ResMergeNet: A Residual Learning Based U-Net for Building Segmentation using Multi-Resolution Data Fusion

Shailja [1], Ramji Dwivedi [1], Manohar Yadav [1]
[1] Motilal Nehru National Institute of Technology Allahabad

Keywords: Dataset Fusion, Remote Sensing, Aerial Images, Deep Learning, Building Segmentation.

Abstract

Data fusion in remote sensing is a critical task for integrating diverse datasets to enhance the accuracy of geospatial analysis. This research aims at building segmentation on a merge data combining the WHU building dataset and Massachusetts Building Dataset, leveraging deep learning for effective feature extraction. A model, ResMergeNet, based on Residual U-Net, is proposed to address challenges such as spatial resolution mismatch, complex building structures, and environmental diversity. The model successfully resolves issues like dataset heterogeneity, noise interference, and occlusions caused by trees and other objects. It also handles variations in building sizes, shapes, and boundaries across different datasets. The model achieves strong performance, with an IoU of 90.63%, accuracy of 95.13%, and an F1-score of 81.00%. The proposed architecture is also compared with other state-of-the art models and can be used in future in applications such as land use monitoring and large-scale building footprint mapping for improved geospatial analysis and smart city development.

1. Introduction

The rapid growth of urban areas and the increasing availability of high-resolution geospatial data are driving the need for accurate and efficient methods for segmenting and analyzing buildings. Building segmentation is widely recognized as a critical task in applications such as urban planning, disaster management, environmental monitoring, and infrastructure development (Agbaje et al., 2024). However, the merging of building datasets from multiple sources is often hindered by differences in data formats, spatial resolutions, and coverage areas. These challenges are addressed by utilizing advanced deep learning techniques that extract meaningful features and harmonize diverse datasets.

Traditional image processing and manual mapping techniques are often time-consuming, labor-intensive, and prone to inconsistencies. To overcome these limitations, recent advancements in deep learning have demonstrated impressive performance in image segmentation tasks, including the delineation of building footprints. Deep learning techniques, such as Convolutional Neural Networks (CNNs) (Bengio & Lecun, 1997) and U-Net architectures (Ronneberger et al., 2015), are commonly employed for segmentation and object detection tasks. However, these models are frequently observed to struggle with the effective integration of spatial and contextual information from complex datasets. Recent advancements in attention mechanisms (Bahdanau et al., 2015) and residual learning (He et al., 2015) are shown to enhance feature extraction and enable models to focus on relevant regions within the data.

The fusion of multi-source remote sensing data has emerged as a critical approach for improving building segmentation accuracy. (Sohn & Dowman, 2007) demonstrated the integration of IKONOS imagery and airborne LiDAR data to compensate for the limitations of single-source data, effectively improving building outline delineation. However, challenges included the alignment of rectilinear lines and the need for hierarchical partitioning. (Ma et al., 2024) tackled precision issues in building height estimation using synthetic aperture radar (SAR) and electro-optical (EO) images. Their multi-level cross-fusion strategy resolved the lack of complementary information in single-modality data, though noise and semantic refinement remained challenging. (Liu et al., 2024) focused on multi-resolution fusion through attention mechanisms, addressing challenges such as edge blurring, small building loss, and occlusion in complex urban environments.

However, a major challenge in building segmentation lies in the variability of building appearances due to differences in geographic location, architectural style, lighting conditions, and occlusions. Relying on a single source of imagery can lead to incomplete or inaccurate segmentation. Data fusion, which involves combining multiple datasets—such as aerial RGB images, satellite images (Dong et al., 2025), LiDAR data (Xu et al., 2025), and multispectral imagery (Zhao et al., 2025)—has emerged as a promising strategy to enhance segmentation accuracy and robustness. Collectively, these studies demonstrate that while multi-resolution and data fusion approaches significantly enhance building segmentation, challenges such as noise, alignment issues, and effective feature integration remain focal areas for improvement

In this research, ResMergeNet, is proposed. This model is based on the Residual U-Net architecture and is designed to merge and analyse two building datasets with varying characteristics. Residual connections are employed to improve gradient flow and mitigate vanishing gradient issues during training. The effectiveness of ResMergeNet is demonstrated on high-resolution building datasets. Superior accuracy and robustness are achieved when compared to conventional approaches. Contributions include a detailed exploration of the dataset fusion, model's performance and validation through experimental analysis. This study investigates how combining various types of spatial data can provide complementary information, enabling models to better distinguish between buildings and background objects. The study also evaluates the

effectiveness of different neural network architectures, such as U-Net, and Attention Residual U-Net, in processing fused data.

2. Method

This section delineates the dataset employed for the building segmentation study and elucidates the processing steps undertaken. Following that, the segmentation models employed in the study are explained, detailing their functionalities and significance in segmenting buildings. Subsequently, the experimental setup is presented, encompassing the tools and configurations essential for the ongoing research.

