






Road Network Mapping from Multispectral Satellite Imagery: Leveraging Deep Learning and Spectral Bands

Samuel Hollendonner ¹, Negar Alinaghi ¹, and Ioannis Giannopoulos ¹

¹Research Division Geoinformation, TU Wien, Vienna, Austria

Correspondence: Samuel Hollendonner (samuel.hollendonner@geo.tuwien.ac.at)

Abstract.

Updating road networks in rapidly changing urban landscapes is an important but difficult task, often challenged by the complexity and errors of manual mapping processes. Traditional methods that primarily use RGB satellite imagery struggle with obstacles in the environment and varying road structures, leading to limitations in global data processing. This paper presents an innovative approach that utilizes deep learning and multispectral satellite imagery to improve road network extraction and mapping. By exploring U-Net models with DenseNet backbones and integrating different spectral bands we apply semantic segmentation and extensive post-processing techniques to create georeferenced road networks. We trained two identical models to evaluate the impact of using images created from specially selected multispectral bands rather than conventional RGB images. Our experiments demonstrate the positive impact of using multispectral bands, by improving the results of the metrics Intersection over Union (IoU) by 6.5%, F1 by 5.4%, and the newly proposed relative graph edit distance (relGED) and topology metrics by 2.2% and 2.6% respectively.

Keywords. Road Network Extraction, Semantic Segmentation, Multispectral Imagery, Remote Sensing

1 Introduction

In the rapidly evolving field of urban and environmental dynamics, the task of updating and creating accurate road network maps is both critical and challenging. The traditional process of creating and updating such maps is a complex and error-prone task that often fails to keep up with rapidly changing urban landscapes. This labor-intensive manual work is not only time-consuming, but is also prone to inaccuracies (Kent, 2010), especially in the face of the development of new infrastructure, the maintenance of existing structures, and changes caused by natural disasters (O’Callaghan et al., 2020). Our research ad-

resses these challenges by using high-resolution multispectral satellite imagery in combination with the latest advances in deep learning, particularly semantic segmentation, to automate and refine road network extraction.

Previously, attempts have been made to extract road networks from satellite images, mainly using RGB bands (Zhu et al., 2021; Lu et al., 2021b, a; Batra et al., 2019; Mei et al., 2021; Ghandorh et al., 2022). These methods included traditional computer vision-based algorithms, which often required extensive user input and reached their limits when processing global datasets (Bajcsy and Tavakoli, 1976; Ünsalan and Sirmacek, 2011; Movaghati et al., 2010; Jin and Davis, 2005; Zhang et al., 2018). Occlusions caused by objects such as trees and buildings, and different road structures and widths posed an additional challenge (Mingjun and Daniel, 2004; Cheng et al., 2014; Mnih and Hinton, 2010). To solve these problems, machine learning techniques such as Support Vector Machines (SVM) and Restricted Boltzmann Machines were introduced, which enable the analysis of larger data sets but still have significant limitations.

The advent of deep learning has introduced a significant change in this area. Models using convolutional neural networks (CNNs) have been widely adapted (Máttyus et al., 2017; Bastani et al., 2018), with architectures such as U-Net (Ronneberger et al., 2015) and DenseNet (Huang et al., 2017) showing promising results on semantic segmentation tasks (Eerapu et al., 2019; Mohanty et al., 2020; Xu et al., 2018; Xin et al., 2019; Ghandorh et al., 2022; Mei et al., 2021). These deep learning methods have better managed the complexity of road network structures and varying environmental conditions.

In this work, we evaluate different deep learning architectures, such as U-Net models with DenseNet backbones for the task of creating road network maps from satellite imagery created by WorldView-3 and provided by SpaceNet. We also investigate the use of different multispectral bands beyond the RGB spectrum to utilize their unique spectral properties for more detailed and accurate road network mapping. To evaluate the extracted networks the graph

edit distance (GED) is adapted into a relative metric, better suited to differently sized graphs (Abu-Aisheh et al., 2015). In addition, we introduce a novel evaluation metric that analyses the topological correctness of a network, to allow for a more comprehensive evaluation.

In summary, our research not only pushes the boundaries of traditional RGB-based mapping techniques, but also introduces novel assessment metrics. Using multispectral satellite imagery and state-of-the-art deep learning, we open new avenues for efficient, automated mapping of road networks, which is essential for modern urban planning and environmental monitoring.

2 Related work

In this section, we present several road extraction techniques, employing various algorithms from the fields of computer vision, machine learning, and deep learning. In addition, we focus on techniques that use multispectral imagery in similar remote sensing applications and elaborate on different approaches in which multispectral bands have been integrated into the processing workflow.

2.1 Computer-Vision-based Extraction Methods

Road network extraction has been a widely researched topic. One of the earliest methods was presented in the work of Bajcsy and Tavakoli (1976) who established specific physical and geometrical requirements that a road must meet for classification. This method used a sequence of low-level operators to detect and extract roads, which were applied to images from the ERTS-1 satellite and compared with a hand-crafted dataset. Another approach by Ünsalan and Sirmacek (2011) included image processing and thresholding to identify the centerline of the road using manually selected pixel values of the road surface as a reference. This method involved additional steps to eliminate false positive road segments and was tested with images from various remote sensing sources, including satellites and aerial platforms. Furthermore, a method proposed by Movaghati et al. (2010), which combines an Extended Kalman filter with a Particle Filter, was used to track the centerline of the roads starting from a manually defined point. Most computer vision-based methods contain a manual component in their workflow that inherently limits their scalability and applicability for extracting large or automated road networks.

2.2 Machine-Learning-based Extraction Methods

In recent years, there has been a clear trend towards the application of machine learning-driven methods for road extraction, moving away from traditional computer-vision-based approaches. These more advanced algorithms enable the processing of large datasets minimizing the need for manual adjustment of parameters.

