Rooftop Segmentation using Geospatial Artificial Intelligence for Rural India

Soham Rangdal*, Prakhar Verma, Vivek Singh Tomar, Shankar Naik Rathod Karamtoth, Kedar Nagnathrao Ghogale, Sivakumar V, Biju C, Jitendra Mhatre, Sajeevan G

Centre for Development of Advanced Computing (C-DAC), Pune, Maharashtra, India
*rangdalsoham1@gmail.com

Keywords: Rooftop, GeoAI, Deep Learning, AI/ML, Semantic Segmentation, HPC

Abstract

Rooftop identification using satellite imagery is having a wide range of practical use cases, which includes habitation identification, disaster management, infrastructure monitoring and solar panel deployment. Recent advances in GeoAI have improved rooftop extraction by integrating spatial technologies and deep learning models. However, it remains a challenge in rural India due to the irregular shapes of rooftops, variations in surface texture, low resolution open-source satellite data, canopy cover, and diverse geographic conditions. This paper focuses on rooftop extraction from satellite imagery utilizing the YOLOv11l segmentation framework trained on CDAC PARAM Siddhi-AI HPC platform. Seamless labels were created over satellite images using QGIS open-source software. We applied the super-resolution model to improve overall quality, thereby up-scale the resolution from 320×320 to 640×640 pixels. The super-resolution images were tiled into 320x320 for training. Additionally, image processing techniques (contour detection, erosion and dilation) were used to reduce noise and maintain consistency of images. The model trained on the pre–super-resolution dataset achieved a mean Intersection over Union (mIoU) of 36.20% and a precision of 76.42%, whereas the model trained on the post–super-resolution dataset achieved a higher mIoU of 61.15% and a precision of 80.92% on the test dataset. It indicates that the super-resolution model performed better prediction. The model is a deployable GeoAI solution for national schemes like PMGSY National GIS and PM *Surya Ghar*.

1. Introduction

The rooftop detection from satellite imagery is significant for various national development needs such as infrastructure planning, disaster management, human settlement mapping and solar panel deployment. There are many challenges in detecting rooftops from satellite imagery in rural areas. The rooftops in rural India do not maintain uniform patterns in shape, size and construction materials and generally built using tin with thatch as well as concrete with clay tiles. The rooftops maybe fully/partially covered by vegetation and shadows. The buildings in villages maintain a low density with scattered locations. Currently, various deep learning algorithms are utilized for rooftop detection in urban areas however; the models trained on urban datasets are not suitable to detect rural rooftops. Moreover, in case of rural rooftop detection the absence of annotated satellite dataset makes it difficult to develop reliable rooftop detection model.

Several studies have explored deep learning models for detecting rooftop from satellite imagery. Zhao et al. (2018) introduced a rooftop detection method using Mask R-CNN with polygon regularization, applied to the Deep-Globe Building Extraction Challenge (DG-BEC) dataset. Their approach effectively simplified polygonal boundaries and was particularly suitable for detecting compact and well-structured buildings. However, the method struggled with irregular rooftops, which requires extensive hyper-parameter tuning and computation to maintain performance. Building on these foundational efforts, Amo-Boateng et al. (2022) utilized Mask R-CNN with the Inception ResNet V2 backbone using transfer learning for rooftop segmentation in 450 aerial images of rural Ghana. Their model not only achieved precise rooftop detection but also enabled accurate area estimation. This was particularly impactful in assessing solar photovoltaic potential and planning infrastructure.