2.1 Dataset

This study utilizes and merges two benchmark datasets for building detection: Massachusetts Buildings Dataset (Mnih, 2013) and the WHU building dataset (Ji et al., 2019). The Massachusetts Buildings Dataset comprises 151 high-resolution aerial images (1500×1500 pixels) covering Boston with a resolution of 1 pixel per square meter. It includes diverse urban and suburban buildings, and building footprints derived from OpenStreetMap. The WHU building dataset contains over 220,000 buildings from Christchurch, New Zealand, with a resolution of 0.3 meters per pixel covering 450 square kilometers. It offers a large variety of building types and has been segmented into smaller tiles (512×512 pixels). Some images from both datasets include their corresponding building masks are used to create a new combined dataset. In the pre-processing stage, all images are resized to 256×256 pixels, and building labels are assigned to each image. The corresponding building masks are also processed similarly, converting RGB labels to 2D grayscale labels representing building or unlabeled areas. One-hot encoding is applied to prepare the labels for semantic segmentation (Table 1). The preprocessed data is split into training, validation and testing sets for model training and evaluation, with approximately 85% for training and validation and 15% for testing. This dataset will serve as input to train and evaluate the model for precise building detection in aerial image The final dataset is split into three subsets: 3542 images for training, 759 images for testing, and 760 images for validation. By merging these datasets, a new, more diverse dataset is created; however, several challenges are encountered, including mismatches in spatial resolution, environmental diversity, and varying levels of noise and image quality. These differences are addressed through careful pre-processing and adaptation to ensure that the model is effectively trained to generalize across different regions, building types, and conditions.

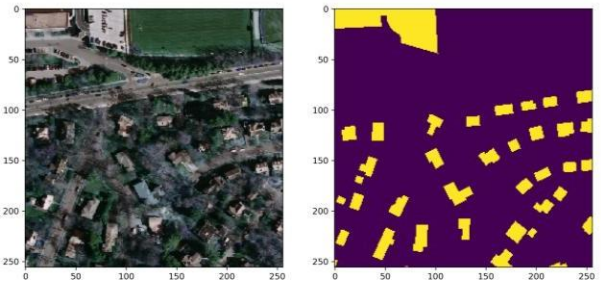
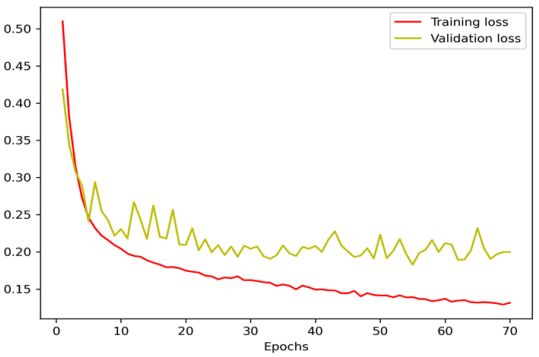


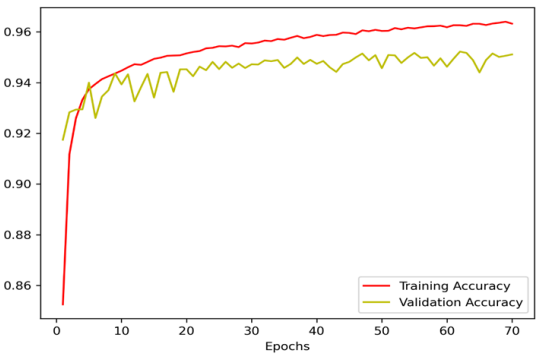
Figure 1. Image of the fused dataset after pre-processing.

2.2 ResMergeNet

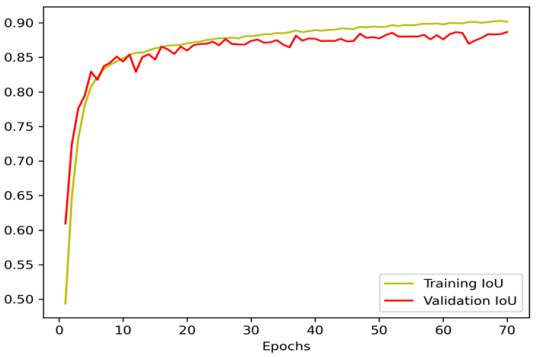
ResMergeNet is a deep learning model for merging and segmenting buildings from fused aerial imagery. It is built upon the Residual U-Net framework, combining the strengths of residual learning with the proven efficiency of encoder-decoder-based segmentation models. By integrating residual connections, ResMergeNet preserves critical multi-scale spatial and semantic features, facilitates better gradient flow during backpropagation, and mitigates the vanishing gradient problem, particularly in deeper layers. Residual learning enhances the model's ability to differentiate buildings from other objects such as roads, vegetation, or shadows, particularly in complex urban environments. The architecture consists of downsampling (encoder) layers to capture hierarchical and abstract features and upsampling (decoder) layers with skip connections to preserve spatial details, improving segmentation accuracy. Skip connections between corresponding encoder and decoder layers retain fine-grained spatial details, leading to more accurate and coherent boundary predictions. This dual-path strategy ensures that the model captures both contextual understanding and local precision, which are essential for reliable building segmentation.