An approach proposed by Mingjun and Daniel (2004) uses a Support Vector Machine (SVM) for a binary classification followed by a Weighted Region-Growing algorithm to refine the labeling of road pixels, leveraging both spectral and spatial road information. This method applied to an image from the Ikonos satellite, proved to be more effective than the Gaussian maximum likelihood classifier. Similarly, Cheng et al. (2014) employed an SVM followed by a graph cut-based probability propagation algorithm to filter out pixel clusters that do not represent road surfaces. When tested on aerial images of Toronto, this method performed better than other techniques that relied on K-means clustering and morphological operations. However, challenges such as occlusion and disconnected road segments were recognized as persistent problems in both studies.

2.3 Semantic-Segmentation-based Extraction Methods

Semantic segmentation is a computer vision task that assigns semantic labels to every pixel of an image and is particularly suited for extracting patterns from remote sensing images (Thoma, 2016). This process, often involving deep learning models like fully CNNs, requires extensive datasets with labeled images. Many approaches use semantic segmentation, as the generalization capabilities of trained models offer more efficient, reliable, and accurate road extraction results. Recent research aims to improve these model architectures and has led to the development of numerous new models.

Máttyus et al. (2017) proposed a method for extracting road networks in the form of graphs from aerial imagery using a CNN. This included semantic segmentation to identify road pixels, which were then simplified to centerlines and formed a graph-based representation of the road network. Building on this work, Bastani et al. (2018) improved their method by implementing a second CNN for iterative extraction of graph components, outperforming both semantic segmentation alone and the method of Máttyus et al. (2017).

Lu et al. (2021a)'s proposal for the Global-Local Adversarial Learning (GOAL) framework, based on ResNet-50 with an additional adversarial learning branch, showed improved generalization of model predictions. When applied to SpaceNet and DeepGlobe road datasets, this framework achieved excellent performance. In this context, Lu et al. (2021b) also introduced the GAMSNet architecture, which has a spatial awareness module to capture spatial context dependencies and a channel awareness module to consider the relationship between image channels. This model outperformed various LinkNet50 models on SpaceNet and DeepGlobe road datasets. As roads usually cover only a small part of the study area, the combination of attention mechanisms with edge detection methods was proposed by Ghandorh et al. (2022). This method improved the segmentation results and emphasized the importance of con-

sidering the characteristics of the dataset during model tuning.

The Connectivity Attention Network (CoANet) proposed by Mei et al. (2021), which is based on a pre-trained ResNet-101 model with additional modules (an Atrous Spatial Pyramid Pooling module and a connectivity attention module) to improve road detection, especially its connectivity, is a step forward. By using strip kernels instead of traditional square kernels, linear road features could be better captured, leading to better results on the SpaceNet and DeepGlobe datasets.

While these methods have achieved remarkable results through complex model architectures and elaborated processing techniques, and primarily use RGB images, they overlook the potential of utilizing more informative image bands available in their datasets. Rather than further complicating models with additional modules or pre-processing steps, exploring and effectively utilizing additional image bands, as suggested in other remote sensing applications (see subsection 2.4), could provide a simpler but effective way to improve performance.

2.4 Inclusion of Multispectral Imagery

Whilst the use of high-resolution multispectral data in remote sensing is well established (Eugenio et al., 2015; Espinoza et al., 2017; Warren and Metternicht, 2005), its integration into deep learning applications has been less explored. As deep learning capabilities grow, new methods are emerging to leverage the unique spectral properties of the diverse multispectral data.

Li et al. (2019) demonstrated the extraction of building footprints from WorldView-2 satellite multispectral imagery, using the eight available bands and integrating GIS map data as an additional channel. They chose a U-Net architecture for semantic segmentation, which required modifications to handle images with more than three channels. This approach showed improved results over others but required changes to the model.

Similarly, Alhassan et al. (2020) modified existing model architectures to include images with additional channels, allowing the incorporation of multispectral bands. Their goal was to produce land-use/land-cover maps, and they modified VGG, ResNet, and GoogLeNet architectures, adding an adversarial learning extension and a context module. These modifications not only improved map extraction accuracy but also significantly streamlined the process.

Taking a different approach, Yuan et al. (2021) combined RGB, Near Infrared (NIR), and Short-Wave Infrared (SWIR) channels as input to their MC-WBDN model, aimed at segmenting water bodies from Sentinel-2 imagery. Considering NIR and SWIR bands' proven usefulness in identifying water bodies in remote sensing, their inclusion enhanced the model's segmentation ability. The MC-WBDN model outperformed other models and

approaches, demonstrating the advantages of integrating multispectral bands in deep learning applications.

In line with these advancements, we plan to adopt a similar strategy of utilizing multiple channels from multispectral imagery to enhance semantic segmentation results for the extraction of road networks, while reducing the need for labor-intensive and complicated model adjustments.

3 Methodology

In this section, we outline the data utilized, explain our methodology for creating ground truth images from vector data, and detail the pre-processing steps. We also describe the selection of multispectral bands and the training process for a semantic segmentation model for road surface detection. Furthermore, we elaborate on post-processing techniques to improve segmented images and to extract a graph-based representation of the road network. We also present additional methods for refining graphs and describe how these graphs are converted into georeferenced GeoJSON formats.

3.1 Data and Software Availability

The SpaceNet dataset¹ contains georeferenced high-resolution multispectral satellite imagery created by WorldView-3 (Etten et al., 2018). These images, with different resolutions and spectral band combinations, show four cities: Las Vegas, Paris, Shanghai, and Khartoum, each with different types of road networks.

The dataset, with a total of 2780 images, is unevenly distributed across these cities: Las Vegas, with 989 images, predominantly shows a grid-shaped road network. Paris, which contributes 310 images, is an example of a radial (star-shaped) network that reflects the historical development of the city. Shanghai, which provides 1198 images, combines grid and organic patterns, reflecting the mix of new and old urban areas. Khartoum, with 283 images, shows a grid and organic pattern as well, influenced by its unique geographical location (Rodrigue, 2020). Each image has a size of 1300×1300 pixels which roughly corresponds to a footprint of 400×400 meters. The corresponding ground truth information is provided as GeoJSON files that contain road centerlines as line string objects in the WGS84 coordinate system. To capture the maximum detail possible, we used the pan-sharpened RGB and multispectral imagery with a resolution of 0.31 meters per pixel. The different spectral properties of the available multispectral bands are listed in Table 1 provided by DigitalGlobe (2017). This diverse collection of imagery, covering a range of road network types in different cities, provides a rich and varied dataset ideal for in-depth analysis and the application of deep learning techniques.

