cover, occlusions and elevation inconsistencies such as those at bridges and overpasses, the similarities between roads and other urban classes, and the impact of varying weather conditions that introduce classification errors in need of further refinement.

Future research will focus on enhancing model robustness and generalization to overcome the challenges and limitations of the U-Net preformance for road detection. Additionally, integrating advanced deep learning techniques, such as attention mechanisms or transformer-based architectures, could refine feature discrimination and classification efficiency, as well as improving the model scalability to other urban environments with different characteristics. Moreover, incorporating various geospatial datasets, such as LiDAR-derived point clouds or multi-temporal imagery is recommended to improve road extraction accuracy across varying urban landscapes. As well, automating post-processing steps for road network regularization and connectivity enhancement would be beneficial in practical applications, particularly for urban planning, navigation systems, and autonomous vehicle technologies.

This study contributes to the growing body of research on deep learning applications in remote sensing and photogrammetry by demonstrating the efficacy of U-Net-based semantic segmentation for road detection. The findings emphasize the potential of multimodal data fusion in geospatial analysis and its implications for improving urban infrastructure mapping. The proposed approach serves as a foundation for future advancements in automated road extraction, paving the way for more accurate and efficient geospatial solutions.

Acknowledgements

The research reported in this paper is part of project no. BME-NVA-02, implemented with the support provided by the Ministry of Innovation and Technology of Hungary from the National Research, Development and Innovation Fund, financed under the TKP2021 funding scheme.

References

Al-Najjar, H. A., B. Kalantar, B. Pradhan, V. Saeidi, A. A. Halin, N. Ueda and S. Mansor, 2019: Land cover classification from fused DSM and UAV images using convolutional neural networks. Remote Sensing 11(12): 1461.

Ayala, C., R. Sesma, C. Aranda and M. Galar, 2021: A deep learning approach to an enhanced building footprint and road detection in high-resolution satellite imagery. Remote Sensing 13(16): 3135.

Fawzy, M. and A. Barsi, 2024: A U-Net Model for Urban Land Cover Classification Using VHR Satellite Images. Periodica Polytechnica Civil Engineering.

Fawzy, M., M. Dowajy, T. Lovas and A. Barsi, 2024: Urban Land Cover Classification Using Deep Neural Networks Based on VHR Multi-Spectral Image and Point Cloud Integration. IGARSS 2024-2024 IEEE International Geoscience and Remote Sensing Symposium, IEEE.

Fawzy, M., G. Szabó and A. Barsi, 2023, A Shallow Neural Network Model for Urban Land Cover

Classification Using VHR Satellite Image Features. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences 10: 57-64.

Gehrke, S., K. Morin, M. Downey, N. Boehrer and T. Fuchs, 2010: Semi-global matching: An alternative to LIDAR for DSM generation. Proceedings of the 2010 Canadian Geomatics Conference and Symposium of Commission I.

Heipke, C., H. Mayer, C. Wiedemann and O. Jamet, 1997: Evaluation of automatic road extraction. International Archives of Photogrammetry and Remote Sensing 32(3 SECT 4W2): 151-160.

Huang, B., B. Zhao and Y. Song, 2018: Urban land-use mapping using a deep convolutional neural network with high spatial resolution multispectral remote sensing imagery. Remote Sensing of Environment 214: 73-86.

Jahan, F., J. Zhou, M. Awrangjeb and Y. Gao, 2018: Fusion of hyperspectral and LiDAR data using discriminant correlation analysis for land cover classification. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 11(10): 3905-3917.

LGLN, 2025: State Office for Geoinformation and Land Surveying of Lower Saxony. Retrieved 8-Febraury, 2025, from https://www.lgln.niedersachsen.de/startseite/vertrieb_support/geodaten_marktplatz/opengeodata/opengeodata-220509.html.

Lin, C.-H. and T.-Y. Wang, 2021: A novel convolutional neural network architecture of multispectral remote sensing images for automatic material classification. Signal Processing: Image Communication 97: 116329.

Liu, W., M. Yang, M. Xie, Z. Guo, E. Li, L. Zhang, T. Pei and D. Wang, 2019: Accurate building extraction from fused DSM and UAV images using a chain fully convolutional neural network. Remote sensing 11(24): 2912.

Ma, L., Y. Liu, X. Zhang, Y. Ye, G. Yin and B. A. Johnson, 2019: Deep learning in remote sensing applications: A meta-analysis and review. ISPRS journal of photogrammetry and remote sensing 152: 166-177.

MathWorks, 2020: Get Started with the Image Labeler. Retrieved 10 January, 2025, from https://nl.mathworks.com/help/vision/ug/get-started-with-the-image-labeler.html.

Neupane, B., T. Horanont and J. Aryal, 2021: Deep learning-based semantic segmentation of urban features in satellite images: A review and meta-analysis. Remote Sensing 13(4): 808.

Osco, L. P., J. M. Junior, A. P. M. Ramos, L. A. de Castro Jorge, S. N. Fatholahi, J. de Andrade Silva, E. T. Matsubara, H. Pistori, W. N. Gonçalves and J. Li, 2021: A review on deep learning in UAV remote sensing. International Journal of Applied Earth Observation and Geoinformation 102: 102456.

Ronneberger, O., P. Fischer and T. Brox, 2015: U-net: Convolutional networks for biomedical image segmentation. Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18, Springer.

Shao, S., L. Xiao, L. Lin, C. Ren and J. Tian, 2022: Road extraction convolutional neural network with embedded attention mechanism for remote sensing imagery. Remote Sensing 14(9): 2061.

Shukla, A. and K. Jain, 2020: Automatic extraction of urban land information from unmanned aerial vehicle (UAV) data. Earth Science Informatics 13(4): 1225-1236.

Wulamu, A., Z. Shi, D. Zhang and Z. He, 2019: Multiscale road extraction in remote sensing images. Computational intelligence and neuroscience 2019(1): 2373798.

Xu, Z., C. Su and X. Zhang, 2021: A semantic segmentation method with category boundary for Land Use and Land Cover (LULC) mapping of Very-High Resolution (VHR) remote sensing image. International Journal of Remote Sensing 42(8): 3146-3165.