(a)



(b)



(c)

Figure 2. Training and Validation performance plot of ResMergeNet showing (a) Training loss and Validation loss (b) Training accuracy and Validation accuracy (c) Training IoU and Validation IoU.

The model is applied to a merged dataset combining benchmark building datasets, enabling it to generalize across different spatial resolutions, environmental conditions, and noise levels. This ability ensures accurate identification of building footprints across diverse urban and rural settings. This fusion process introduces diversity in terms of spatial resolution, building styles, environmental conditions, and background clutter, thus encouraging the model to generalize effectively across varied urban and rural landscapes. By learning from this rich and heterogeneous dataset, ResMergeNet becomes robust to variability in rooftop textures, building sizes, shapes, occlusions (e.g., by trees), and imaging conditions.

In practical applications, this enhanced generalization capability allows ResMergeNet to deliver consistent, high-accuracy segmentation outputs across a broad range of geospatial scenarios. Its ability to preserve detailed structures while reducing misclassification makes it especially suitable for urban planning, disaster response, land use mapping, and other geospatial intelligence tasks that require precise delineation of man-made structures.

The model is trained for 70 epochs with a batch size of 4. The binary cross-entropy focal dice loss function (Wazir & Fraz, 2022) is used, while accuracy (Story & Congalton, 1986), IoU score (Rezatofighi et al., 2019), F1-score (Lipton et al., 2014), and Jaccard coefficient (Niwattanakul et al., 2013) serve as performance metrics. These metrics evaluate the model's segmentation quality and overall performance.

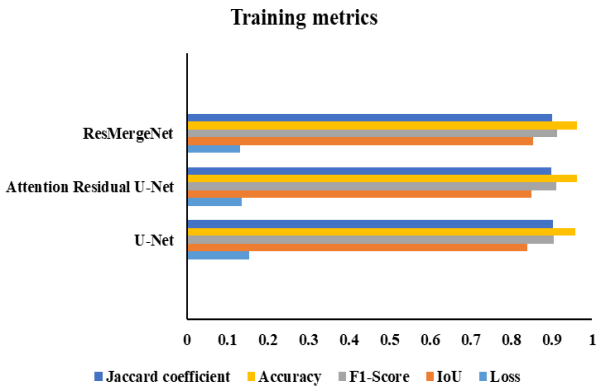
2.3 Training performance

ResMergeNet demonstrates strong performance on the merged dataset. The model achieves a training loss of 0.1316 as shown in Figure 2(a), an IoU of 0.8534 as depicted in Figure 2(c), and an F1-score of 0.9139. Training accuracy reaches 96.32% shown in Figure 2(b), and the Jaccard coefficient is 0.9015. On validation data, the model attains a loss of 0.1999, an IoU of 0.7997, and an F1-score of 0.8766. Validation accuracy reaches 95.11% and the Jaccard coefficient is 0.8868 highlighting the model's ability to generalize to unseen data. ResMergeNet surpassed the training performance of U-Net and Attention Residual U-Net as depicted in Figure 3. U-Net attained a training loss of 0.1536 which was higher than that of Attention Residual U-Net having training loss of 0.1359. Attention Residual U-Net performed well in comparison to U-Net in case of training but U-Net out shown its performance in the case of validation. Attention Residual U-Net had training IoU of 0.8499, F1-Score of 0.9051, accuracy of 0.9626 and jaccard coefficient of 0.8997 which was higher than U-Net having IoU as 0.8402, F1-Score as 0.9051, accuracy as 0.9589 and jaccard coefficient as 0.9024. Whereas, in case of validation U-Net had a validation loss of 0.2088, IoU of 0.7998, F1-Score of 0.877, accuracy of 0.9506 and jaccard coefficient of 0.891 which was greater than Attention Residual U-Net having validation loss of 0.2122, IoU of 0.7869, F1-Score of 0.867, accuracy of 0.947 and jaccard coefficient of 0.8717.