¹<https://spacenet.ai/spacenet-roads-dataset/>

The code developed for this paper, including a detailed workflow, is publicly available at ². It can be executed via a set of numbered scripts and is written entirely in Python.

| Band | Wavelength in nm |
|----------|------------------|
| Coastal | 397 – 454 |
| Blue | 445 – 517 |
| Green | 507 – 586 |
| Yellow | 580 – 629 |
| Red | 626 – 696 |
| Red Edge | 698 – 749 |
| Near-IR1 | 765 – 899 |
| Near-IR2 | 857 – 1039 |

Table 1. Spectral properties of WorldView-3 image bands.

3.2 Data Pre-Processing

To prepare the data for a semantic segmentation model, several pre-processing steps are required. First, we create ground truth images by converting GeoJSON vector road centerlines into dataframes and then applying a two-meter buffer to simulate road widths. While this assumption of uniform width does not always reflect variability in the real world, it is consistent with the results of our data inspection. The dataframes are then converted to binary arrays and then to single-channel images for model training. The original 16-bit remote sensing images were down-scaled to 8-bits to facilitate processing and reduce memory requirements.

To avoid memory issues when loading the images, the 1300×1300 pixel images are segmented into smaller 512×512 -pixel sub-images, each creating nine overlapping segments. These can later be merged to reconstruct complete images. The pre-processing uses and extends functions from the APLS Python library (Etten and Le, 2020).

3.3 Selection of Multispectral Bands

The eight-band multispectral images of the SpaceNet dataset, which go beyond RGB, provide a unique opportunity to utilize the spectral properties of road surfaces for improved segmentation. As described in subsection 2.4, several approaches have been proposed to include multispectral bands, which mainly require adapting the semantic segmentation model (Li et al., 2019). To avoid this complication and maintain the standard three-channel image structure, we chose a trio of multispectral bands that emphasize the spectral characteristics of road surfaces. These bands were merged to create a false-color image for model input.

Our band selection process involved evaluating the provided imagery and pertinent literature. Guided by

²<https://geoinfo.geo.tuwien.ac.at/resources/>

Shahi et al. (2015), we incorporated an infrared band (WorldView-3’s Near-IR2 from 857 nm to 1039 nm) due to its efficacy in depicting road surfaces (DigitalGlobe, 2017).

To determine the other two image bands, we conducted comparisons to maximize differences between bands, thus enhancing the information content in the composite three-band image. This analysis yielded various combinations, predominantly in the longer wavelengths (Green to Near-IR2), as urban surfaces typically reflect these wavelengths more (Jensen, 2009). Thus, we chose the Red edge band (698 nm to 749 nm) alongside the Near-IR2 band to improve road surface detection. The green image band (507 nm to 586 nm) was selected to complement other bands, improve vegetation detection, and help distinguish vegetation from roads, often located adjacently.

Due to computational constraints, we could only assess a semantic segmentation model trained on RGB images against one trained on a single non-RGB image band combination. Training additional models with diverse band combinations exceeded our resources. Figure 1 displays an example, showing an RGB image alongside the resulting multispectral image from our band selection, highlighting the improved differentiation between the roads and the surrounding area.

3.4 Semantic Segmentation Model

To investigate the effectiveness of semantic segmentation for the automatic extraction of road networks, different model architectures were tested. However, as U-Net’s potential is well studied in the literature (see subsection 2.3), and as its decoder-encoder structure is known to be essential for complex structures such as roads (Ronneberger et al., 2015), we focused on leveraging existing model backbones that are compatible with the U-Net structure. We tested backbones such as ResNet, VGG, and DenseNet and settled on a simple U-Net model with a DenseNet201 backbone pre-trained with ImageNet (Krizhevsky et al., 2017) weights. We deliberately opted for a simpler network than in recent research (e.g., in (Mei et al., 2021; Lu et al., 2021b; Zhu et al., 2021; Ghandorh et al., 2022)) to analyze in detail how multispectral bands affect the segmentation process.

A major challenge in our task is the significant imbalance between road and non-road pixels. To tackle this problem, we selected specific hyperparameters for our model through extensive experiments. Since the road pixels account for only 6.33% of the total pixels, we combined the loss functions Focal Tversky Loss (FTL) (Abraham and Khan, 2019) and the Intersection over Union (IoU) (Rahman and Wang, 2016, p. 234-244) to cope with the imbalance between the classes. The FTL (in equation 1) is based on the Tversky Loss (TL) (Salehi et al., 2017) and applies inversely proportional weights to the classes, which helps improve overall results. The IoU (in equation 2) is calculated using the values for true positives (TP), false posi-



(a) RGB image



(b) Multispectral image

Figure 1. Comparison of an RGB and false-color multispectral image with the Green, Red Edge, and Near-IR2 bands (Paris_img290). As it can be seen in 1b, the road surfaces are better reflected in the multispectral image.

tives (FP), and false negatives (FN) and helps to reduce the negative effects of class imbalance. During training, both loss functions were added together to form a hybrid loss function.

$$FTL = \sum (1 - TL)^{\frac{1}{\gamma}} \quad (1)$$

$$IoU = \frac{TP}{FP + TP + FN} \quad (2)$$

Additionally, we used a batch size of four and the Adam optimizer with an initial learning rate of 0.0001, coupled with a learning rate scheduler. Whenever the loss functions did not improve for a certain number of epochs, the learning rate was reduced by a factor of 0.4, allowing the model to learn finer features before finishing its training procedure. To increase efficiency, early stopping and weight-saving recall were implemented. The training was performed on an NVIDIA GeForce GTX 1080Ti 11GB GPU. Two final models were trained with identical hyperparameters on different multispectral and RGB datasets and run through a training (80%), testing (10%), and validation (10%) split, as seen in Table 2. Normalizations and random augmentations such as brightness and contrast adjustments and random flipping were performed. The rotation of the data sets was avoided due to computational constraints.

