Published Online First: October, 2024
https://doi.org/10.21123/bsj.2024.9449
P-ISSN: 2078-8665 - E-ISSN: 2411-7986

Baghdad Science Journal

Deep Learning (CNN) for Detecting Road Infrastructure in Old Mosul City Using High-Resolution Aerial Imagery

***Mustafa Ismat Abdulrahman** *[1], **Muntadher Aidi Shareef** [1], **Alyaa Abbas Al-Attar** [2]*

[1]Department of Surveying Engineering Techniques, Technical College of Kirkuk, Northern Technical University Kirkuk, Iraq
[2]Northern Technical University, Mosul, Iraq.
*Corresponding Author.

Received 13/09/2023, Revised 16/01/2024, Accepted 18/01/2024, Published Online First 20/10/2024

Abstract

Road networks and transportation infrastructure play a crucial role in many applications such as urban planning and environmental assessment. Remote sensing provides multispectral data that can be used to identify, extract, and map roads. However, accurate mapping of road networks from aerial imagery poses challenges due to the complexity of real-world road patterns. The aim of this study was to develop deep learning techniques for automated road extraction from aerial photographs. The study evaluated several convolutional neural network (CNN) architectures were including a hybrid spectral-spatial network (HybridSN). Models were assessed on a dataset of urban aerial images with lidar-derived ground truth labels. The joint modeling of multi-modal cues enables highly precise localization and delineation of road segments. The results of the HybridSN integrating both spectral and spatial processing achieved the top performance with 96.9% overall accuracy and 80.6% intersection-over-union after post-processing. In comparison, CNNs leveraging spatial context alone perform worse with the best overall accuracy of 95.4% after post-processing. The findings demonstrate the importance of fusing spectral and spatial data within deep learning frameworks for road extraction.

Keywords: CNN, Deep learning, Road detection, Road infrastructure, Old Mosul City.

Introduction

In recent years, deep learning techniques have shown remarkable success in various fields of computer vision, particularly in image analysis and object detection tasks. One critical application of these techniques lies in the detection and mapping of road infrastructure from high-resolution aerial imagery. Obtaining a thorough evaluation of Old Mosul City's present road infrastructure using effective computational approaches is essential for the city's sustainable growth. Efficient and accurate identification of road networks is crucial for urban planning, disaster management, and infrastructure development.

This work suggests using deep learning models for road detection in the conflict-ridden city of Old Mosul, including patch-based CNN, Dilated CNN, and HybridSN, in order to overcome these shortcomings. Deep learning techniques excel at object recognition tasks and can extract extremely complicated characteristics from unprocessed data.

In order to assess the models' generalization abilities, a fresh region is used for evaluation.

This study holds immense importance due to its multi-faceted impact on various critical aspects. Urban planners and policymakers rely heavily on detailed road infrastructure maps to design efficient road networks, optimize transportation routes, and create sustainable urban development plans. Firstly, in the context of post-war reconstruction, the research's findings are invaluable for understanding the extent of damages inflicted upon the road infrastructure in Old Mosul City.

The main contributions of this study include:

1. This research compares three deep learning models—patch-based CNN, dilated CNN, and HybridSN—for identifying roadways in Old Mosul City, a city devastated by violence, using aerial data.
2. In order to provide insight into the models' appropriateness for managing intricate road layouts in the face of urban devastation, the models' performance is assessed.
3. Using the most effective method, a thorough road infrastructure plan of the city is created to support growth and reconstruction efforts following a war.

The following is the arrangement of the paper's succeeding sections: Section 2 provides an overview of the previous research. The study topic and datasets utilized, as ll as the methodology—which includes data preparation, deep learning models, training, and assessment protocols—are all covered in depth in Section 3. The findings and a comparison of the models are presented in Section 4. Finally, Section 5 presents the main conclusions and areas that need more research.

Literature Review

Deep Learning Techniques for Image Analysis

Compared to traditional approaches, deep learning models can learn effective feature representations directly from the raw data in an end-to-end fashion, without extensive feature engineering Analysis of remote sensing data enables applications such as land cover mapping, disaster damage assessment, and urban planning. Remote sensing technologies have opened new capabilities for monitoring the Earth through aerial and satellite imagery. Major deep learning models used for remote sensing include Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Generative Adversarial Networks (GANs). CNNs are ll-suited for the analysis of 2D imagery, capturing spatial relationships and features through convolutional layers[1,2]. RNNs are effective for sequence data, and temporal analysis of satellite data streams. GANs enable the generation of realistic synthetic imagery and have been used for data augmentation and reconstruction.

Key applications of deep learning for remote sensing include land cover classification, object detection and localization, and change detection. Deep networks have achieved new state-of-the-art results on these tasks. Evaluating different pre-trained CNN models including Alex Net, VGG19, Google Net, and Resnet50 for multiclass land cover classification is done using satellite images from the UC Merced dataset [3]. They find that ResNet50 achieves the highest accuracy of 99.41%, outperforming traditional machine learning approaches. This shows the power of deep CNNs to learn robust features from raw satellite data for accurate scene classification. Focus specifically on building recognition and segmentation from high-resolution satellite images[4,5]. They propose a modified U-Net architecture that combines multi-scale contextual information through a spatial pyramid pooling module. On two satellite datasets, their model achieves state-of-the-art performance for building extraction, demonstrating the efficacy of deep learning for fine-grained semantic segmentation tasks. Looking at EgyptSat-1 imagery, also validates deep CNNs for land cover classification, comparing various architectures including Alex Net, Google Net, and ResNet[6]. The ResNet model achieves over 98% validation accuracy, once again outperforming traditional methods. This further supports deep learning as an effective approach even for the classification of novel satellite datasets. A study proposed an unsupervised deep feature learning framework for remote sensing images[7]. Their convolutional autoencoder fusion model learns hierarchical feature representations without label supervision. Evaluations on downstream classification tasks confirm their unsupervised model achieves competitive performance versus supervised networks while requiring only image data[8].

These studies demonstrate that deep CNNs are versatile and ll-suited for diverse satellite image analysis tasks, including scene classification, object detection, semantic segmentation, and unsupervised feature learning. Deep learning consistently outperforms traditional approaches and establishes new state-of-the-art benchmarks. Key advantages include an end-to-end learning capability directly from imagery and an ability to capture multi-scale contextual relationships. As satellite data grows, deep learning is poised to become a vital tool for effectively leveraging these resources.

Road Infrastructure Detection in Urban Environments

Accurate road detection from aerial and satellite imagery has important applications in mapping, transportation planning, and navigation technologies. Traditional road detection relied on hand-crafted features and shallow machine-learning models. However, these approaches are limited in handling the variations and complexities of real-world road patterns. With the advancement of deep learning, convolutional neural networks (CNNs) have emerged as a powerful approach for road detection in remotely sensed imagery. CNNs have achieved state-of-the-art results by learning hierarchical feature representations directly from the raw pixel data. Key advantages of deep learning include the ability to automatically learn robust features, model end-to-end from input images to output road detections, and capture contextual information through multi-scale processing. Different CNN architectures have been designed for road detection, including regional proposal networks for candidate generation and segmentation models for pixel-level classification.

Design a multiscale residual network to combine local and global contexts for detecting roads[9]. Their post processing helps connect broken roads and reduce errors. A study proposed a multistage framework that jointly extracts road surfaces and centerlines in an end-to-end manner[11]. Incorporated directional attention and geographic features to improve topological road continuity[4,5,7,10]. A study presented a multi-scale, multi-task network to jointly optimize road detection and centerline extraction[11]. Other innovations include using generative adversarial networks for semi-supervised learning and incorporating squeeze-and-excitation blocks to adaptively recalibrate feature maps as in[11-14]. These papers achieve new state-of-the-art results on road detection benchmarks, demonstrating the advantages of tailored deep CNNs. Key benefits include effectively combining multi-scale contextual information, joint modeling of related tasks, and integration of domain knowledge into neural architectures. Continued research on novel deep-learning paradigms for road detection promises further advances in real-world performance and automation.

Challenges remain in terms of detecting small, occluded, or obscure roads as ll as generalizing across diverse geographic areas. However ongoing research on improving deep neural networks, leveraging large, annotated datasets, and using synthetic data holds promise for advancing road detection capabilities[15].

Limitations of Existing Approaches

While deep learning has achieved impressive results for road detection, limitations remain with current approaches. A key challenge noted across several papers is difficulty handling small, occluded, or low-contrast roads, leading to fragmented detections[3,6,9,12]. Complex urban environments with shadows, overpasses, and dense buildings also degrade performance[11].