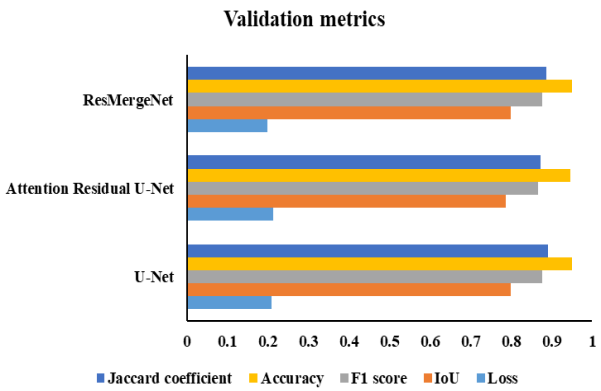
3. Results

The merged dataset combining two benchmark building datasets differ in spatial resolution as WHU building dataset is having high resolution than Massachusetts Buildings Dataset. So, this combination creates a diverse and comprehensive dataset having certain challenges such as spatial resolution mismatch, environmental diversity, complex building surrounding, different building types and their structure. Even considering these challenges, the model is able to segment buildings under variable conditions and performs well. The performance evaluation of this study is detailed in Table 1, highlighting metrics such as Mean IoU, IoU, accuracy, precision (Menditto et al., 2007), recall (Martin & Powers, 2011), and F1-Score for building segmentation results.

The ResMergeNet achieves a Mean IoU of 77.72% and an accuracy of 95.13%, reflecting high precision in building segmentation. It attains a precision of 87.83% and recall of 75.15%, balancing detection accuracy and completeness. This architecture for fused dataset provides an IoU Of 90.63% and F1-Score of 81.00%



(a)



(b)

Figure 3. Performance Metrics of U-Net, Attention Residual U-Net and ResMergeNet showing (a) Training and (b)Validation.

Mean IoU	77.72
IoU	90.63
Accuracy	95.13

Precision	87.83
Recall	75.15
F1-Score	81.00

Table 1. Evaluation Metrics in (%) of ResMergeNet

The proposed ResMergeNet model is evaluated on the testing dataset to assess its performance in real-world conditions. Four sample images are selected as representative examples to illustrate the model's prediction results. For each example, the testing image, the corresponding ground truth building mask (testing label), and the predicted output are presented. These visual results demonstrate the model's ability to detect building footprints accurately, even in challenging scenarios. Test image numbers 80**,** 345**,** 581**,** and 585 have been selected as examples to demonstrate the performance of the ResMergeNet model.

Figure 4 (a) and Figure 4 (b) primarily represent images from the WHU building dataset, while Figure 4 (c) and Figure 4 (d) correspond to images from the Massachusetts Buildings Dataset. Figure 4 (a) showcases buildings with diverse roof types, surrounded by dense vegetation. While the model successfully detects most buildings and reduces unnecessary noise by avoiding the misclassification of other objects as buildings, some challenges persist in accurately delineating the shapes of all the buildings. This highlights both the strengths and limitations of the model in complex environments. Figure 4 (b) presents a cluttered environment with non-uniform building structures, typical of a complex commercial area. While the model performs well in accurately delineating the shapes of many buildings, it has missed some buildings due to the dense and heterogeneous nature of the surroundings. Despite this, the model has made significant efforts to accurately define the building boundaries in a challenging and visually complex setting.

In Figure 4 (c), the lower resolution of the image posed challenges for the model in accurately identifying the boundaries of large buildings. However, the model successfully detected small buildings with high accuracy, rarely missing any. This performance can be attributed to the feature learning from the fused dataset, which enabled the model to generalize well across varying building scales and conditions. Figure 4 (d) contains the highest number of buildings among the four sample images. Despite the structured and organized surroundings, the buildings vary significantly in size, including both large and small structures. While the model successfully detected the buildings, it struggled to differentiate between them in some cases, resulting in converged or merged building detections. This highlights a limitation of the model and indicates the need for further improvement. Among all the examples, this image presents the most complex and challenging case for building detection.

4. Discussion & Conclusion

4.1 Quantitative analysis

Despite differences in spatial resolution and environmental conditions, ResMergeNet effectively segments buildings across diverse datasets. The ResMergeNet achieves highest mean IoU, IoU, accuracy, precision, recall and F1-Score reflecting high precision in building segmentation in comparison to U-Net and attention mechanism-based Attention Residual U-Net architecture. Figure 5 perfectly describes about evaluation metrics of U-Net, Attention Residual U-Net and ResMergeNet in segmenting buildings from fused dataset.

The results indicate that the U-Net slightly outperforms the Attention Residual U-Net across most metrics. U-Net achieved a mean IoU of 77.05%, compared to 76.22% for the Attention Residual U-Net. Similarly, U-Net attained a marginally higher IoU of 90.53% versus 90.12%. In terms of overall pixel-wise segmentation, U-Net recorded an accuracy of 95.03%, exceeding the 94.82% of the Attention Residual U-Net. U-Net achieved a precision of 89.13% and a recall of 72.91%, while the Attention Residual U-Net reached 88.11% and 72.20%, respectively. Consequently, the F1-Score, which balances precision and recall, was also higher for U-Net (80.21%) compared to the Attention Residual U-Net (79.36%). These findings suggest that while attention mechanisms and residual connections contribute to learning contextual features, the standard U-Net architecture remains slightly more effective for this specific task and dataset.

