Journal of Information Systems Engineering and Management
2025, 10(46s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Extraction of Road from Super-Resolved Satellite Imagery using Semantic Segmentation

Nisha Shamsudin (Researcher)[1*], Bindu V R (Professor)[2]
[1]*School of Computer Sciences, Mahatma Gandhi University, Kottayam, Kerala*
[2]*School of Computer Sciences, Mahatma Gandhi University, Kottayam, Kerala*

ARTICLE INFO

Received: 29 Dec 2024

Revised: 15 Feb 2025

Accepted: 24 Feb 2025

ABSTRACT

Accurate road extraction from satellite imagery is essential for urban planning, disaster management, and navigation. Our research uses the super-resolved Massachusetts roads dataset to develop an automated road extraction framework. We employ the enhanced U-Net model for semantic segmentation, incorporating preprocessing techniques such as normalization and resizing to improve input quality. The U-Net architecture effectively captures spatial and contextual information, ensuring accurate road delineation even in complex environments. Post-processing techniques refine predictions by reducing noise and improving connectivity in the extracted road networks. Our experiments demonstrate the robustness of enhanced U-Net on super-resolved data, achieving high accuracy across urban, suburban, and rural areas. This study advances satellite image analysis by emphasizing the role of super-resolution in improving geospatial predictions and enhancing road extraction accuracy.

Keywords: Semantic segmentation, Road network extraction, Super-resolved satellite imagery, U-Net architecture, Massachusetts roads dataset.

INTRODUCTION

The extraction of road networks from satellite imagery is a fundamental task in geospatial analysis, with significant applications in urban planning, disaster management, navigation systems, and infrastructure monitoring [1,13,14,15]. As global urbanization expands, the need for accurate, up-to-date, and scalable road network data has become increasingly critical [2,27]. Satellite imagery offers a valuable resource with its wide coverage and periodic updates. The automated extraction of roads from satellite imagery has always been challenging due to the inherent complexity of landscapes, varying road widths, occlusions caused by vegetation or shadows, and heterogeneous structures [3,19,20,21]. Despite these advancements, several challenges persist. One major limitation is the reliance on low-resolution satellite imagery, which often lacks the necessary detail for accurate road extraction, particularly in complex environments [16,17,18]. Super-resolution techniques, which enhance the spatial resolution of images, offer a promising solution to this problem [4,22,23,24]. By generating high-resolution imagery from low-resolution inputs, these methods can improve the visibility of fine road features, thereby enhancing the performance of deep learning models [5,25,27]. The integration of super-resolution with deep learning for road extraction remains underexplored, and standardized methodologies for preprocessing and postprocessing super-resolved data are lacking. This paper presents a comprehensive deep learning-based approach to road extraction that integrates super-resolution techniques with the U-Net architecture. The proposed methodology includes standardized preprocessing steps, such as image normalization and resizing, to enhance the quality of super-resolved input data and ensure optimal compatibility with the U-Net model. The U-Net architecture is optimized to capture spatial and contextual information in complex environments. Post-processing techniques are applied to refine the model's predictions, improving connectivity and reducing noise in the extracted road networks. The **contributions** of this research are as follows:

- To address the challenge of inconsistent road extraction accuracy in diverse environments caused by low-resolu tion inputs and inadequate feature representation, we propose a novel model by integrating the ESPCN super-r esolution technique with the U-Net architecture, enhancing input clarity and spatial feature extraction to achiev

Journal of Information Systems Engineering and Management
2025, 10(46s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

e robust, high-precision road detection across varied terrains

- Inconsistent preprocessing of super-resolved data and arbitrary post-processing of predictions hinder compatib ility with deep learning frameworks and degrade road network quality, prompting our development of standard ized pipelines to unify data alignment and refine outputs, ensuring robustness and coherence in extracted road networks
- Traditional road network mapping approaches often struggle with computational inefficiency and scalability li mitations, which our framework addresses by integrating advanced image processing techniques with standard ized datasets and architectures
- The underutilization of super-resolution in geospatial prediction models restricts the accuracy of extracted road networks, a gap mitigated by our research through enhanced satellite image analysis pipelines that use super-re solution to deliver robust, high-precision frameworks for diverse environmental conditions

Through these contributions, this research addresses current challenges in road extraction and provides a foundation for future research in automating geospatial tasks for real-world applications. The rest of the paper was shown as follows. The methodology is elaborated in Section 2. Section 3 depicts the experimental analysis, results, and comments, while Section 4 provides the discussion and conclusion.