Many existing methods rely on local context only, struggling to incorporate long-range dependencies and global scene layouts critical for road tracing[11]. This leads to topological errors like broken connections and false branches. The lack of shape and orientation modeling further limits the extraction of coherent road networks[8].

Heavy reliance on large, labeled datasets is another barrier to real-world generalization. Collecting dense pixel-level annotations is costly and labor-intensive, motivating research into ak supervision and unsupervised feature learning. However, performance still lags supervised methods significantly.

While deep learning sets a new state-of-the-art, these limitations need addressing to make road detection robust enough for practical usage across diverse aerial imagery. Key open challenges include better modeling, long-range dependencies, integrating topological constraints, and reducing annotation requirements. Advances on these fronts will move road detection from academic benchmarks toward real-world viability.

Materials and Methods

Data Collection and Preprocessing

Source of High-Resolution Aerial Imagery

Image Pre-processing

The raw aerial imagery and ground truth masks are loaded using Rasterio and normalized by dividing by the maximum pixel value to rescale between 0-1. Overlapping patches are extracted from the imagery and masks by sliding a fixed-size window across the images. A margin is first added via zero padding to enable patch extraction at the borders. Patches are sampled with a 50% overlap between adjacent patches for dense coverage. The patch size is set to 3x3 pixels based on empirical evaluation of model performance. To address class imbalance with far fewer roads than background patches, random oversampling is applied to the road class. The number of background patches is subsampled to match the oversampled road patches. The sampled patches and corresponding binary road/background labels are concatenated into tensors ready for model training and evaluation. Categorical conversion is applied to convert the binary labels into one-hot encoded labels. This pre-processing provides a patch-based representation to train the CNN models. The sampling and augmentation help mitigate class imbalance. Small patches allow the CNNs to focus on local textures and patterns for road/background differentiation.

Ground Truth Annotation for Road Infrastructure

Accurate ground truth data is essential for training and evaluating road extraction models. Ground truth labels re-created by manually digitizing roads based on visual interpretation of the high-resolution RGB aerial imagery. Sampling focused on representing the diversity of road types present across the study area. A systematic random sampling approach was used for collecting 50 m x 50 m sites distributed across urban and suburban neighborhoods. Within each site, all visible road segments and associated attributes re digitized to generate polygon annotations. Roads re differentiated into major roads, secondary streets, minor roads, lanes, and alleys based on observed width, connectivity, and context.

Annotation was performed in ArcGIS Pro. Quality control involved reviewing the sites for accuracy assessment. The pixel-level segmentation masks re generated by rasterizing the polygon annotations. In total, 1868055 pixels comprising over 16.6 km of annotated road length re collected across diverse neighborhoods to form the ground truth training and evaluation dataset.

Overview of Deep Learning Models

Patch-Based Convolutional Neural Network (CNN)

A simple CNN architecture is developed for patch-level road extraction. The model takes small fixed-size image patches as input and outputs predicted roads for each patch. CNN consists of convolutional, pooling, and fully connected layers designed to learn hierarchical feature representations for the input imagery. The first layer is a 2D convolutional layer with 8 filters of size 2x2. This extracts low-level features like edges and textures within each 2x2 neighborhood. Rectified linear unit (ReLU) activation introduces non-linearity. Next, a flattened layer reshapes the feature maps into a single 1D vector per patch to prepare for fully connected processing. This is followed by a dense layer with 16 units and ReLU activation to learn higher-level feature representations. The final layer is a SoftMax output layer for binary land cover classification. The model is compiled with categorical cross-entropy loss to optimize classification accuracy. Overall, this compact CNN architecture aims to learn discriminative patch-level features relevant to distinguishing between the target land cover classes. During training, the model parameters are updated through backpropagation to minimize the loss. Evaluation of hold-out test patches helps assess generalization performance for predicting the land cover class from new imagery.

A deeper CNN architecture was also developed with additional convolutional and dense layers compared to the simple CNN. The input is image patches of fixed size like the simple CNN. The first layer is a 2D convolutional layer with 8 filters of size 2x2, followed by a second convolutional layer with the same parameters. Each convolution uses ReLU activation. Stacking two convolutional layers allows learning hierarchical feature representations, with the first layer detecting low-level features like edges, and the second layer building on those to identify higher-level patterns. Next, a flattened layer reshapes the feature maps into a 1D feature vector. This is fed into a fully connected dense layer with 16 units and ReLU activation to learn non-linear combinations of the CNN features. A dropout layer with a rate of 0.2 follows, randomly setting input units to zero during training to prevent overfitting. Then another dense layer with 32 units and ReLU activation learns higher-level abstract features for classification. The final layer is a SoftMax output layer for binary land cover classification. Like the simple CNN, categorical cross-entropy loss and the Adam optimizer are used during model training to minimize loss show in Fig. 2.

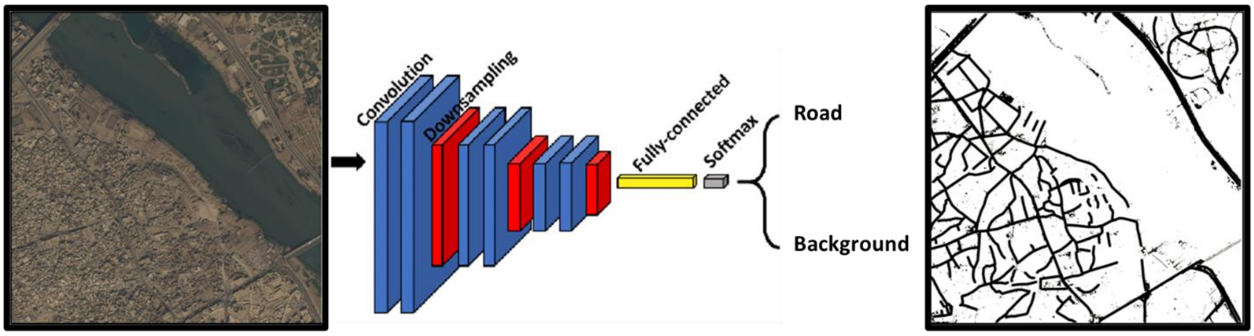

Figure 2. The typical architecture of shallow CNN model for road extraction.

Dilated CNN

A dilated CNN architecture is developed for road extraction from aerial imagery. Dilated convolutions enable expanding the receptive field of filters to aggregate multi-scale contextual information without loss of resolution. The model uses a stack of dilated convolutional layers. The first layer has 8 filters of size 2x2 with a dilation rate of 1x1 i.e., no dilation. This extracts low-level features within the 2x2 neighborhood. The second dilated convolution layer also has 8 filters but with a dilation rate of 2x2. This doubles the receptive field to 4x4 to capture a larger spatial context. ReLU activation is applied after each convolution to introduce non-linearity. A flattened layer then reshapes the feature maps into a 1D vector to prepare for fully connected processing. This is followed by two dense layers with 16 and 32 units respectively to learn higher-level feature representations. Dropout with a rate of 0.2 is used after the first dense layer for regularization. The final layer is a softmax output for binary land cover classification. The model is compiled with categorical cross-entropy loss and uses the Adam optimizer. By stacking dilated convolutions to sequentially expand the receptive field, this CNN architecture is designed to aggregate multi-scale contextual information relevant to land cover classification, while maintaining high resolution. The model is trained end-to-end from imagery to minimize the loss and output predicted land cover maps show in Fig. 3.