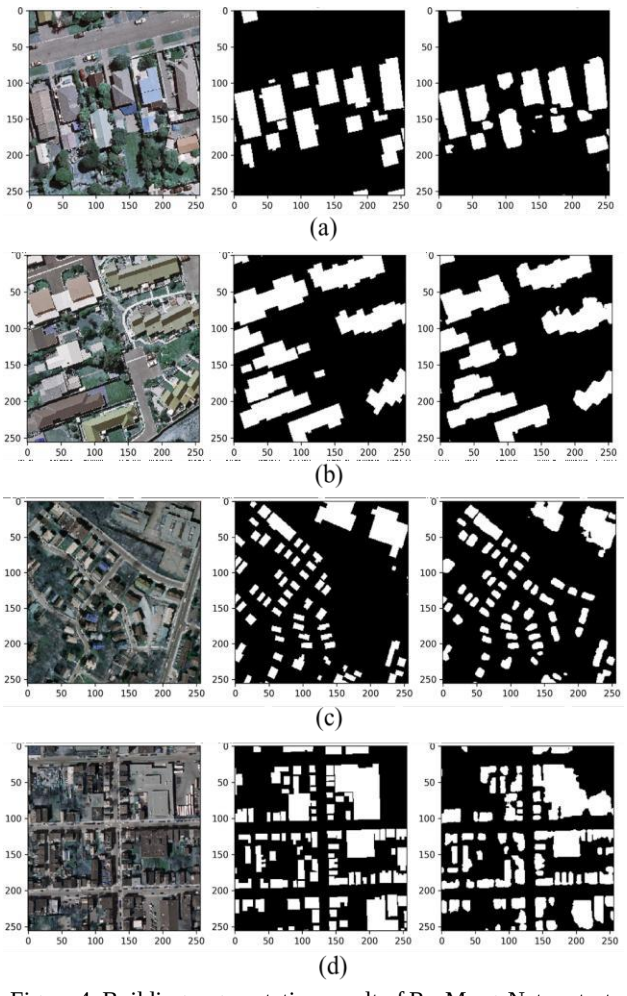


Figure 4. Building segmentation result of ResMergeNet on test image number (a) 80 (b) 345 (c) 581 (d) 585.

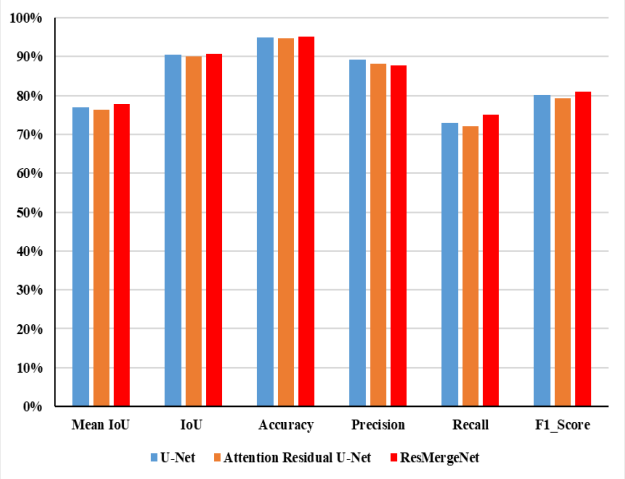


Figure 5. Evaluation Metrics performance of U-Net, Attention Residual U-Net and ResMergeNet.

Overall, ResMergeNet outperformed both U-Net and Attention Residual U-Net across most evaluation metrics. It achieved a 1.5% higher mean IoU and 1.64% higher F1-Score compared to the Attention Residual U-Net, and was 0.87% better in mean IoU and 0.79% better in F1-Score than U-Net. Additionally, it improved recall by over 2% compared to both models, demonstrating its enhanced ability to detect building regions more completely. These results confirm that ResMergeNet offers a more accurate and balanced segmentation performance for fused aerial imagery.

4.2 Qualitative analysis

This section describes how ResMergeNet performed well in segmenting buildings on all 4 test images in comparison to U-Net and Attention Residual U-Net architectures. Qualitative analysis shows that the model performs well in identifying buildings in vegetated, clustered, and dense urban environments, though challenges remain in delineating structures in low-resolution images. Figure 6(a) presents the results for test image no. 80. In this image, the U-Net model missed several buildings, produced inaccurate building shapes, and mistakenly identified a tree as a building, showing limitations in distinguishing between built and natural features. The Attention Residual U-Net performed slightly better, successfully identifying more buildings; however, it still missed some structures, misclassified another surface, and generated less precise outlines. In comparison, the ResMergeNet delivered the most accurate results, with more complete building coverage, fewer false positives, and well-preserved building shapes, closely aligning with the ground truth. Figure 6(b) illustrates the results for test image no. 345. In this case, the U-Net model segmented a single building into separate parts, likely due to color variations, and missed the overall structure of the building where it was occluded by trees. Additionally, the predicted building shapes were inaccurate, and a nearby building was completely missed due to vegetation cover. The Attention Residual U-Net performed better by addressing some of these issues; it managed to preserve more continuous structures and avoided missing buildings solely due to tree occlusion, although it still failed to detect certain smaller buildings. The ResMergeNet produced the most accurate building shapes among the three, maintaining better boundary definition and completeness. However, all three models struggled with partial occlusions, particularly where trees overlapped the rooftops, resulting in missed parts of buildings in those areas. This highlights a common limitation in segmentation performance under dense vegetation cover, despite the improvements shown by ResMergeNet.