3.5 Post-Processing Procedures

The binary output of the model categorizes pixels as either *road* or *background*. To improve the segmentation results, extract the road network, and georeference it, a series of post-processing steps are applied: First, the smaller segmented images are stitched back together. Overlapping areas, in which segmentation information is available from

| Model name | Image bands | Epochs |
|-------------------|---------------------------|--------|
| UNetDense_RGB_512 | Red, Green, Blue | 177 |
| UNetDense_MS_512 | Green, Red Edge, Near-IR2 | 161 |

Table 2. Information regarding the model name, used multispectral image bands, and training duration.

up to four images, are merged with the use of an 'OR' operator. This operator prioritizes the labeling of road pixels, as pixels in overlapping areas only need to be predicted as a road once to be classified as such. Then, morphological operations such as Gaussian filtering and thresholding are used to smooth the boundaries and eliminate small, non-road-related features. This process, shown in Fig. 2, refines the images for accurate extraction and georeferencing of the road network.

After post-processing, images are converted into graphs through skeletonization, reducing roads to single-pixel-wide centerlines. Then a multigraph is formed with nodes denoting endpoints and intersections, and edges representing the connecting road segments. To enhance the connectivity and topological correctness of the generated graph, multiple steps are carried out. Using a buffer operation with a radius of 15 pixels, nearby nodes are merged. Depending on the geographical latitude, the selected radius is large enough to merge spatially related nodes, but small enough to prevent the merging of nodes belonging to separate roads. In cases where buffered regions of nodes overlap, they are replaced by a centroid-positioned node, preserving previous edge connections to maintain their original topological state. This maintains topological accuracy and visual coherence. The process, illustrated in Fig. 3, shows how node reduction and connection enhancements



Figure 2. Example of the applied post-processing steps on a binary segmentation result. Blue rectangles highlight areas where connections were established and the red rectangle shows the positive effect of boundary smoothing for graph extraction (Vegas_img1001).

improve the graph structure. In this figure, the blue bounding box shows a case where post-processing established a connection between two previously disconnected nodes.

As a result of the graph creation, each edge contains all the pixel values that the skeletonized road has traversed. This is not a problem for the graph representation, but the derived GeoJSONs contain an unnecessary amount of data. Furthermore, subsequent graph evaluation processes would benefit from simplified edges and lower computational requirements. To simplify the edges, the Ramer-Douglas-Peucker algorithm (Douglas and Peucker, 1973) is used, which preserves the shape of the edges while reducing their complexity. This not only optimizes the representation but also contributes to a more efficient downstream graph analysis.

Finally, the graphs are converted to GeoJSON format by retrieving the coordinate reference frame, pixel size, width, and height of the graphs from the corresponding georeferenced training images. The WGS84 coordinates for each node and edge are calculated and assigned by a planar transformation. The transformed graphs are then saved, with the nodes displayed as points and the edges as lines.

4 Results

In this section, we present our results and describe the evaluation methods used to assess the quality of our results. We discuss the performance of the UNet-Dense_RGB_512 and UNetDense_MS_512 models and examine their results for the entire dataset and in each city. The analysis includes both the quality of the binary segmentation masks and the derived graph networks.

4.1 Evaluation Techniques

While the predicted segmentation mask and the extracted graph represent similar content, their structural differences require separate evaluation methods. To evaluate the performance of the semantic segmentation results, two metrics were used: The IoU, which is integrated into the loss function during the training process; and the F1 metric, which is calculated as the harmonic mean of precision and recall. The F1 metric is widely used in semantic segmentation applications (Máttyus et al., 2017; Xu et al., 2018; Xin et al., 2019) and enables the evaluation of imbalanced datasets.

Evaluating graph networks is a more resource-intensive and complex task than the pixel-wise comparison of images. Graphs consist of nodes and edges that create a relational network that displays the topology and connectivity between nodes. Graph comparison is an actively researched topic with various existing algorithms (Sedgewick, 1998). However, not all of these algorithms are capable of inexact graph matching, a common problem when comparing graphs with different numbers of nodes (Bengoetxea, 2002).

To solve the problem, the graph edit distance (GED) was selected. The GED calculates the number of operations required to transform a proposed graph into its corresponding ground truth graph (Abu-Aisheh et al., 2015). A lower GED value indicates a higher similarity between the graphs. However, this measure does not always reflect the actual complexity, as some graphs may naturally have fewer nodes and edges, resulting in a lower GED score.

To mitigate this problem, we have introduced the relative GED (relGED), which is calculated using the proposal graph (GP), the ground truth graph (GGT), and an empty comparison graph (G0), as shown in equation 3. For each



(a) Different issues in the original graph that need to be solved.

(b) Solved issues after the post-processing.

Figure 3. Example of the applied post-processing steps on the extracted graph network with the impact of node reduction highlighted in red and blue rectangles (Shanghai_img472).

of the GED calculations, the first approximated result of the GED, provided by the Python library NetworkX (Hagberg et al., 2008), is used. This metric offers a relative value that provides context to the graph matching inexactness. Despite its usefulness in fair comparison of inexact graphs, relGED has limitations: It weighs all edges of the graphs equally, which does not match the varying importance of roads in a real-world network. A low relGED value still indicates a good fit of the graph, but it does not differentiate the importance of the different road connections.

$$\text{relGED} = \frac{\text{GED}(\text{GP}, \text{GGT})}{\text{GED}(\text{GP}, \text{G0}) + \text{GED}(\text{GGT}, \text{G0})} \quad (3)$$

The relGED metric measures the similarity of the graphs, but not the topological accuracy of the proposed graph. To capture this, a Topology metric is introduced. This metric compares the connectivity between proposal and ground truth graphs by matching their nodes and calculating the path length similarity. As we could not assume that corresponding nodes from the proposed and ground truth graphs were indexed similarly, we first applied a node-matching algorithm by calculating the Euclidean distance between spatially related nodes and filtering them through a thresholding distance. After the nodes have been matched, all possible shortest paths from each matched node are calculated within its graph using the Dijkstra algorithm (Dijkstra, 2022). The metric then calculates the mean normalized absolute difference in path length and the number of nodes along these paths between the proposed and the ground truth graphs.