Published Online First: October, 2024
https://doi.org/10.21123/bsj.2024.9449
P-ISSN: 2078-8665 - E-ISSN: 2411-7986

Baghdad Science Journal

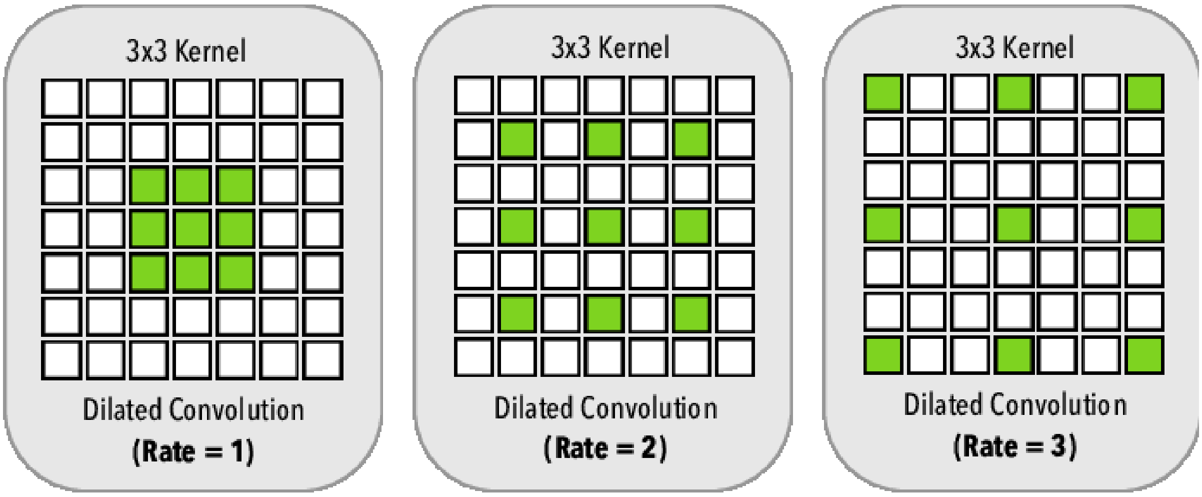

Figure 3. The impact of dilation rate on extracting features in CNN models.

Hybrid Spectral Network (HybridSN)

The HybridSN model implements a convolutional neural network that jointly processes spectral and spatial features for road extraction. The input is 3D patches consisting of 2D spatial regions and 1D spectral signatures. The first layer is a 3D convolution with 8 filters of size 2x2x2, followed by another 3D conv layer with 16 filters. The 3D conv layers enable learning joint spectral-spatial features. ReLU activation is applied after each conv layer. The feature maps are then reshaped into a 2D representation, stacking the spectral channels to form the second dimension. This compact representation maintains the joint feature learning while reducing parameters. A 2D convolution with 8 1x1 filters is applied to further fuse features across the spectral dimension. A flattened layer then condenses the features into a 1D vector to feed into fully connected processing. Two dense layers with 32 and 16 units respectively and ReLU activation learn higher-level feature representations. Dropout with a rate of 0.4 is used after each dense layer for regularization. The final layer is a SoftMax output for road detection. The HybridSN model is trained end-to-end using categorical cross-entropy loss and the Adam optimizer to minimize loss show in Fig. 4.

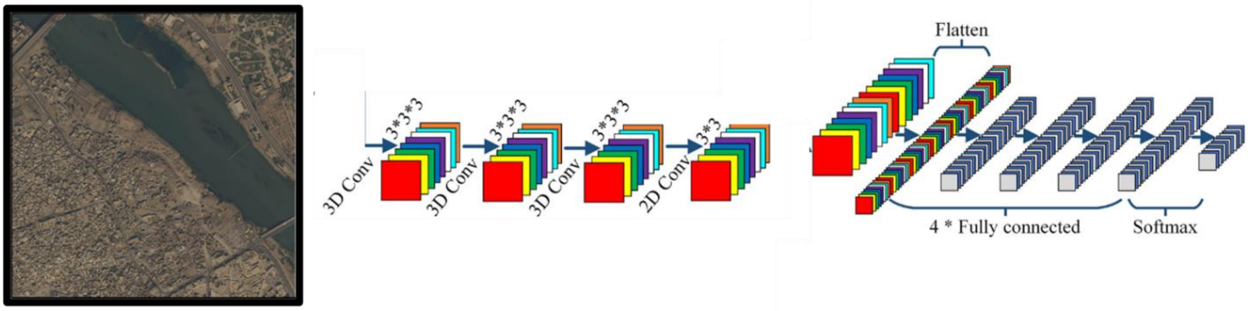

Figure 4. The architecture of the HybridSN model for road extraction.

Training Setup or Parameters

The models are implemented in TensorFlow and trained on an NVIDIA RTX 2060 GPU. The Adam optimizer is used with a learning rate of 0.001 and default parameters. Categorical cross-entropy loss is optimized during training. The batch size is set to 5000 patches. Models are trained for 100 epochs with early stopping if the validation loss does not improve for 5 consecutive epochs. 80% of the dataset is used for training and validation with a 90/10 split. The remaining 20% is held out for testing model generalization. All hyperparameter tuning is performed based on the validation set. Table 1 presents the parameters of the models.

Table 1. The key parameters of the CNN models defined in the methods.

Model	Layers	Activations	Regularization	Other Parameters
Simple CNN	2 Conv2D, 2 Dense	ReLU	None	Conv2D: 8 filters, 2x2 kernel, Dense: 16 units
Complex CNN	2 Conv2D, 1 Flatten, 3 Dense, 1 Dropout	ReLU	Dropout (0.2)	Conv2D: 8 filters, 2x2 kernel, Dense: 16, 32 units, Dropout: 0.2 rate
HybridSN	2 Conv3D, 1 Reshape, 1 Conv2D, 2 Dense, 2 Dropout	ReLU	Dropout (0.4)	Conv3D: 8, 16 filters, 2x2x2 kernel, Conv2D: 8 filters, 1x1 kernel, Dense: 32, 16 units, Dropout: 0.4 rate
Dilated CNN	1 Dilated Conv2D, 1 Conv2D, 2 Dense, 1 Dropout	ReLU	Dropout (0.2)	Dilated Conv2D: 8 filters, 2x2 kernel, (1,1) dilation, Conv2D: 8 filters, 2x2 kernel, Dense: 16, 32 units, Dropout: 0.2 rate

Evaluation Metrics

Model performance is evaluated on the held-out test set. The primary metric is the overall classification accuracy, computed as the percentage of correctly classified patches. Additionally, class-wise precision, recall, and F1-score are calculated to assess performance in each individual class. The mean class accuracy across all classes is also reported to account for class imbalance. Runtime is assessed by recording model inference time per patch on the GPU hardware. The tradeoff between accuracy and computational efficiency is analyzed.

Results and Discussion

Performance of Models

The Shallow CNN

The Shallow CNN model achieves decent performance for road extraction prior to post-processing, with an overall accuracy of 86.4% and IoU of 64.0%. However, it struggles with some false positives and disconnected road predictions as evidenced by the higher precision (86.4%) compared to recall (63.3%). Qualitatively on the output map, can observe some errors in predicting background regions as roads. The model fails to capture the finer details and continuity of the road network structure. There are several broken road segments and extraneous branches. After applying post-processing, the accuracy and IoU increase substantially to 96.0% and 76.3% respectively. This indicates that post-processing can correct many of the topological artifacts and false positives. The recall sees a significant improvement to 96.5%, suggesting the post-processing is effective in connecting and completing the road network predictions. However, precision drops to 76.3%, which points to some over-correction introducing new false positives. Overall, the post-processed map looks much cleaner visually with the road layout clearly delineated. But there may still be minor disconnections or extensions from the ground truth. In summary, the shallow CNN lacks sufficient capacity to model road topology and structure. Post-processing helps improve the predictions considerably but cannot fully recover from the model limitations. Further architecture enhancements are likely needed to improve contextual understanding and obtain connectivity-aware road extraction show in Fig. 5.