Figure 7(a) presents the segmentation results for test image no. 581. The U-Net model showed several limitations in this image, where multiple buildings were partially segmented or entirely missed, and some were incorrectly divided or merged into a single structure due to shape distortion and misclassification. The predicted building outlines lacked accuracy, and smaller structures were often overlooked. The Attention Residual U-Net improved upon U-Net by preserving better separation between adjacent buildings, avoiding the merging issue. However, it missed a major large building entirely and detected another large building with a distorted shape, failing to preserve structural boundaries. Additionally, small-sized buildings were still missed. In contrast, ResMergeNet successfully detected most of the buildings, providing more complete and accurate building shapes, and better handling separation between adjacent structures. While some minor inaccuracies remained, ResMergeNet demonstrated stronger robustness in capturing both large and small buildings across varying urban textures.

Figure 7(b) shows the segmentation results for test image no. 585, which represents one of the most challenging cases in the test set. The image contains a dense urban layout with a mix of building types, sizes, colours, and orientations, as well as significant visual clutter and occlusions. The U-Net model struggled considerably in this scenario, producing poor segmentation performance, where multiple buildings were merged, shapes were highly inaccurate, and smaller structures were either distorted or missed entirely. The high level of congestion and variability in the image made it particularly difficult for the model to distinguish individual building boundaries. Both the Attention Residual U-Net and ResMergeNet also faced challenges in this complex scene. However, Attention Residual U-Net delivered slightly better results in this case. The incorporation of the attention mechanism allowed the model to focus more effectively on relevant features, helping it better differentiate between closely packed buildings. While it still misclassified some areas and missed fine details, the segmentation outlines were comparatively cleaner and more accurate than those of U-Net. ResMergeNet, although effective in many other cases, showed limitations here, likely due to the overwhelming variety of textures and overlaps, which diminished the impact of its fusion strategy.

This comparison highlights that even advanced models can struggle in highly congested urban environments, and integrating attention mechanisms can provide marginal advantages in extracting meaningful patterns from visually complex inputs.

4.3 Conclusion

The proposed ResMergeNet model demonstrates significant advancements in building segmentation by integrating residual learning within a U-Net architecture. By leveraging a fused dataset comprising the Massachusetts Buildings Dataset and WHU building dataset, the model effectively adapts to diverse urban scenarios characterized by variations in building size, shape, environmental conditions, and spatial resolutions. Quantitative evaluations reveal the model's robustness,

achieving an IoU of 90.63% and an accuracy of 95.13%, underscoring its capability to precisely detect building footprints despite challenges such as occlusions, shadow interference, and non-uniform building layouts. Qualitative analysis further validates the model's strength in handling complex scenarios, from vegetated and clustered environments to low-resolution and dense structured urban areas. While the model achieves commendable performance, limitations such as difficulty in delineating building boundaries in low-resolution images and distinguishing converged structures highlight areas for future improvement. The proposed methodology can be further enhanced by integrating advanced data augmentation techniques, adoption of attention and transformer mechanism for multi-scale feature fusion approach, more advanced context understanding, better feature reuse, improved spatial awareness, preserve building boundaries and improved performance.