4.2 Semantic Segmentation Evaluation

Using the F1 and IoU metrics, we evaluated the performance of each semantic segmentation model, which differs only in the training images. To facilitate detailed analysis and a better understanding of the results, Table 3 presents these metrics for each model together with other city-specific metrics. In addition, examples of the predicted segmentation masks are compared to their corresponding ground truth images in Fig. 4a and Fig. 4b respectively.

4.3 Extracted Road Network Evaluation

As mentioned in subsection 4.1, evaluating the extracted road network requires more advanced methods. Table 3 shows the results for each model and city. This information helps to better understand how the graph extraction methods were influenced by city-specific properties. Furthermore, visual comparisons of the extracted graphs and the georeferenced graphs overlaid on their respective ground truth images are presented in Fig. 4c and Fig. 4d.

5 Discussions

In this work, we addressed the automatic extraction of road networks from multispectral satellite imagery and evaluated the improvements attributed to the selection of non-RGB image channels compared to conventional RGB images. We fine-tuned a U-Net-based semantic segmentation

| Model | F1 in % | IoU in % | GED in steps | relGED in % | Topology in % |
|--------------------------|---------|----------|--------------|-------------|---------------|
| UNetDense_RGB_512 | 82.81 | 73.93 | 66.80 | 46.02 | 86.13 |
| Las Vegas | 89.93 | 82.33 | 67.64 | 43.23 | 89.71 |
| Paris | 44.11 | 31.76 | 61.74 | 59.65 | 72.78 |
| Shanghai | 86.66 | 78.13 | 64.62 | 44.37 | 86.44 |
| Khartoum | 80.94 | 69.65 | 76.63 | 48.77 | 82.06 |
| UNetDense_MS_512 | 88.21 | 80.47 | 61.37 | 43.85 | 88.77 |
| Las Vegas | 90.73 | 83.95 | 65.26 | 42.68 | 90.74 |
| Paris | 85.58 | 77.10 | 33.30 | 43.03 | 90.56 |
| Shanghai | 88.45 | 80.75 | 62.16 | 44.01 | 87.91 |
| Khartoum | 81.22 | 70.74 | 71.65 | 48.04 | 83.55 |

Table 3. Evaluation metric results for the semantic segmentation models and extracted graph networks. Results for the whole dataset are displayed next to the model's names with city-specific results presented in the lines below.

model and performed post-processing steps to improve the extraction of road networks.

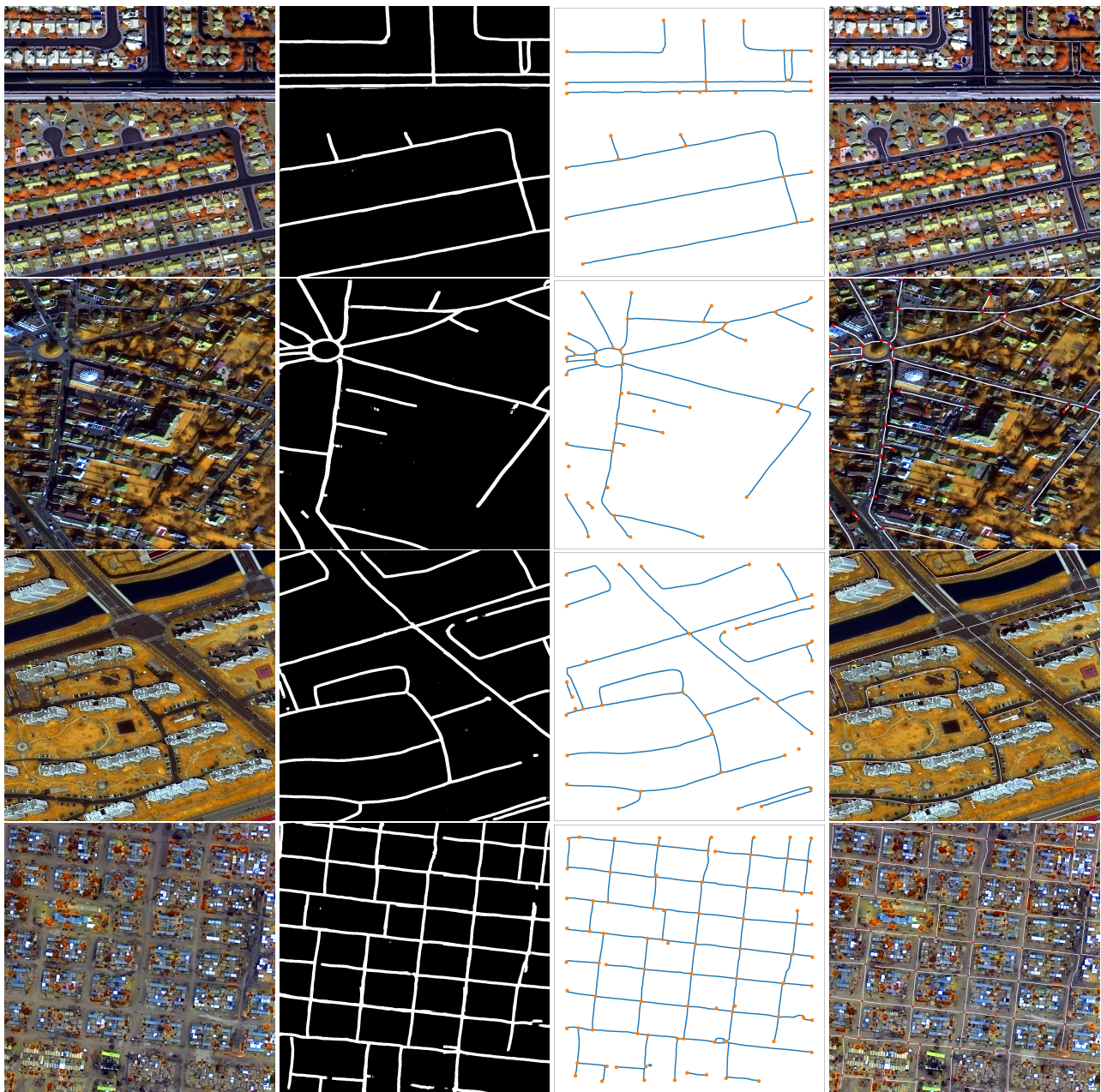
By analyzing the standard GED metric as shown in Table 3, we found that this metric is not entirely reliable, especially in Paris where both models resulted in a low GED value, suggesting a very good graph extraction, while all other metrics reflected the opposite. Meanwhile, the images from the other cities are more or less correctly evaluated by the GED, receiving a high value and thus a poor graph matching, as verified by the other graph evaluating metrics. This issue is in line with observations of Hu et al. (2020) and Zhao et al. (2023) who adapted the GED themselves to enable its use for their specific research question. We assume that this discrepancy in the GED, particularly for Paris, appears due to the low number of nodes and edges contained in the relatively rural road network depicted in the images of Paris. This further highlights the benefit of including the relGED during graph comparisons, as it allows a more thorough and complete analysis.