In the context of humanitarian applications, Gella et al. (2022) used Mask R-CNN to detect shelters in refugee and Internally Displaced Person (IDP) settlements using very high-resolution (VHR) satellite imagery. Their model demonstrated strong performance even in environments with significant occlusions and irregular layouts. Ghorbanzadeh et al. (2022) further extended this work by employing CNN-based methods optimized through preprocessing and data augmentation, enabling better detection of informal and non-standardized shelter types. Yang et al. (2023) addressed class imbalance and segmentation precision by enhancing the U-Net architecture. Using the Inria Aerial Image Labeling Dataset, they integrated a VGG19-ASPP backbone with an ESPCN decoder and hybrid loss function, achieving a dice index of 0.955. Chen et al. (2025) resolved rooftop detection issue in degraded black-and-white aerial images. Their two-stage approach used DeOldify for colorization and Real-ESRGAN for enhancement, followed by YOLOv11n for detection, reaching 85.2% accuracy. Further improving structural precision, Mei et al. (2025) employed a YOLOv8-based Oriented Bounding Box (OBB) technique for extracting roof edges. Their approach, combining deep learning, geometric rules, and advanced post-processing, showed better results on the SGA and Melville datasets outperforming Segment Anything Model (SAM) model in terms of mIoU and vector IoU. The method was effective for automatic roof detection in urban region.

The challenges in rooftop detection include foreground-background class imbalance, occlusions, noisy imagery, irregular rooftop shapes and closely packed structures. These factors continue to limit detection performance, which demand more attention for robust models that are suited for both urban and rural areas.

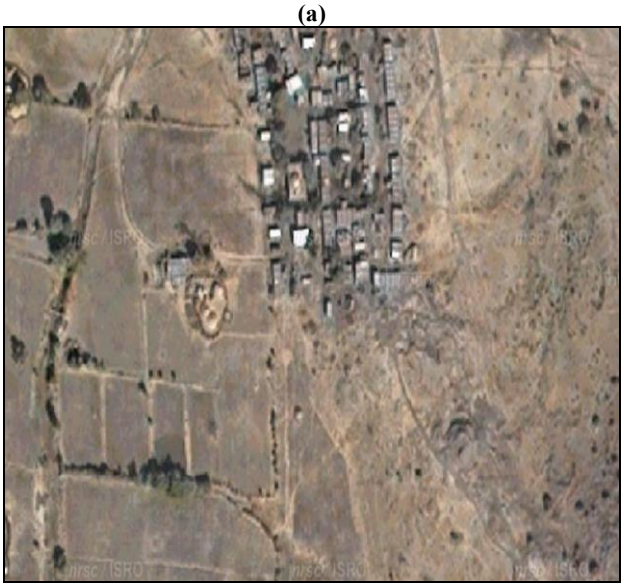
This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-5-W2-2025-537-2025

2. Methodology

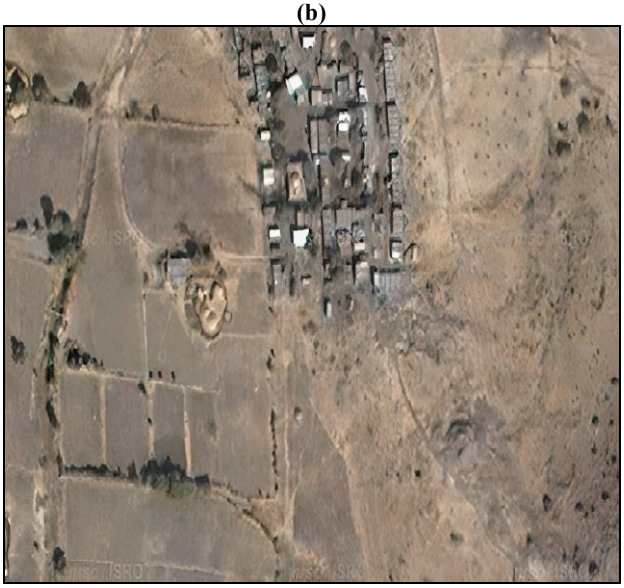
2.1 Data Preparation

The proposed method is to identify rooftop from satellite imagery sourced from ISRO-Bhuvan. The annotated images for rooftop detection were tiled into 320x320 pixels. We used the Real-ESRGAN super-resolution model to enhance image quality by upscaling the images to 640x640 pixels. The super-resolution images were further cropped into 320x320 tiles. For experimental purpose 600 images were utilised for training. Figure 1(a) shows a pre-super-resolution image, Figure 1(b) shows a post-super-resolution image and Figure 1(c) represents mask.