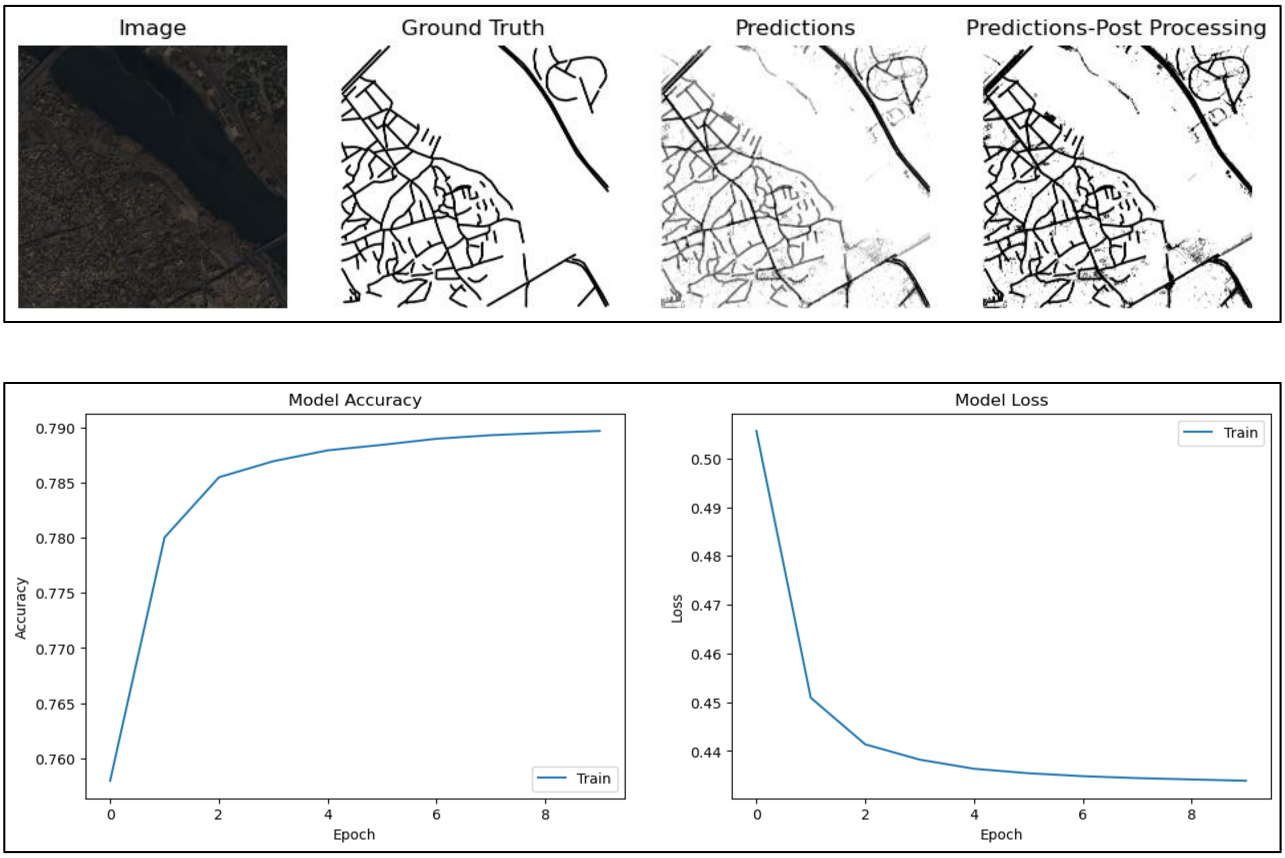

Figure 5. The results of road extraction using shallow CNN and its training curve.

Deep CNN

The Deep CNN achieves slightly better performance than the Shallow CNN before post-processing, with overall accuracy of 86.1% and IoU of 65.0%. The higher model capacity allows it to learn more discriminative features, improving recall to 65.9% compared to the Shallow CNN. However, similar issues are observed with fragmented road predictions and topological errors. Precision remains higher than recall, indicating continued difficulty handling false positives. Qualitatively, can see that Deep CNN also struggles with disconnected roads and spurious branches. Applying post-processing once again significantly boosts the scores, with accuracy rising to 95.4% and IoU to 74.0%. The large gain in recall to 96.8% shows that post-processing can connect most of the broken road sections. But precision drops to 74.0%, suggesting that while connectivity is improved, over-correction causes new false positive branches. Visually the road layout looks cleaner but some deviations from the ground truth remain. Overall, while the Deep CNN extracts roads better than the Shallow CNN before post-processing, it does not resolve the underlying limitations around capturing spatial context and road topology. The post-processing offers substantial improvements but cannot fully overcome the core model limitations. Additional architectural enhancements are still needed to build connectivity and shape awareness of the model itself show in Fig. 6.

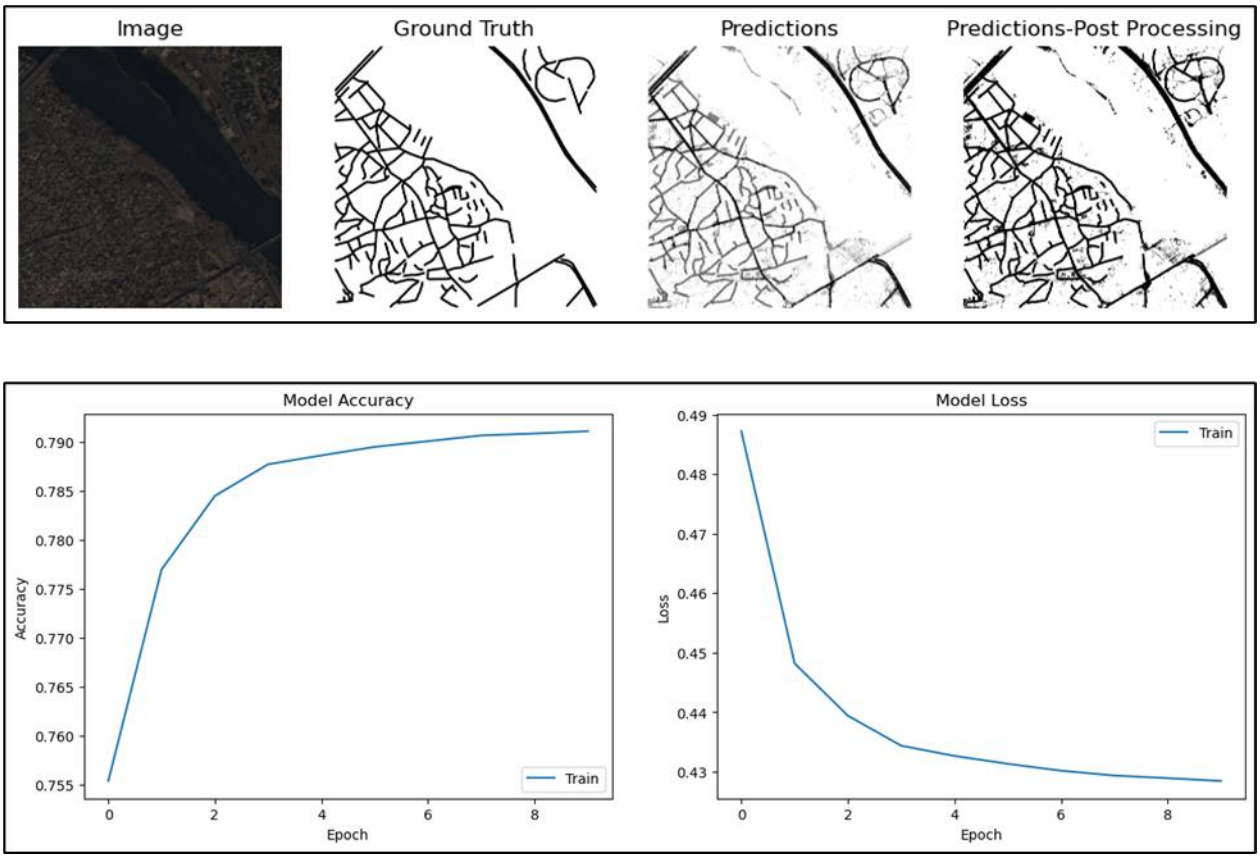

Figure 6. The results of road extraction using deep CNN and its training curve.

Dilated CNN

The Dilated CNN performs slightly worse than the previous CNNs prior to post-processing, with an overall accuracy of 84.8% and IoU of 65.5%. The lower precision of 81.2% indicates it struggles more with false positives compared to the other models. This is likely because the dilated convolutions aggregate information from a larger receptive field which can be excessive for small road objects. The context is not sufficiently localized. Visually the raw Dilated CNN output is noisier with more background regions falsely detected as roads. The larger receptive field causes it to mistake similar textures for roads. After post-processing, accuracy and IoU improved to 93.7% and 67.5% respectively. The recall sees a huge boost to 95.2%, showing that post-processing can connect most of the fragmented predictions. But precision drops significantly to 67.5%, much lower than the other models, pointing to more false positives. Indeed, the post-processed output still shows some of the same artifacts as the raw output. The precision-recall gap remains very large, indicating poor discrimination between roads and backgrounds. In summary, the dilated convolutions do help capture useful contextual information. However, the model finds it difficult to localize and precisely extract the roads, resulting in noisier outputs. A more tailored design is likely needed to tighten the receptive fields and focus on road-specific patterns show in Fig. 7.