The number of images from each city provided for training is another important factor to consider when analyzing the results. The inevitable imbalance in the data resulting from the size of the cities meant that the two cities of Las Vegas and Shanghai dominated 79% of the training data. In such a case, even when trying to compensate for minority cities by augmentation techniques or adjusting the loss function as analyzed by Johnson and Khoshgoftaar (2019), it is inevitable that the model automatically learns to adjust its prediction behavior in favor of images from the larger cities. This can be seen in the semantic segmentation metrics of the two cities, which are superior to the images from Paris and Khartoum. The unique characteristics of the road network layout (see subsection 3.1), road width, and occluding objects further influence the results of the metrics. While the Las Vegas images have a uniform, grid-like road layout and few obscuring objects, the Shanghai images have more high-rise buildings and thus more obscured areas, which makes segmentation more difficult and is reflected in the metrics. Whilst the roads of Khartoum follow

a grid-like road layout, these roads are not always paved and/or covered with sand or debris, which leads to an inhomogeneity of all road surface pixels. The road network of Paris follows a rather random distribution, and as there is a lot of vegetation close to the roads, it obscures a large part of the road surface, leading to more non-uniform segmentation results. These effects can be observed, albeit to different degrees, in both evaluated models.

It is important to note that a high score on the segmentation metrics does not always translate into equally good results on the graph metrics, which focus on connectivity and topological accuracy and are not visible in segmented masks. Experiments conducted by Mei et al. (2021) confirm this assumption. The promising graph evaluation metrics of the UNetDense_MS_512 model for Paris imagery could be due to the richer information of the multispectral imagery, which allows better differentiation between roads and other elements. However, the reasons for the poor performance of the UNetDense_RGB_512 model in Paris could go beyond spectral limitations and point to the possible effects of other anomalies that require further investigation.

In the overall evaluation of the model across all cities, the use of multispectral image channels has been shown to notably improve all metric scores. The difference is more pronounced in the segmentation result, where the pixel-based evaluation benefits more from the use of multispectral data. The content displayed in the images is the same regardless of whether multispectral or RGB imagery is used, and thus occluding objects have a similar effect in both image types. This in turn influences the graph evaluation metric, leading to a less pronounced improvement in the graph evaluation metrics. The improvement of 5.4% in the F1, 6.5% in the IoU, 2.2% in the relGED, and 2.6% in the topology metric underline the great potential of the proposed method of including multispectral bands, compared to relying solely on RGB bands. All cities processed by the UNetDense_MS_512 model consistently exhibit superior metric results in comparison to



(a) Ground truth images (b) Predicted segmentation mask (c) Extracted graph network (d) Georeferenced road network

Figure 4. Road extraction results using our proposed pipeline with the UNetDense_MS_512 model as the backbone. From top to bottom the images are from Las Vegas, Paris, Shanghai, and Khartoum.

UNetDense_RGB_512. Notably, the outcomes for the city of Paris stand out prominently.

6 Conclusion

In this work, we have demonstrated the viability of our approach in using multispectral image bands for the automatic extraction of road network graphs using semantic segmentation of road surfaces from high-resolution satellite images. This approach yields satisfactory results and can be utilized to generate up-to-date georeferenced road

networks, suitable for many applications. To utilize the full capabilities of semantic segmentation, we explored different model architectures, backbones, and training techniques. Loss functions were chosen specifically to mitigate problems related to the prevalent class imbalance. Through the analysis, we have shown that our approach of integrating multispectral image bands into the training process leads to very promising results without requiring any changes to the architecture of the model. This enables the use of diverse spectral information in many other areas of application, with the selection of image bands depending on the task at hand. For road segmentation, we recom-

mend the spectral bands Green, Red Edge, and Near-IR2, although future work might benefit from a more in-depth analysis and comparison of different image band combinations. Apart from conducting experiments with more diverse combinations, the potential for enhancing results by applying transfer learning to models trained on different image bands deserves thorough exploration. Furthermore, future work could include the implementation of more advanced semantic segmentation models that are uniquely modified to improve the connectivity of the extracted road networks. To enhance the generalization capabilities of the trained model, the dataset could be enlarged by including additional data, such as the DeepGlobe road dataset (Demir et al., 2018) or by applying additional image augmentations. Further work could include the consideration of different types of road networks during the processing of image data and road graphs, using available semantic information.