METHODS

The section outlines the systematic approach developed to address the challenges of automated road extraction from satellite imagery, integrating the super-resolution technique ESPCN with the U-Net architecture for enhanced performance. Due to the problems in existing methods, we need to combine SR with segmentation models, standardize preprocessing/postprocessing pipelines, and improve generalization across diverse environments. This research proposes a unified framework structured into data preprocessing, including super-resolution enhancement (comparison of different SR models like SRCNN, SRGAN, and ESPCN, of which ESPCN outperforms) and normalization of the Massachusetts Roads Dataset, model architecture design, using a modified U-Net optimized for spatial and contextual feature extraction, and post-processing employing morphological operations to refine connectivity and reduce noise in extracted road networks [8,9,10].

The Efficient Sub-Pixel Convolutional Neural Network (ESPCN) is a state-of-the-art super-resolution model designed to efficiently enhance the resolution of low-quality images by performing upscaling in the low-resolution space and reconstructing the high-resolution image using a sub-pixel convolution layer, and is illustrated in Algorithm 1. This approach significantly reduces computational complexity while maintaining high-quality results. The core innovation of ESPCN lies in its sub-pixel convolution layer, which rearranges the channels of low-resolution feature maps into the spatial dimensions of the high-resolution output, enabling efficient upscaling without computationally expensive operations. ESPCN is computationally efficient, as most processing occurs in the low-resolution space, reducing memory and processing requirements, and it is trained to predict super-resolved images. Its superior performance in terms of speed and accuracy compared to models like SRCNN and SRGAN [11,12,26] makes it an ideal choice for preprocessing satellite imagery before segmentation. In our research, ESPCN enhances the resolution of satellite images from the Massachusetts Roads Dataset, providing clearer and more detailed inputs for the U-Net segmentation model, thereby enabling more accurate road network extraction.

Algorithm 1: Enhanced ESPCN

Input: Low-resolution image I_{LR} of size $H \times W \times C$, Upscaling factor r

Output: High-resolution image I_{HR} of size $(rH) \times (rW) \times C$.

1. Pass I_{LR} through convolutional layers to extract feature maps.
2. Feature maps: $F = f(I_{LR})$, where F has dimensions $H \times W \times D$.
3. Apply a non-linear activation function (e.g., ReLU) to feature maps.
4. Expand feature maps to $D = r^2 \times C$ channels.
5. Rearrange channels into spatial dimensions $(rH) \times (rW) \times C$ using pixel-shuffling.
6. Combine the rearranged feature maps to generate IHR.
7. Compute loss using Mean Squared Error (MSE) or L1 loss:

Journal of Information Systems Engineering and Management
2025, 10(46s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

$$L = \frac{1}{N}\sum(I_{HR} - I_{GT})2 \quad \textbf{(1)}$$

8. Update model parameters using backpropagation and gradient descent.
9. Return the final high-resolution image.

Algorithm 2 was designed and implemented for road segmentation using the Massachusetts Roads Dataset. The architecture follows a fully convolutional encoder-decoder design with skip connections, allowing for precise segmentation by preserving spatial information.

Algorithm 2: Road Extraction Using Super-Resolution and U-Net Segmentation

Input: Satellite image I from the Massachusetts Roads Dataset

Output: Predicted road segmentation mask $\widehat{M}$

1. Extract low-resolution image I_{LR} by down-sampling the original image I
2. Normalize images:

$$I_{LR} = \frac{I_{LR}}{255},\ I_{HR} = \frac{I}{255} \quad \textbf{(2)}$$

3. Pair low-resolution images I_{LR} with corresponding high-resolution ground truth images I_{HR} for the SR model ESPCN training. Train the ESPCN model M_{SR} to enhance low-resolution images:

$$I_{SR} = M_{SR}(I_{LR}) \quad \textbf{(3)}$$

4. Minimize ESPCN model loss function:

$$£_{SR} = £_{MSE}\,(I_{HR}, I_{SR}) + \lambda £_{Perceptual}\,(I_{HR}, I_{SR}) \quad \textbf{(4)}$$

5. Apply the ESPCN model to generate an enhanced high-resolution image I_{SR}.
6. Pass the I_{SR} through the U-Net segmentation model $M_{U\text{-}Net}$.
7. Minimize binary cross-entropy:

$$£_{BCE} = -\frac{1}{N}\sum_{i=1}^{N} [y_i \log \hat{y}_i + (1 - y_i)\log(1 - \hat{y}_i)] \quad \textbf{(5)}$$

8. Output the predicted road segmentation mask $\widehat{M}$.
9. Threshold predicted mask $\widehat{M}$ at 0.5 for the binary mask.
10. Apply morphological operations (closing, opening) to refine the mask.
11. Use graph-based algorithms to connect fragmented road segments.
12. Calculate performance metrics: IoU, Dice Coefficient, and pixel accuracy.