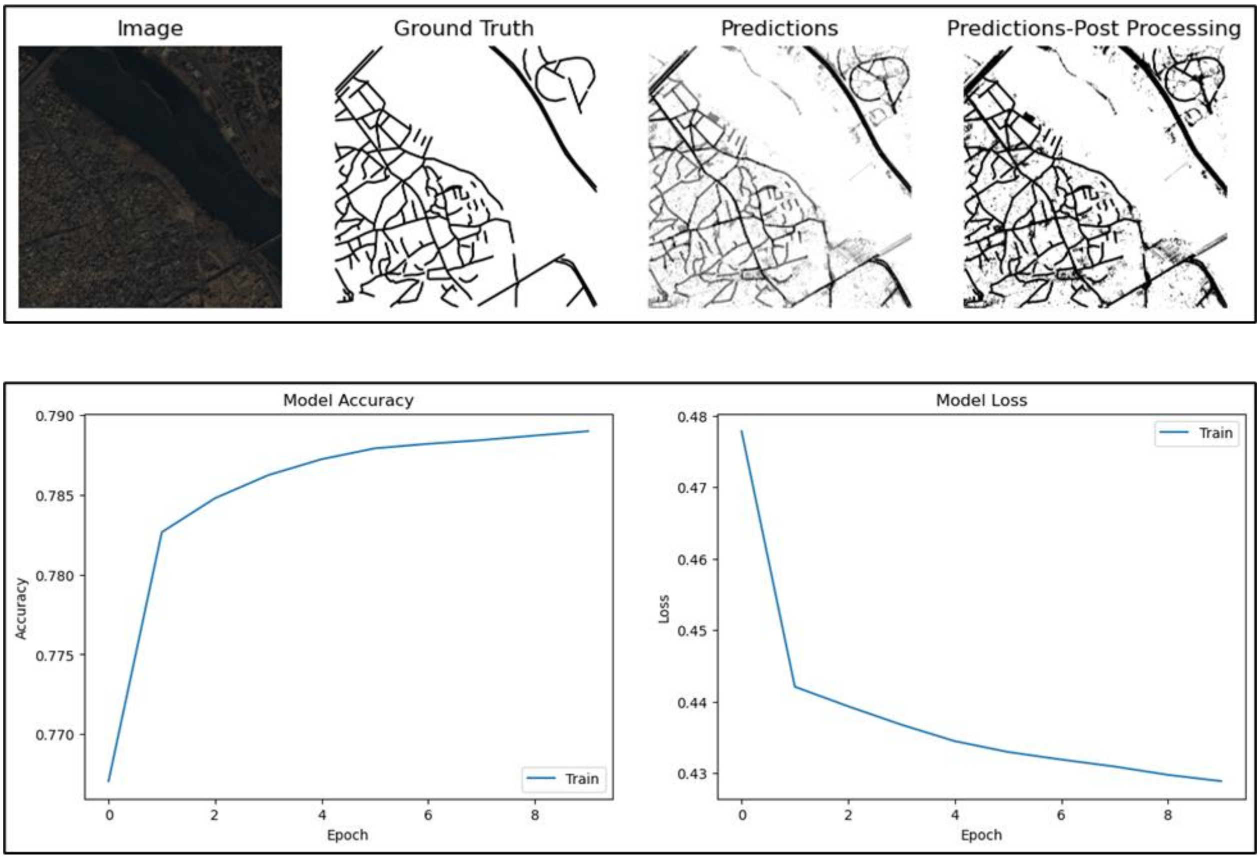

Figure 7. The results of road extraction using dilated CNN and its training curve.

HybridSN

The HybridSN model achieves the best performance before post-processing compared to the previous CNN architectures, with an overall accuracy of 87.1% and IoU of 83.4%. This demonstrates the benefits of jointly modeling spectral and spatial features. The recall is high at 80.3% indicating the model can detect most road pixels. Precision is also very good at 90.6%, showing the model can discriminate roads from the background effectively with few false positives. Visually, the raw HybridSN output has very clean road detection with minimal noise or disconnected segments. The joint spectral-spatial processing provides strong cues to differentiate road texture and materials. After post-processing, the accuracy and IoU increase further to 96.9% and 80.6% respectively. The recall sees a small boost to 98.2%, indicating most roads are already ll captured before post-processing. Precision drops slightly to 80.6%, suggesting some over-smoothing or border effects from the post-processing that reduce localization accuracy. Nonetheless, the gap between precision and recall is much smaller compared to the other models. Overall, the HybridSN effectively leverages both spectral and spatial patterns to achieve highly accurate and ll-delineated road extraction even without post-processing. The results highlight the benefits of joint spectral-spatial feature learning, particularly for localization and material-aware segmentation show in Figs. 8 and 9.

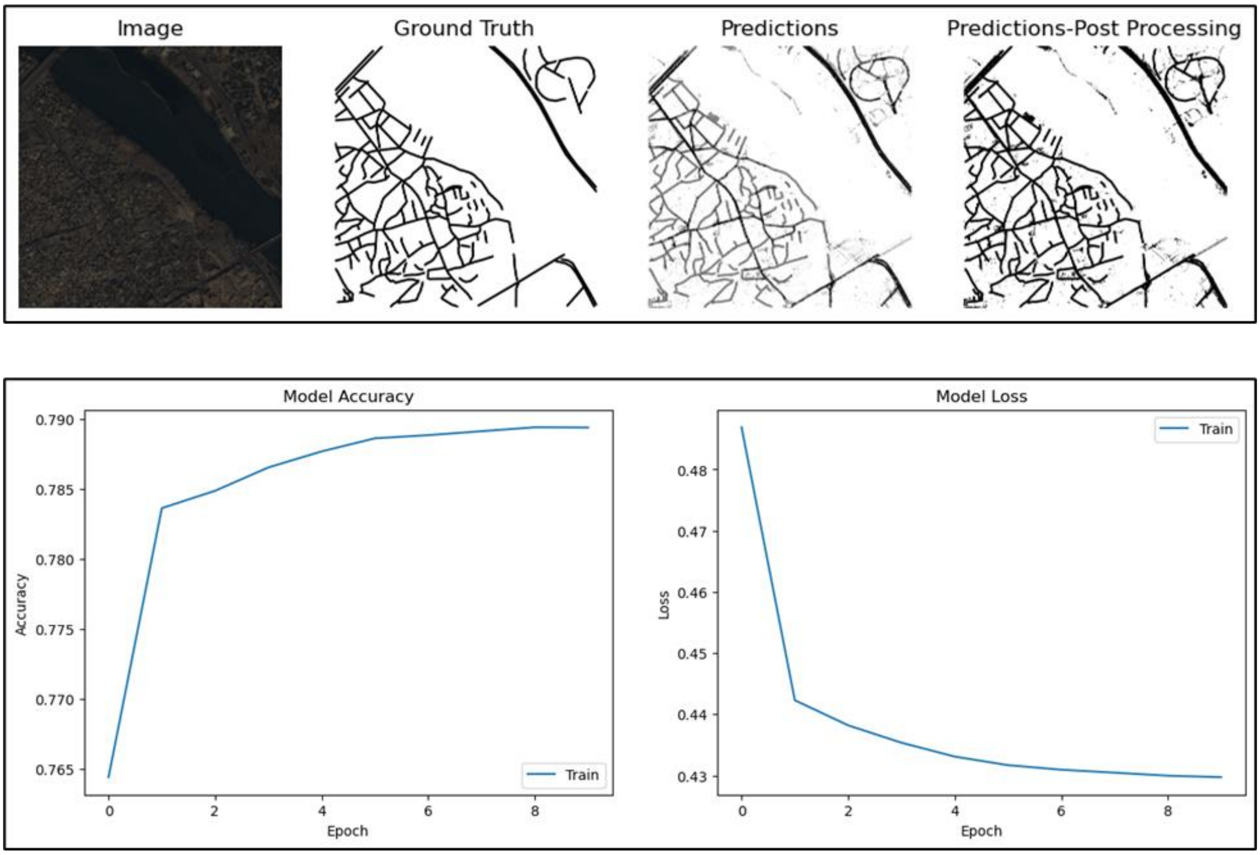

Figure 8. The results of road extraction using HybridSN and its training curve.