References

- Abraham, N. and Khan, N. M.: A Novel Focal Tversky Loss Function With Improved Attention U-Net for Lesion Segmentation, in: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), pp. 683–687, <https://doi.org/10.1109/ISBI.2019.8759329>, 2019.
- Abu-Aisheh, Z., Raveaux, R., Ramel, J.-Y., and Martineau, P.: An Exact Graph Edit Distance Algorithm for Solving Pattern Recognition Problems, in: 4th International Conference on Pattern Recognition Applications and Methods 2015, Lisbon, Portugal, <https://doi.org/10.5220/0005209202710278>, 2015.
- Alhassan, V., Henry, C., Ramanna, S., and Storie, C.: A deep learning framework for land-use/land-cover mapping and analysis using multispectral satellite imagery, *Neural Computing and Applications*, 32, 8529–8544, 2020.
- Bajcsy, R. and Tavakoli, M.: *Computer Recognition of Roads from Satellite Pictures*, 1976.
- Bastani, F., He, S., Abbar, S., Alizadeh, M., Balakrishnan, H., Chawla, S., Madden, S., and Dewitt, D.: RoadTracer: Automatic Extraction of Road Networks from Aerial Images, <https://roadmaps.>, 2018.
- Batra, A., Singh, S., Pang, G., Basu, S., Jawahar, C., and Paluri, M.: Improved road connectivity by joint learning of orientation and segmentation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10 385–10 393, 2019.
- Bengoetxea, E.: Inexact graph matching using estimation of distribution algorithms, Ph.D. thesis, PhD thesis, Ecole Nationale Supérieure des Télécommunications, Paris, France, 2002.
- Cheng, G., Wang, Y., Gong, Y., Zhu, F., and Pan, C.: Urban road extraction via graph cuts based probability propagation, in: 2014 IEEE International Conference on Image Processing (ICIP), pp. 5072–5076, IEEE, 2014.
- Demir, I., Koperski, K., Lindenbaum, D., Pang, G., Huang, J., Basu, S., Hughes, F., Tuia, D., and Raskar, R.: Deepglobe 2018: A challenge to parse the earth through satellite images, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 172–181, 2018.
- DigitalGlobe: WorldView-3, 2017.
- Dijkstra, E. W.: A note on two problems in connexion with graphs, in: Edsger Wybe Dijkstra: His Life, Work, and Legacy, pp. 287–290, 2022.
- Douglas, D. H. and Peucker, T. K.: Algorithms for the reduction of the number of points required to represent a digitized line or its caricature, *Cartographica: the international journal for geographic information and geovisualization*, 10, 112–122, 1973.
- Eerapu, K. K., Ashwath, B., Lal, S., Dell’Acqua, F., and Dhan, A. V. N.: Dense refinement residual network for road extraction from aerial imagery data, *IEEE Access*, 7, 151 764–151 782, <https://doi.org/10.1109/ACCESS.2019.2928882>, 2019.
- Espinoza, C. Z., Khot, L. R., Sankaran, S., and Jacoby, P. W.: High Resolution Multispectral and Thermal Remote Sensing-Based Water Stress Assessment in Subsurface Irrigated Grapevines, *Remote Sensing*, 9, <https://doi.org/10.3390/rs9090961>, 2017.
- Etten, A. V. and Le, M.: CosmiQ/apls, <https://github.com/CosmiQ/apls>, 2020.
- Etten, A. V., Lindenbaum, D., and Bacastow, T. M.: SpaceNet: A Remote Sensing Dataset and Challenge Series, <http://arxiv.org/abs/1807.01232>, 2018.
- Eugenio, F., Marcello, J., and Martin, J.: High-Resolution Maps of Bathymetry and Benthic Habitats in Shallow-Water Environments Using Multispectral Remote Sensing Imagery, *IEEE Transactions on Geoscience and Remote Sensing*, 53, 3539–3549, <https://doi.org/10.1109/TGRS.2014.2377300>, 2015.
- Ghandorh, H., Boulila, W., Masood, S., Koubaa, A., Ahmed, F., and Ahmad, J.: Semantic segmentation and edge detection—Approach to road detection in very high resolution satellite images, *Remote Sensing*, 14, 613, 2022.
- Hagberg, A. A., Schult, D. A., and Swart, P. J.: Exploring Network Structure, Dynamics, and Function using NetworkX, in: Proceedings of the 7th Python in Science Conference, edited by Varoquaux, G., Vaught, T., and Millman, J., pp. 11 – 15, Pasadena, CA USA, 2008.
- Hu, D. K., Mower, A., Shrey, D. W., and Lopour, B. A.: Effect of interictal epileptiform discharges on EEG-based functional connectivity networks, *Clinical Neurophysiology*, 131, 1087–1098, 2020.
- Huang, G., Liu, Z., van der Maaten, L., and Weinberger, K. Q.: Densely Connected Convolutional Networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.
- Jensen, J. R.: *Remote sensing of the environment: An earth resource perspective 2/e*, Pearson Education India, 2009.
- Jin, X. and Davis, C. H.: An integrated system for automatic road mapping from high-resolution multi-spectral satellite imagery by information fusion, *Information Fusion*, 6, 257–273, 2005.
- Johnson, J. M. and Khoshgoftaar, T. M.: Survey on deep learning with class imbalance, *Journal of Big Data*, 6, 1–54, 2019.
- Kent, A. J.: Helping Haiti: some reflections on contributing to a global disaster relief effort, *The Bulletin of the Society of Cartographers*, 44, 2, 2010.
- Krizhevsky, A., Sutskever, I., and Hinton, G. E.: ImageNet classification with deep convolutional neural