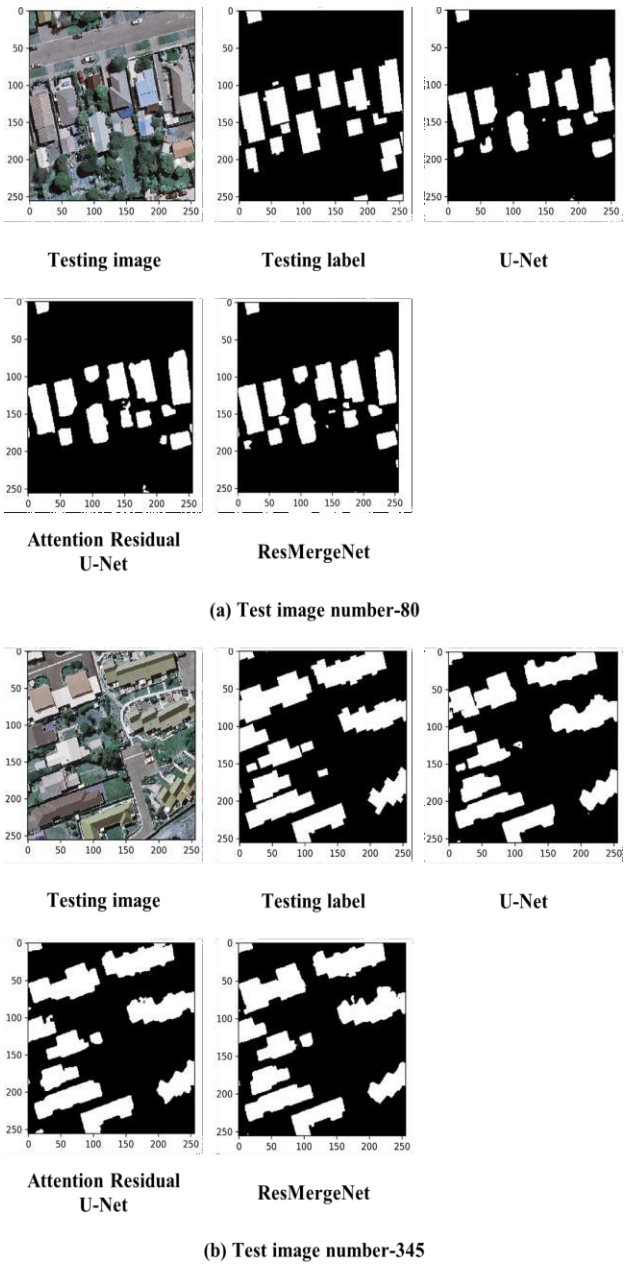


Figure 6. Building segmentation result of U-Net, Attention Residual U-Net and ResMergeNet on test image number (a) 80 and (b) 345.

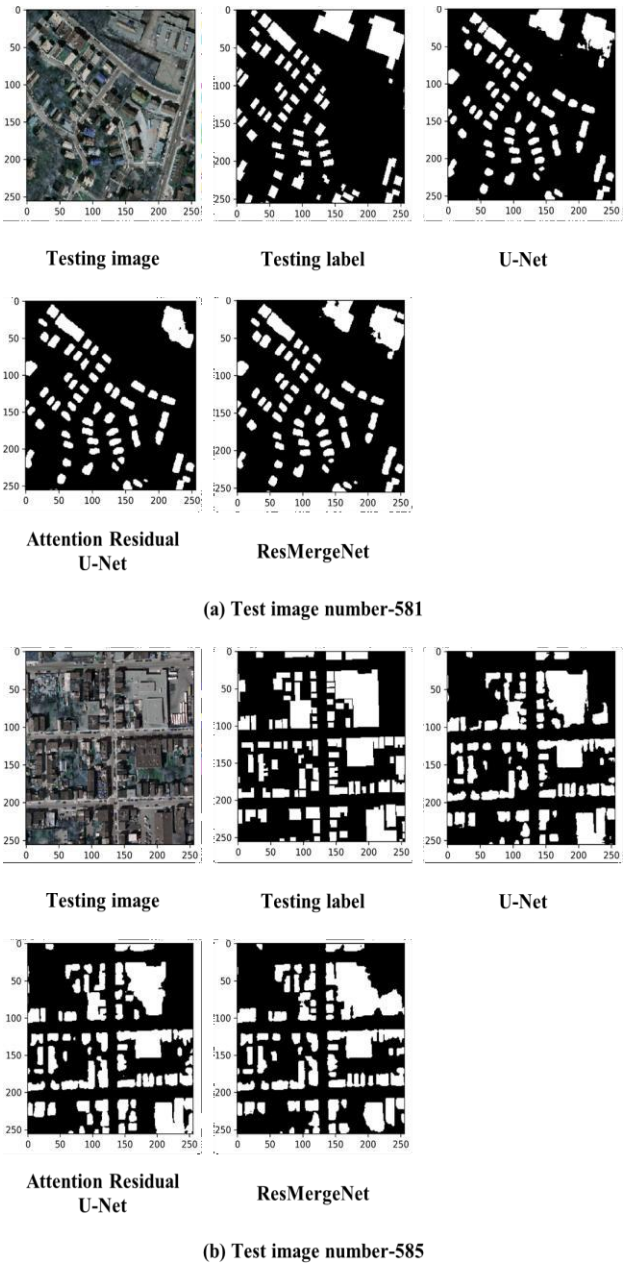


Figure 7. Building segmentation result of U-Net, Attention Residual U-Net and ResMergeNet on test image number (a) 581 and (b) 585.