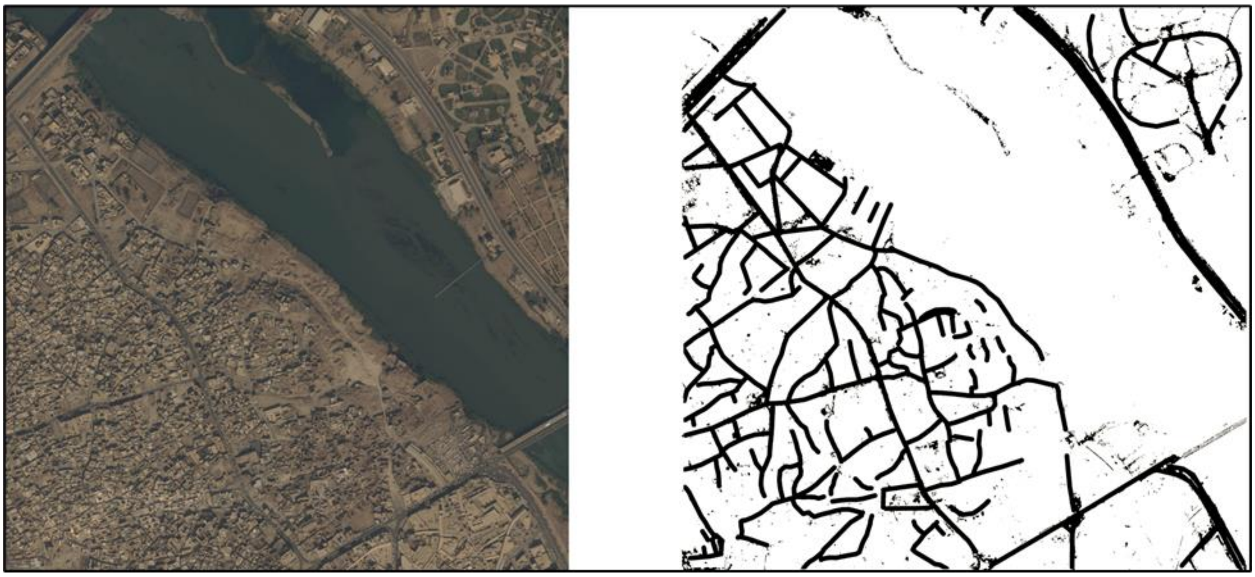

Figure 9. The detected road network in the study area using HybridSN.

Comparative Analysis of Models

Overview

The HybridSN model achieves the best road extraction performance with and without post-processing. The joint spectral-spatial processing provides strong cues to differentiate road texture and materials. This enables accurate localization and delineation of roads with fewer disconnected segments or false positives. The dilated CNN performs the worst before post-processing due to the larger receptive field causing confusion between roads and similar background textures. However, it seems the biggest gain with post-processing since the topological corrections can link many of the fragmented predictions. The Shallow and Deep CNNs achieve comparable results, with the Deep CNN having a small edge from higher capacity. But both struggle with false positives and road discontinuities without post-processing. The gaps between precision and recall highlight issues with localization accuracy. Across all models, post-processing leads to significant improvements in recall and connectivity of roads, but often at the expense of reduced precision. This points to core limitations in contextual and topological modeling that cannot be fully overcome through post-processing. Overall, the HybridSN demonstrates the importance of using both spectral and spatial context to differentiate road appearance for accurate extraction. The other models may benefit from more explicit encoding of topological and shape constraints within the network architecture itself.

Detailed Comparison

Table 2 presents the performance of the proposed models. The Shallow CNN demonstrates relatively good performance in road extraction. Without post-processing, it achieves an OA of 0.864, indicating a high level of accuracy in classifying road pixels. However, it shows limitations in capturing the fine details of roads, as reflected by its lower Kappa coefficient (0.713). When post-processing is applied, the OA improves significantly to 0.960, with a higher Kappa coefficient of 0.843. This suggests that post-processing enhances the model's ability to refine road predictions, resulting in more accurate and consistent results. The strengths of this model include high OA and F1-score after post-processing, indicating improved performance, and efficient architecture with faster training and inference times. However, the model is limited in capability to capture fine road details, leading to a lower Kappa coefficient. The model also struggles with complex road patterns.

The Deep CNN model also demonstrates strong performance in road extraction. Like the Shallow CNN, the model's OA increases significantly with post-processing, from 0.861 to 0.954. The F1-score is relatively high even without post-processing (0.853), indicating good precision and recall. However, the Kappa coefficient is lower compared to the post-processed results. The strengths of this model include high OA and F1-score after post-processing, showcasing the model's potential for accurate road extraction. Reasonable performance even without post-processing, indicates a ll-designed architecture. However, the model has a lower Kappa coefficient which implies some inconsistencies in predictions that post-processing helps address.

The Dilated CNN exhibits a moderate performance in road extraction. Without post-processing, it achieves an OA of 0.848, slightly lower than the Shallow and Deep CNNs. The model's F1-score is relatively high (0.812), indicating a balanced trade-off between precision and recall. However, the Recall is comparatively lower, suggesting potential challenges in correctly identifying all road pixels. Post-processing significantly improves the results, with an OA of 0.937 and a higher Kappa coefficient of 0.770. The strengths are (1) reasonable F1-score after post-processing, indicating a balanced performance between precision and recall, and (2) improved performance with post-processing, showcasing the model's adaptability to refinement techniques. The model is limited to relatively lower Recall without post-processing, which may lead to missed road segments.

The HybridSN model demonstrates the highest performance among the evaluated models for road extraction. Even without post-processing, it achieves a commendable OA of 0.871, which is higher than other models without post-processing. Its F1-score is also impressive at 0.906, indicating a ll-balanced precision and recall. After post-processing, the OA significantly increases to 0.969, showcasing the model's compatibility with post-processing techniques. This model has high OA and F1-score without post-processing, reflecting the model's ability to accurately extract roads, and excellent overall performance, even compared to other models after post-processing. Although high, the Recall and Precision could still be improved, as seen from the post-processed results.

When comparing the models, observe that post-processing consistently improves their performance, enhancing OA and Kappa coefficients across the board. The HybridSN model stands out with the highest OA and F1-score, indicating its superiority in road extraction. However, it's important to note that post-processing benefits all models, potentially leveling the playing field to some extent. Each model has its strengths and aknesses. The Shallow CNN and Deep CNN demonstrate respectable performance but may struggle with capturing fine road details. The Dilated CNN performs moderately ll but exhibits room for improvement in Recall. The HybridSN outperforms others but can still be optimized for better Recall and Precision.

Table 2. Performance assessment of different models for road extraction.

Model	Post-processing	OA	Kappa	Recall	Precision	F1-score	IoU
Shallow CNN	No	0.864	0.713	0.633	0.864	0.730	0.640
	Yes	0.960	0.843	0.965	0.763	0.852	0.763
Deep CNN	No	0.861	0.723	0.659	0.853	0.743	0.650
	Yes	0.954	0.824	0.968	0.740	0.838	0.740
Dilated CNN	No	0.848	0.720	0.602	0.812	0.691	0.655
	Yes	0.937	0.770	0.952	0.675	0.789	0.675
HybridSN	No	0.871	0.813	0.803	0.906	0.851	0.834
	Yes	0.969	0.875	0.982	0.806	0.885	0.806

Training Time

The training time per epoch provides insights into the computational complexity and efficiency of the different models (Table 3). The Shallow CNN has the fastest training time at 5.3 seconds per epoch. This is expected as it has the simplest architecture with fewer layers compared to the other models. The low training time makes it more suitable for quick prototyping and iteration during development. The Deep CNN takes 15 seconds, nearly 3x longer than the Shallow CNN. The increased depth and parameters lead to higher computational costs for the forward and backward passes during training. The Dilated CNN requires 17 seconds per epoch. The dilated convolutions increase the receptive field for aggregating broader context but add computational overhead. Finally, the HybridSN model takes the longest at 20 seconds per epoch. Jointly processing the spectral and spatial data in 3D convolutional layers adds substantial computation. The spectral modeling also increases the input feature dimensionality compared to 2D spatial-only CNNs.Overall, the training times reflect model complexity in terms of depth, parameters, and input/feature dimensionality. There are clear tradeoffs between model performance and efficiency. The RTX 2060 GPU provides sufficient computing power to train the models in a reasonable time frame. The training times help guide appropriate model selection based on the application constraints and priorities.

Table 3. Training time of different models for road extraction.

Model	Training Time (seconds/epoch)
Shallow CNN	5.3
Deep CNN	15
Dilated CNN	17
HybridSN	20

Discussions

The Importance of New Ideas in Study

Previous studies have extracted roads that contain constant shapes and measurements or parameters, while this study has extracted twisted and variable methods of measurement, which have been somewhat difficult. The results demonstrate that the proposed model was more efficient for road extraction. Additionally, data augmentation procedures were employed to effectively increase the size of the dataset. An encoder–decoder SegNet model was used for the generative part to generate a high-resolution segmentation map[16]. The accuracy that they achieved for recall, precision, and F1 score was 91.01%, 88.31%, and 89.63%, respectively, which shows the superiority of the proposed model for road extraction.