- networks, *Communications of the ACM*, 60, 84–90, <https://doi.org/10.1145/3065386>, 2017.
- Li, W., He, C., Fang, J., Zheng, J., Fu, H., and Yu, L.: Semantic segmentation-based building footprint extraction using very high-resolution satellite images and multi-source GIS data, *Remote Sensing*, 11, 403, 2019.
- Lu, X., Zhong, Y., Zheng, Z., and Wang, J.: Cross-domain road detection based on global-local adversarial learning framework from very high resolution satellite imagery, *ISPRS Journal of Photogrammetry and Remote Sensing*, 180, 296–312, 2021a.
- Lu, X., Zhong, Y., Zheng, Z., and Zhang, L.: GAMSNet: Globally aware road detection network with multi-scale residual learning, *ISPRS Journal of Photogrammetry and Remote Sensing*, 175, 340–352, 2021b.
- Mátyus, G., Luo, W., and Urtasun, R.: DeepRoadMapper: Extracting Road Topology From Aerial Images, in: *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017.
- Mei, J., Li, R.-J., Gao, W., and Cheng, M.-M.: CoANet: Connectivity attention network for road extraction from satellite imagery, *IEEE Transactions on Image Processing*, 30, 8540–8552, 2021.
- Mingjun, S. and Daniel, C.: Road Extraction Using SVM and Image Segmentation, 2004.
- Mnih, V. and Hinton, G. E.: LNCS 6316 - Learning to Detect Roads in High-Resolution Aerial Images, 2010.
- Mohanty, S. P., Czakon, J., Kaczmarek, K. A., Pyskir, A., Tarasiewicz, P., Kunwar, S., Rohrbach, J., Luo, D., Prasad, M., Fleer, S., Göpfert, J. P., Tandon, A., Mollard, G., Rayaprolu, N., Salathe, M., and Schilling, M.: Deep Learning for Understanding Satellite Imagery: An Experimental Survey, *Frontiers in Artificial Intelligence*, 3, <https://doi.org/10.3389/frai.2020.534696>, 2020.
- Movaghati, S., Moghaddamjoo, A., and Tavakoli, A.: Road extraction from satellite images using particle filtering and extended Kalman filtering, *IEEE Transactions on Geoscience and Remote Sensing*, 48, 2807–2817, <https://doi.org/10.1109/TGRS.2010.2041783>, 2010.
- O’Callaghan, J., McKinnon, K., Sandev, R., Stanley, S., Žarić, S., Cikamatana, E., Stevens, D., Katsanakis, R., Fysh, A. R., Uribe Perez, C. A., Poussin, F., Ramzy, N., Bray, H., Wejuli, W., Lee, H. S., Keskinen, A., Vita, L., Bussink, C., and Ravan, S.: BLUEPRINT Geospatial for a Better World Transforming the Lives of People, Places and Planet, http://ggim.un.org/meetings/GGIM-committee/10th-Session/documents/2020_UN-Geospatial-Network-Blueprint-Landscape.pdf, 2020.
- Rahman, M. A. and Wang, Y.: Optimizing Intersection-Over-Union in Deep Neural Networks for Image Segmentation, in: *Advances in Visual Computing*, edited by Bebis, G., Boyle, R., Parvin, B., Koracin, D., Porikli, F., Skaff, S., Entezari, A., Min, J., Iwai, D., Sadagic, A., Scheidegger, C., and Isenberg, T., pp. 234–244, Springer International Publishing, Cham, 2016.
- Rodrigue, J.-P.: *The geography of transport systems*, Routledge, 2020.
- Ronneberger, O., Fischer, P., and Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation, <http://arxiv.org/abs/1505.04597>, 2015.
- Salehi, S. S. M., Erdogmus, D., and Gholipour, A.: Tversky loss function for image segmentation using 3D fully convolutional deep networks, in: *International workshop on machine learning in medical imaging*, pp. 379–387, Springer, 2017.
- Sedgewick, R.: *Algorithms in C++, parts 1-4: fundamentals, data structure, sorting, searching*, Pearson Education, 1998.
- Shahi, K., Shafri, H. Z., Taherzadeh, E., Mansor, S., and Mu-niandy, R.: A novel spectral index to automatically extract road networks from WorldView-2 satellite imagery, *The Egyptian Journal of Remote Sensing and Space Science*, 18, 27–33, 2015.
- Thoma, M.: A Survey of Semantic Segmentation, <http://arxiv.org/abs/1602.06541>, 2016.
- Warren, G. and Metternicht, G.: Agricultural applications of high-resolution digital multispectral imagery, *Photogrammetric Engineering & Remote Sensing*, 71, 595–602, 2005.
- Xin, J., Zhang, X., Zhang, Z., and Fang, W.: Road extraction of high-resolution remote sensing images derived from DenseUNet, *Remote Sensing*, 11, <https://doi.org/10.3390/rs11212499>, 2019.
- Xu, Y., Xie, Z., Feng, Y., and Chen, Z.: Road extraction from high-resolution remote sensing imagery using deep learning, *Remote Sensing*, 10, <https://doi.org/10.3390/rs10091461>, 2018.
- Yuan, K., Zhuang, X., Schaefer, G., Feng, J., Guan, L., and Fang, H.: Deep-learning-based multispectral satellite image segmentation for water body detection, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14, 7422–7434, 2021.
- Zhang, J., Chen, L., Zhuo, L., Geng, W., and Wang, C.: Multiple Saliency Features Based Automatic Road Extraction from High-Resolution Multispectral Satellite Images, *Chinese Journal of Electronics*, 27, 133–139, 2018.
- Zhao, Y., Qi, J., Trisedya, B. D., Su, Y., Zhang, R., and Ren, H.: Learning Region Similarities via Graph-based Deep Metric Learning, *IEEE Transactions on Knowledge and Data Engineering*, 2023.
- Zhu, Q., Zhang, Y., Wang, L., Zhong, Y., Guan, Q., Lu, X., Zhang, L., and Li, D.: A global context-aware and batch-independent network for road extraction from VHR satellite imagery, *ISPRS Journal of Photogrammetry and Remote Sensing*, 175, 353–365, 2021.
- Ünsalan, C. and Sirmacek, B.: *Road Network Detection using Probabilistic and Graph Theoretical Methods*, 2011.