The predicted mask is then refined using post-processing techniques, including binarization, Gaussian blur, and morphological operations like closing and opening to remove noise and fill gaps. Edge detection with the Canny algorithm and contour filling further enhances the segmentation quality, ensuring continuous road structures.

RESULTS

To evaluate the performance of our U-Net-based super-resolved road segmentation model, we conducted extensive testing using the Massachusetts Roads Dataset. The Massachusetts Roads Dataset consists of 1171 aerial images of the state of Massachusetts [7]. Each image is 1500×1500 pixels, covering an area of 2.25 square kilometers. The enhanced U-Net trained model was applied to a set of test images, and segmentation performance was measured using two key metrics: the Intersection over Union (IoU) and the Dice coefficient

$$IOU = \frac{|Prediction \cup Ground\ Truth|}{|Prediction \cap Ground\ Truth|} \quad DICE = \frac{2|Prediction \cap Ground\ Truth|}{|Prediction| + |Ground\ Truth|} \quad \textbf{(6)}$$

- ∩ denotes the intersection (overlapping pixels between predicted and ground truth masks)
- ∪ denotes the union (total number of pixels classified as roads in either the prediction or ground truth).

The model was then evaluated on a separate test dataset, where segmentation performance was assessed using test loss and accuracy. The evaluation was conducted using the following formula for accuracy:

Journal of Information Systems Engineering and Management
2025, 10(46s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

$$\text{Accuracy} = \frac{\text{Number of Correctly Classified Pixels}}{\text{Total Number of Pixels}} \quad \textbf{(7)}$$

Table 1 summarizes the performance metrics obtained from the test dataset, achieving better accuracy and segmenting roads from satellite imagery.

Table 1. Quantitative evaluation metrics for an enhanced U-Net Model with super-resolved dataset

Metric	Value
Average Dice Coefficient	0.87
Average IoU	0.78
Test Accuracy	0.9612
Test MSE	0.0274

Table 2 summarizes the performance of super-resolution models, highlighting ESPCN's effectiveness. The results indicate that ESPCN achieves the highest Peak signal to noise ratio (PSNR) and Structural similarity index measure (SSIM) while maintaining the lowest Mean square error (MSE), confirming its superiority in reconstructing high-quality images from low-resolution inputs.

Table 2. Super resolution model comparison for the Massachusetts dataset

Model	PSNR	SSIM	MSE
SRCNN	28.45	0.912	0.0023
SRGAN	30.21	0.940	0.0018
ESPCN	41.69	0.9969	0.0005

To further analyze the performance of the model, we visualized the segmentation results with post-processing techniques like,

Binarization (Thresholding): The predicted mask from the U-Net model contains probability values between 0 and 1. To convert it into a binary mask (0 or 1), thresholding is applied:

$$M_b(x,y) = \begin{cases} 1, if\ Mp(x,y) > T \\ 0, otherwise \end{cases} \quad \textbf{(8)}$$

Gaussian Blurring: To smooth the edges and reduce noise, a Gaussian blur is applied, which convolves the image with a Gaussian kernel:

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{-x^2+y^2} / 2\sigma^2 \quad \textbf{(9)}$$

where:

• G(x, y) is the Gaussian kernel,

• σ is the standard deviation controlling blur intensity.

Morphological Operations: Morphological operations help refine the segmented regions by eliminating small noise and filling gaps.

Edge Detection using Canny Algorithm: The Canny edge detector extracts edges from the refined mask. The gradient magnitude G is computed as:

$$G = \sqrt{G_x^2 + G}y^2 \quad \textbf{(10)}$$

where:

Journal of Information Systems Engineering and Management
2025, 10(46s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

• Gx and Gy are the gradients in the x and y directions, computed using Sobel filters.

Contour Detection and Filling: Contours are identified using connected edge pixels and then filled to make the segmented regions more continuous:

$$M_{filled} = \sum_{i=1}^{N} C_i] \quad \textbf{(11)}$$

Where: C_i represents individual contours detected in the image.

After these post-processing steps, the final segmentation mask is refined, with smoother edges, fewer noise artifacts, and well-connected road structures. These steps significantly improve the model's accuracy in road segmentation tasks, and Figure 1 presents sample test images alongside their corresponding ground truth masks and predicted masks. The predicted masks closely resemble the ground truth masks, indicating effective road segmentation, but in some cases, the model struggles with faint or narrow roads, leading to slight misclassifications. The graphs in Figure 2 illustrate the training progress of the model over 50 epochs, displaying accuracy and loss trends.