Factors Affecting Model Performance

Several key factors that significantly influence the performance of road extraction models in this challenging task have emerged from the comparative evaluation. Firstly, the size of the receptive field plays a crucial role. The dilated CNN, with its larger receptive field, struggled with precision, as it tended to mistake background regions for roads. This highlights the importance of incorporating more localized context modeling to focus on road-specific spatial patterns and textures effectively. Secondly, the HybridSN model, which leveraged both spectral and spatial cues, achieved superior accuracy and localization compared to CNNs that only utilized spatial context. This demonstrates the importance of joint spectral-spatial modeling and multi-modal data fusion for enhancing road extraction results. Thirdly, the capacity of the model also impacts its performance. The Deep CNN, with its greater capacity and ability to learn deeper hierarchical feature representations, outperformed the Shallow CNN. This emphasizes the benefits of having sufficient model capacity to capture intricate road features. However, it is essential to note that merely increasing depth alone did not fully address the underlying challenges of road extraction. Fourthly, all CNN models faced difficulties in producing continuous and connected road predictions, indicating a lack of explicit topological modeling in their architectures. Although post-processing techniques helped improve connectivity, they came at the cost of reduced precision. Therefore, there is a need to incorporate topological modeling within the architectures to address fragmented road predictions more effectively. Finally, end-to-end learning proved advantageous in the case of the HybridSN model, as it could directly predict roads from pixels without heavily relying on post-processing corrections. This highlights the importance of incorporating more inductive bias into models to build topological and contextual awareness directly into their architecture.

Challenges in Road Infrastructure Detection

Road extraction from overhead imagery poses numerous complex challenges that push the boundaries of remote sensing capabilities. One of the primary hurdles is dealing with the diverse and intricate appearance of roads. They come in various materials, textures, widths, orientations, and topological structures, making it essential for models to learn robust and expressive features that can effectively capture this wide variability and generalize ll. Another significant challenge is occlusions caused by various objects like trees, buildings, and vehicles, which frequently obscure parts of the roads. These masking and incomplete observations hinder the detection process and require the modeling of longer-range dependencies to enable accurate road detection even in obstructed areas. Furthermore, the presence of small targets, such as narrow roads, lanes, and sidewalks, adds complexity to the extraction task. Given the limited number of pixels occupied by these narrow road segments, precise localization at higher resolutions is crucial. Models must be equipped to accurately identify and delineate these small road features. Background confusion is yet another obstacle in road extraction. Many background classes, such as parking lots, buildings, and trails, share visual cues similar to roads, leading to false positives in the results. It is essential for the models to perform discriminative learning of fine-grained differences to avoid misclassifications between roads and visually similar background elements. Moreover, the topological relations between different road segments pose a challenging problem. Roads form an interconnected network with dependencies and constraints between various parts. Effectively modeling these complex topological relationships remains a daunting task for road extraction algorithms. The broader scene context, including terrain, land use, and socioeconomic factors, significantly influences road patterns and can aid in road detection. However, incorporating this context effectively into models is a complex and ongoing research problem that requires innovative approaches. A key obstacle to developing accurate road extraction models lies in the availability of suitable training data. Supervised learning for road extraction demands pixel-level annotated road data, which is both labor-intensive and time-consuming to collect and annotate. Despite available datasets, there may still be a lack of diversity in road types and conditions, which can limit the model's ability to generalize to various scenarios.

Practical Applications of the Developed Models

The models' applications in road extraction from overhead imagery are diverse and have the potential to revolutionize various domains. One significant application is Transportation Planning, where precise road network maps play a crucial role in analyzing transportation connectivity. These maps help identify gaps, redundancies, and opportunities for new infrastructure development, providing valuable support for urban planning initiatives. Navigation applications also benefit greatly from up-to-date road maps generated by these models. In-car GPS, ride-sharing platforms, and autonomous vehicles can use these maps to plan optimal legal routes, leading to improved routing algorithms and more efficient navigation experiences for users. During Emergency Response scenarios, the rapid mapping of roads using aerial imagery becomes critical. The models enable the quick coordination of evacuations, deliveries, and deployment of responders by optimizing travel routes in disaster-affected areas, aiding in timely and effective emergency response efforts. Population Mapping is another valuable application where road layouts provide insights into population distribution and land use patterns. This information can be used to target infrastructure and services effectively, leading to better resource allocation and development planning. For Development Monitoring, the ability to repeatedly map roads over time becomes essential. These models can detect newly constructed roads in rapidly growing areas, facilitating the tracking and monitoring of development progress. In Military Operations, analyzing road networks becomes vital for armed forces. It provides key logistical intelligence, helping in the coordination of movements for personnel and supplies, ultimately enhancing military operations' efficiency and effectiveness. Additionally, Automated Mapping using these models allows for accurate and scalable road mapping. This approach efficiently leverages large volumes of aerial data, ensuring that maps remain up-to-date and reflective of real-world road infrastructure changes.

Potential Improvements in Urban Planning and Development

The applications of advanced AI techniques for road extraction offer valuable insights and data for evidence-based infrastructure planning. One notable benefit lies in Informed Infrastructure Planning, where precisely mapped road networks aid in identifying neighborhoods lacking connectivity to essential destinations like employment centers. This information can guide targeted infrastructure investments and the formulation of policies that promote equitable growth, ensuring better accessibility for all residents. Moreover, detailed road information allows for optimized land use analyses, helping to understand transportation capacity and travel patterns. This, in turn, informs complementary land use planning, such as densification along transit corridors and the promotion of mixed-use development, leading to more efficient and sustainable urban development. The automated mapping of new and informal roads in rapidly growing regions contributes to Cost-Effective Expansion strategies. By identifying areas that require upgrades and expansions, urban planners can maximize the returns on infrastructure investments, making the most of limited resources in developing areas. Sustainable Development initiatives can benefit from monitoring long-term road network changes. By evaluating the impacts of sustainable transportation policies like transit-oriented development, urban sprawl can be curbed, leading to more environmentally friendly and resilient cities. Resilience Planning is another crucial application re updating road maps before and after disasters prove invaluable. This facilitates damage assessments and helps in planning redevelopment strategies, enhancing the resilience of critical transportation networks, and improving disaster response.

The automation of road extraction using deep learning analysis of regularly collected aerial imagery brings the advantage of Scaling. Frequent large-scale assessments of road infrastructure become possible, enabling dynamic, data-driven urban planning and development processes.

Conclusion

This study presented a comparative evaluation of deep learning models for extracting road infrastructure from aerial imagery. The evaluated models included CNNs, dilated CNNs, and a hybrid spectral-spatial network. Experiments demonstrated that the hybrid network integrating both spectral and spatial processing achieved the highest accuracy and intersection-over-union score of 96.9% and 80.6% respectively after post-processing. The joint modeling of spectral and spatial cues enabled precise localization and delineation of roads. While all models benefited from topological improvements of post-processing, they struggled with false positives and disconnected predictions indicating limitations in contextual modeling. The results highlight the importance of multi-modal data fusion and encoding domain knowledge of road topology into network architectures.

This research makes several significant contributions to the field of road extraction from overhead imagery. Firstly, it demonstrates the superiority of joint spectral-spatial modeling over spatial-only methods for achieving accurate road extraction. By leveraging both spectral and spatial cues, the proposed approach outperforms existing methods and highlights the importance of incorporating multi-modal data fusion for improved results. Secondly, the research provides valuable quantitative benchmarking of different deep-learning architectures specifically tailored for road detection. This comparative analysis sheds light on the strengths and aknesses of each model, enabling researchers and practitioners to make informed choices based on the specific requirements of their applications. Furthermore, the study thoroughly investigates the relative impacts of various factors on road extraction performance. These factors include receptive field sizes, model capacities, and topological constraints. By analyzing their effects, the research offers insights into how to design and fine-tune deep learning models for optimal road extraction results. Lastly, the research provides compelling evidence for the importance of end-to-end learning that integrates multi-modal cues and domain knowledge directly into the models. This approach proves superior to relying solely on post-processing techniques for refining predictions. By incorporating inductive biases and contextual awareness during the learning process, the proposed method achieves better accuracy and efficiency in road extraction tasks.

Many areas might be revolutionized by the models' broad uses in road extraction from aerial photos. Accurate road network maps are essential for assessing transportation connections in one important use, which is transportation planning. Urban planning activities benefit greatly from the identification of gaps, redundancies, and potential for future infrastructure development that these maps serve to provide.