(a)



(b)



(c)

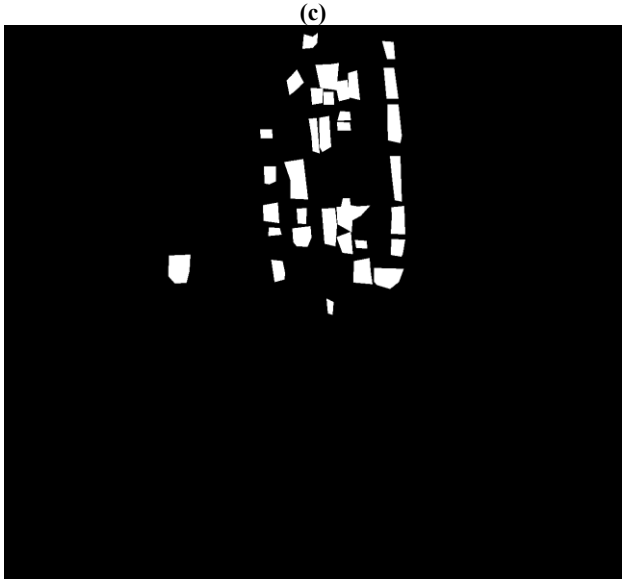
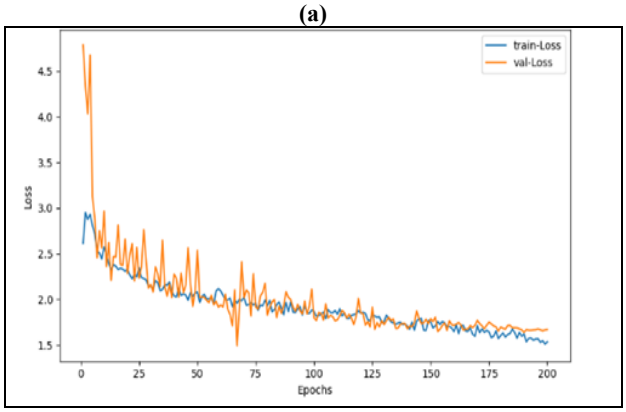


Figure 1. Input data (Source: ISRO-Bhuvan) (a) Pre-Super-Resolution Image (b) Post-Super-Resolution Image and (c) Mask

2.2 Model Training

The YOLOv11l model training utilizes 320x320 tiles. The model's ability to generalize and perform robust rooftop segmentation from satellite imagery received several advanced training approaches. Multi-scale training (multi_scale=True) was used to present the model with different resolution images which improved its capacity to work with multiple scales.

(a)



(b)

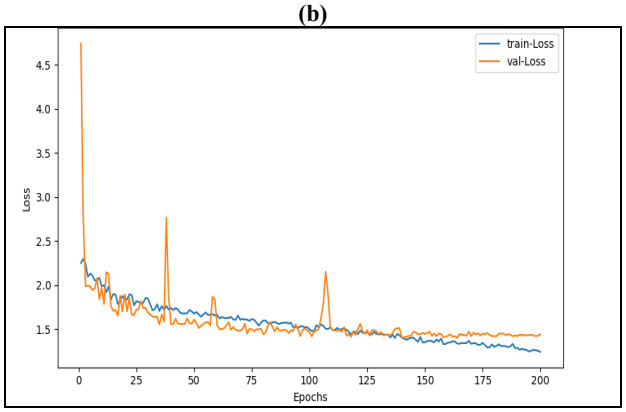


Figure 2. Training and validation loss over epochs of (a) Pre-SR-model and (b) Post-SR-model