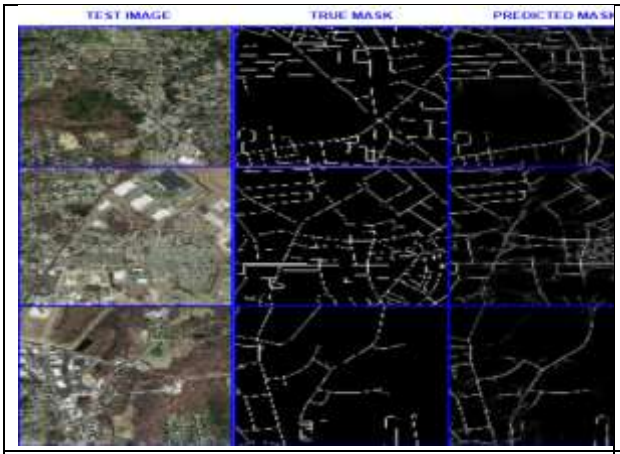

Fig. 1. Comparison of Input Images, Ground Truth Masks, and Predicted Masks with super-resolved Massachusetts road dataset (applied enhanced U-Net segmentation Model)

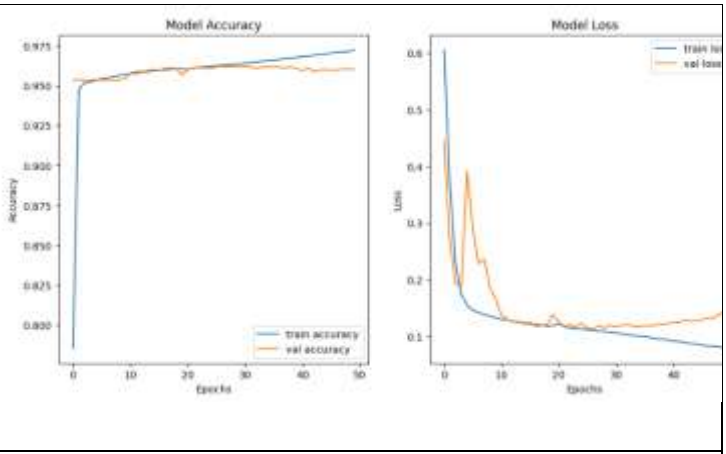

Fig. 2. Training and validation accuracy and loss curves. Accuracy trends in the left plot and loss variations during training and validation in the right plot.

DISCUSSION

This research demonstrated the effectiveness of U-Net for road segmentation using the Massachusetts Roads Dataset, producing high-quality segmentation masks with strong quantitative and qualitative performance. The merging of the super-resolution model, ESPCN, further enhanced the process by improving the resolution of low-quality satellite imagery, providing clearer and more detailed inputs for segmentation. The U-Net architecture achieved robust performance, with an IoU of 0.78 and a Dice coefficient of 0.87, highlighting its capability for accurate road detection. The challenges remain in segmenting smaller roads, particularly in areas with weak contrast between the road and the background. Future improvements could focus on refining the architecture, integrating additional post-processing techniques, and incorporating advanced methods such as data augmentation, multi-scale feature fusion, and specialized loss functions like Dice loss to handle complex road structures more effectively. Overall, the combined approach of super-resolution and segmentation offers a promising solution for robust and accurate road network extraction from satellite imagery, with applications in urban planning, navigation, and geographic information systems.

REFRENCES

[1] Amirgan, B. and Erener, A. (2024). Semantic segmentation of satellite images with different building types using deep learning methods. Remote Sensing Applications: Society and Environment, 34:101176.