References

Agbaje, T. H., Abomaye-nimenibo, N., Ezeh, C. J., Bello, A., & Olorunnishola, A. (2024). Building Damage Assessment in Aftermath of Disaster Events by Leveraging Geoai (Geospatial Artificial Intelligence): Review. *World Journal of Advanced Research and Reviews*, *23*(July), 667–687. https://doi.org/10.30574/wjarr.2024.23.1.2000

Bahdanau, D., Cho, K. H., & Bengio, Y. (2015). Neural machine translation by jointly learning to align and translate. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 1–15.

Bengio, Y., & Lecun, Y. (1997). *Convolutional Networks for Images , Speech , and Time-Series*.

Dong, X., Li, J., Chang, Q., Miao, S., & Wan, H. (2025). Data Fusion and Models Integration for Enhanced Semantic Segmentation in Remote Sensing. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *18*, 7134–7151. https://doi.org/10.1109/JSTARS.2025.3528650

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep Residual Learning for Image Recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *2016-Decem*, 770–778. https://doi.org/10.48550

Ji, S., Wei, S., & Lu, M. (2019). Fully Convolutional Networks for Multisource Building Extraction From an Open Aerial and Satellite Imagery Data Set. *IEEE Transactions on Geoscience and Remote Sensing*, *57*(1), 574–586. https://doi.org/10.1109/TGRS.2018.2858817

Lipton, Z. C., Elkan, C., & Narayanaswamy, B. (2014). *Thresholding Classifiers to Maximize F1 Score*. http://arxiv.org/abs/1402.1892

Liu, J., Gu, H., Li, Z., Chen, H., & Chen, H. (2024). Multi-Scale Feature Fusion Attention Network for Building Extraction in Remote Sensing Images. *Electronics (Switzerland)*, *13*(5). https://doi.org/10.3390/electronics13050923

Ma, C., Zhang, Y., Guo, J., Zhou, G., & Geng, X. (2024). FusionHeightNet: A Multi-Level Cross-Fusion Method from Multi-Source Remote Sensing Images for Urban Building Height Estimation. *Remote Sensing*, *16*(6). https://doi.org/10.3390/rs16060958

Martin, D., & Powers, W. (2011). Evaluation : From precision , recall and F-measure to ROC , informedness , markedness & correlation. *Journal of Machine Learning Technologies*, *2*(May), 2229–3981. https://doi.org/10.9735/2229-3981

Menditto, A., Patriarca, M., & Magnusson, B. (2007). Understanding the meaning of accuracy , trueness and precision. *Accreditation and Quality Assurance*, *12*(1), 45–47. https://doi.org/10.1007/s00769-006-0191-z

Mnih, V. (2013). *Machine Learning for Aerial Image Labeling*.

Niwattanakul, S., Singthongchai, J., Naenudorn, E., & Wanapu, S. (2013). Using of Jaccard Coefficient for Keywords Similarity. *The 2013 IAENG International Conference on Internet Computing and Web Services (ICICWS'13)*, *March*.

Rezatofighi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., & Savarese, S. (2019). Generalized intersection over union: A metric and a loss for bounding box regression. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *2019-June*, 658–666. https://doi.org/10.1109/CVPR.2019.00075

Ronneberger, O., Fischer, P., & Brox, T. (2015). UNet: Convolutional Networks for Biomedical Image Segmentation. *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 234–241. https://doi.org/10.48550/arXiv.1505.04597

Sohn, G., & Dowman, I. (2007). Data fusion of high-resolution satellite imagery and LiDAR data for automatic building extraction. *ISPRS Journal of Photogrammetry and Remote Sensing*, *62*(1), 43–63. https://doi.org/10.1016/j.isprsjprs.2007.01.001

Story, M., & Congalton, R. G. (1986). Accuracy Assessment: A User's Perspective. *Photogrammetric Engineering and Remote Sensing*, *52*(3), 397–399. https://www.asprs.org/wp-content/uploads/pers/1986journal/mar/1986_mar_397-399.pdf

Wazir, S., & Fraz, M. M. (2022). HistoSeg: Quick attention with multi-loss function for multi-structure segmentation in digital histology images. *2022 12th International Conference on Pattern Recognition Systems, ICPRS 2022*. https://doi.org/10.1109/ICPRS54038.2022.9854067

Xu, X., Sun, Z., Mao, Z., Feng, Y., Zhang, H., Bai, J., Yang, J., Ran, Y., & Tan, Z. (2025). Evaluation method of rooftop photovoltaic resources of distributed buildings based on the fusion of ResFAUnet and MAS-PointMLP. *Energy and Buildings*, *337*(March). https://doi.org/10.1016/j.enbuild.2025.115680

Zhao, Z., Zhao, B., Wu, Y., He, Z., & Gao, L. (2025). Building extraction from high-resolution multispectral and SAR images using a boundary-link multimodal fusion network. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *18*, 3864–3878. https://doi.org/10.1109/JSTARS.2025.3525709