Future work should explore incorporating topological priors and constraints directly into network architectures to improve connectivity modeling. Hybrid spatial-spectral networks could also be enhanced to learn scale-adaptive features across multiple input resolutions. akly supervised and semi-supervised learning paradigms provide opportunities to improve generalizability and relax annotation requirements. Architectural search could help automate learning specialized designs tailored for road extraction tasks. More powerful deep learning frameworks offer abundant opportunities to advance the state-of-the-art in automated mapping of transportation infrastructure from overhead imagery.

Authors' Declaration

- Conflicts of Interest: None.
- We hereby confirm that all the Figures and Tables in the manuscript are ours. Furthermore, any Figures and images, that are not ours, have been included with the necessary permission for re-publication, which is attached to the manuscript.

- No human studies are present in the manuscript.
- The author has signed an animal welfare statement.
- Ethical Clearance: The project was approved by the local ethical committee at Northern Technical University.

Authors' Contribution Statement

M. I. A. was responsible for data collection from the study site, data analysis, modeling, and interpretation. He also played a key role in writing, drafting, revising, and editing the research paper. Additionally, he secured the necessary funding to complete this research.

M. A. Sh. and A. A. Al. collaborated in generating the core research idea, comprehending its intricacies, and designing the study. Together, they played instrumental roles in developing the research methodology, offering invaluable insights, and providing critical comments and critiques that enhanced the overall quality of the paper. Throughout the research process, both authors actively supervised and guided the study, ensuring its rigorous execution. Finally, they jointly reviewed and approved the final version of the paper.

References

1. Jassim OA, Abed MJ, Saied ZH. Indoor/Outdoor Deep Learning Based Image Classification for Object Recognition Applications. Baghdad Sic J. 2023 Dec 5; 20(6 (Suppl.)): 2540. https://orcid.org/0009-0004-1607-2778
2. Abd Alsammed SM. Advanced GIS-based multi-function support system for identifying the best route. Baghdad Sic J. 2022 Jun 1; 19(3): 0631. https://doi.org/10.21123/bsj.2022.19.3.0631
3. Kadhim MA, Abed MH. Convolutional Neural Network for Satellite Image Classification. Intelligent Information and Database Systems: Recent Developments. 2020; 11: 165-178. https://doi.org/10.1007/978-3-030-14132-5_13
4. Zeng Y, Guo Y, Li J. Recognition and extraction of high-resolution satellite remote sensing image buildings based on deep learning. Neural Comput Appl. Feb 2022; 34(4): 2691-2706. https://doi.org/10.1007/s00521-021-06027-1
5. Shareef MA, Toumi A, Khenchaf A. Estimation of water quality parameters using the regression model with fuzzy k-means clustering. Int J Adv Comput Sci Appl. 2014; 5(6) : 151-157. https://doi.org/10.14569/IJACSA.2014.050624
6. Keshk H, Yin X. Classification of EgyptSat-1 Images Using Deep Learning Methods. Int J Sens Wirel Commun Control. Feb 2020; 10: 37-46. https://doi.org/10.2174/22103279096661902071538 58
7. Yu Y, Gong Z, Zhong P. An Unsupervised Convolutional Feature Fusion Network for Deep Representation of Remote Sensing Images. IEEE Trans Geosci Remote Sens. Dec 2017; 15: 23-27. https://doi.org/10.1109/LGRS.2017.2767626
8. Shareef MA, Ameen MH, Ajaj QM. Change detection and GIS-based fuzzy AHP to evaluate the degradation and reclamation land of Tikrit City Iraq. Geodesy Cartogr. Dec 2020; 46(4): 194-203. https://doi.org/10.3846/gac.2020.11616
9. Dai J, Du Y, Zhu T, Wang Y, Gao L. Multiscale Residual Convolution Neural Network and Sector Descriptor-Based Road Detection Method. IEEE Access. Nov 2019; 7: 173377-173392. https://doi.org/10.1109/ACCESS.2019.2956725
10. Wei Y, Zhang K, Ji S. Simultaneous Road Surface and Centerline Extraction From Large-Scale Remote Sensing Images Using CNN-Based Segmentation and Tracing. IEEE Trans Geosci Remote Sens. May 2020; 58(12): 8919-8931. https://doi.org/10.1109/TGRS.2020.2991733
11. Lu X, Zhong Y, Zheng Z, Liu Y, Zhao J, Ma A, et al. Multi-Scale and Multi-Task Deep Learning Framework for Automatic Road Extraction. IEEE Trans Geosci Remote Sens. Aug 2019; 57(11): 9362-9377. https://doi.org/10.1109/TGRS.2019.2926397
12. Han X, Lu J, Zhao C, You S, Li H. Semisupervised and akly Supervised Road Detection Based on Generative Adversarial Networks. IEEE Signal Process. Lett. Feb. 2018; 25(4): 551-555. https://doi.org/10.1109/LSP.2018.2809685
13. Li X, Wang Y, Zhang L, Liu S, Mei J, Li Y. Topology-Enhanced Urban Road Extraction via a Geographic Feature-Enhanced Network. IEEE Trans Geosci Remote Sens.May 2020; 58(12): 8819-8830. https://doi.org/10.1109/TGRS.2020.2991006
14. Sofla R, Alipour-Fard T, Arefi H. Road extraction from satellite and aerial image using SE-Unet. J Appl Remote Sens. . Jan 2021; 15(1): 014512 – 014512. https://doi.org/10.1117/1.JRS.15.014512

15. Fan R, Bocus MJ, Zhu Y, Jiao J, Wang L, Ma F, Cheng S, Liu M. Road crack detection using deep convolutional neural network and adaptive thresholding. In2019 IEEE Intelligent Vehicles Symposium (IV) 2019 Jun 9 (pp. 474-479). https://doi.org/10.1109/IVS.2019.8814000

16. Shi Q, Liu X, Li X. Road detection from remote sensing images by generative adversarial networks. IEEE access. 2017 Nov 13; 6: 25486-94.https://doi.org/10.1109/ACCESS.2017.2773142

التعلم العميق (CNN) للكشف عن البنية التحتية للطرق في مدينة الموصل القديمة باستخدام الصور الجوية عالية الدقة

مصطفى عصمت عبدالرحمن[1]، منتظر عيدي شريف[1]، علياء عباس العطار[2]

[1]قسم تقنيات هندسة المساحة، الكلية التقنية الهندسية، الجامعة التقنية الشمالية، كركوك، العراق.
[2]الجامعة التقنية الشمالية، الموصل، العراق.

الخلاصة

يعد رسم خرائط دقيقة للبنية التحتية للطرق من الصور الجوية أمرًا بالغ الأهمية لمختلف التطبيقات ولكنه يطرح تحديات عديدة بسبب تعقيد أنماط الطرق في الواقع الحقيقي. يبحث هذا البحث في تقنيات التعلم العميق لاستخراج الطرق بشكل الي من البيانات العامة. يتم تقييم العديد من بنى الشبكات العصبية التلافيفية (CNN) بما في ذلك الشبكة الطيفية المكانية الهجينة (HybridSN) التي تجمع بين للصور البصرية وبيانات الليدار. يتم تقييم النماذج على مجموعة بيانات من الصور الجوية الحضرية باستخدام علامات حقيقة أرضية مشتقة من تقنية الليدار. ويحقق HybridSN الذي يدمج كلاً من المعالجة الطيفية والمكانية أعلى أداء بدقة إجمالية تبلغ 96.9% و80.6% من التقاطع بعد المعالجة اللاحقة. تتيح النمذجة المشتركة للإشارات متعددة الوسائط إمكانية تحديد أجزاء الطريق وتحديدها بدقة عالية. وبالمقارنة، فإن شبكات CNN التي تستفيد من السياق المكاني وحده تؤدي أداءً أسوأ مع أفضل دقة إجمالية تبلغ 95.4% بعد المعالجة اللاحقة. تظهر جميع النماذج أوجه قصور في استخراج شبكات الطرق المتماسكة. وهذا يدل على أهمية دمج البيانات الطيفية والمكانية ضمن أطر التعلم العميق لاستخراج الطرق. تسلط النتائج الضوء على الفرص المتاحة لتطوير أحدث الخرائط من خلال أجهزة الاستشعار الهجينة وتصميم بنى عصبية ذات وعي طوبولوجي أقوى. يقوم بالتحليل الآلي للبيانات الجوية متعددة الوسائط مع التعلم العميق والحفاظ على كفاءة قوائم جرد حديثة للبنية التحتية الحيوية للنقل على نطاق المدن.

الكلمات المفتاحية: CNN، التعلم العميق، كشف الطرق، البنية التحتية للطرق، مدينة الموصل القديمة.