[2] Kim, B., An, E.-J., Kim, S., Sri Preethaa, K., Lee, D.-E., and Lukacs, R. (2024). Srgan-enhanced unsafe operation detection and classification of heavy construction machinery using cascade learning. Artificial Intelligence Review, 57(8).
[3] Guo, J., Hong, D., Liu, Z., and Zhu, X. X. (2024). Continent-wide urban tree canopy fine-scale mapping and coverage assessment in south america with high-resolution satellite images. ISPRS Journal of Photogrammetry and Remote Sensing, 212:251–273.
[4] Maiseli, B. and Abdalla, A. T. (2024). Seven decades of image super-resolution: achievements, challenges, and opportunities. Eurasip Journal on Advances in Signal Processing, 2024(1).
[5] Karwowska, K. and Wierzbicki, D. (2023). Mcwesrgan: Improving enhanced super-resolution generative adversarial network for satellite images. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing.
[6] Guo, J., Lv, F., Shen, J., Liu, J., and Wang, M. (2023). An improved generative adversarial network for remote sensing image super-resolution. IET Image Processing, 17(6):1852 – 1863.
[7] Mnih, V. (2013). Machine learning for aerial image labeling (Doctoral dissertation, University of Toronto)
[8] Arun, P. V., Buddhiraju, K. M., Porwal, A., and Chanussot, J. (2020). Cnn-based super-resolution of hyperspectral images. IEEE Transactions on Geoscience and Remote Sensing.
[9] Chu, M., Xie, Y., Mayer, J., Leal-Taix´e, L., and Thuerey, N. (2020). Learning temporal coherence via self-supervision for gan-based video generation. ACM Trans. Graph., 39(4).
[10] Diakogiannis, F. I., Waldner, F., Caccetta, P., and Wu, C. (2020). Resunet-a: A deep learning framework for semantic segmentation of remotely sensed data. ISPRS Journal of Photogrammetry and Remote Sensing, 162:94–114.
[11] Dong, C., Loy, C. C., He, K., and Tang, X. (2016). Image super-resolution using deep convolutional networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 38(2):295–307.
[12] Ledig, C., Theis, L., Huszar, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., and Shi, W. (2017). Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 105–114, Honolulu, HI, USA.
[13] Malczewska, A. and Wielgosz, M. (2024). How does super-resolution for satellite imagery affect different types of land cover? sentinel-2 case. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 17:340–363.
[14] Min, J., Lee, Y., Kim, D., and Yoo, J. (2024). Bridging the domain gap: A simple domain matching method for reference-based image super-resolution in remote sensing. IEEE Geoscience and Remote Sensing Letters, 21:1–5.
[15] Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N. Y., Kainz, B., Glocker, B., and Rueckert, D. (2018). Attention u-net: Learning where to look for the pancreas. In Medical Image Computing and Computer-Assisted Intervention (MICCAI), pages 605–612. Springer.
[16] Pan, Y., Tang, J., and Tjahjadi, T. (2024). Lpsrgan: Generative adversarial networks for super-resolution of license plate image. Neurocomputing, 580.
[17] Pandey, B. and Pandey, M. S. (2024). Enhanced satellite image classification using deep convolutional neural network. In 2024 International Conference on Communication, Computer Sciences and Engineering (IC3SE), pages 1–6.
[18] Rasheed, M. P. A., Rajeev, R., Shamsudin, N., Haroon, R. P., Al Mansoori, S., and Panthakkan, A. (2023). Enhancing khalifa satellite imagery resolution with ai-powered super resolution generative adversarial networks (srgan). In 2023 International Conference on Innovations in Engineering and Technology (ICIET), pages 1–4.
[19] Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In Medical Image Computing and Computer-Assisted Intervention (MICCAI), pages 234–241. Springer.
[20] Shi, W., Caballero, J., Husz´ar, F., Totz, J., Aitken, A. P., Bishop, R., Rueckert, D., and Wang, Z. (2016). Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 1874–1883.

Research Article

[21] Sreelakshmi, S. and Vinod Chandra, S. (2024). Visual saliency-based landslide identification using super-resolution remote sensing data. Results in Engineering, 21:101656.
[22] Wang, C., Zhang, X., Yang, W., Li, X., Lu, B., and Wang, J. (2023a). Msagan: A new super-resolution algorithm for multispectral remote sensing image based on a multiscale attention gan network. IEEE Geoscience and Remote Sensing Letters, 20:1–5.
[23] Wang, P., Bayram, B., and Sertel, E. (2022). A comprehensive review on deep learning based remote sensing image super-resolution methods. Earth-Science Reviews, 232:104110.
[24] Wang, P. and Sertel, E. (2023). Multi-frame super-resolution of remote sensing images using attention-based gan models. Knowledge-Based Systems, 266:110387.
[25] Wang, X., Xie, L., Dong, C., and Shan, Y. (2021). Real-ESRGAN: Training real-world blind super-resolution with pure synthetic data. 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), pages 1905–1914.
[26] Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Qiao, Y., and Change Loy, C. (2018). Esrgan: Enhanced super-resolution generative adversarial networks. In Proceedings of the European Conference on Computer Vision (ECCV), pages 0–0.
[27] Ayala, C., Aranda, C. M., & Galar, M. (2021). Sub-Pixel Width Road Network Extraction Using Sentinel-2 Imagery. https://doi.org/10.1109/igarss47720.2021.9555128