Training process included regularisation methods which combined dropout at 50% with weight decay at 0.0005 and momentum at 0.9 for optimisation stability. In addition to that Stochastic Gradient Descent (SGD) optimiser was used with a cosine learning rate scheduler. Using these parameters we trained two separate models: the Pre-SR-model, based on the pre-super-resolution dataset, and the Post-SR-model, based on the post-super-resolution dataset. The training and validation loss curves for the two models reveal distinct differences in learning stability and generalization performance. The loss graph of the Pre-SR model (Figure 2a) shows that both training and validation losses start at high levels and drop sharply within the first 20 epochs. The validation curve exhibits persistent fluctuations, indicating instability and poor generalization attributed to low image resolution. However, the Post-SR-model loss graph (Figure 2b) shows an initial drop in loss, followed by a steady trend with reduced validation loss. The enhanced model performance shows super-resolution techniques delivered sharper rooftop boundaries and textures which allowed the model to develop stronger learning capabilities and better handle super resolution data. Both models reach comparable final states with only a minor difference between training and validation losses. However, the Post-SR-model exhibits reduced validation loss variability, indicating that super-resolution preprocessing enhances the effectiveness of rooftop segmentation.

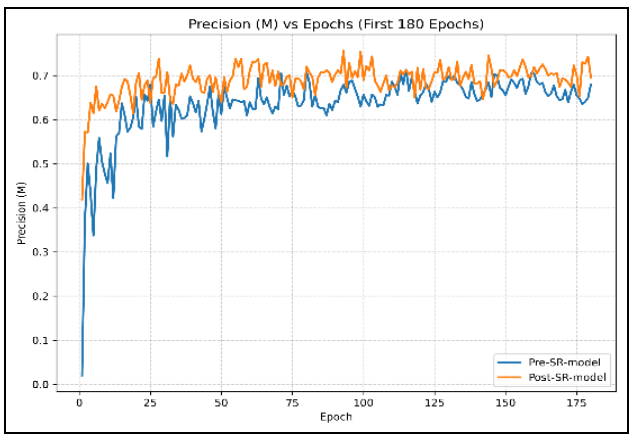


Figure 3. Precision (M) vs. Epoch

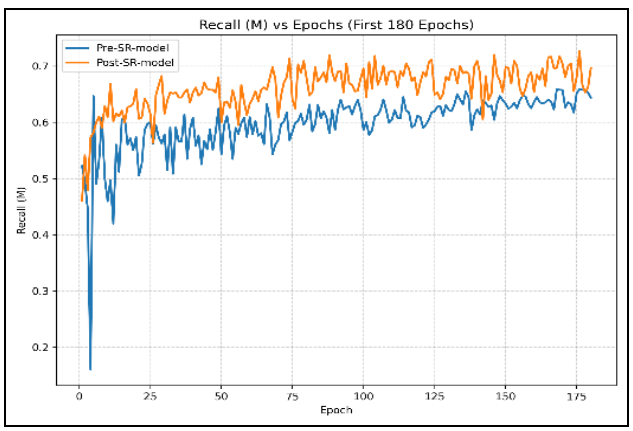


Figure 4. Recall (M) vs. Epoch

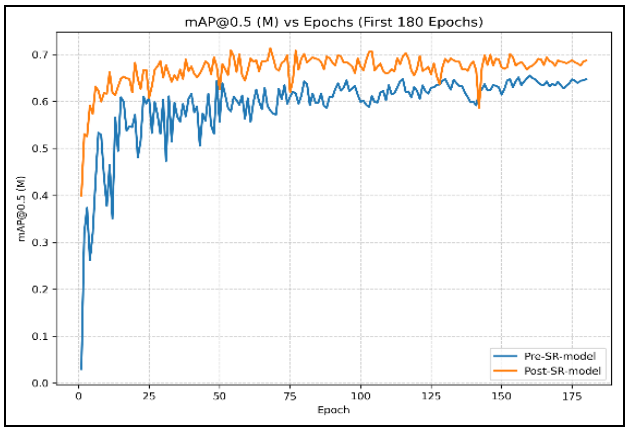


Figure 5. mAP@0.5 (M) vs. Epoch

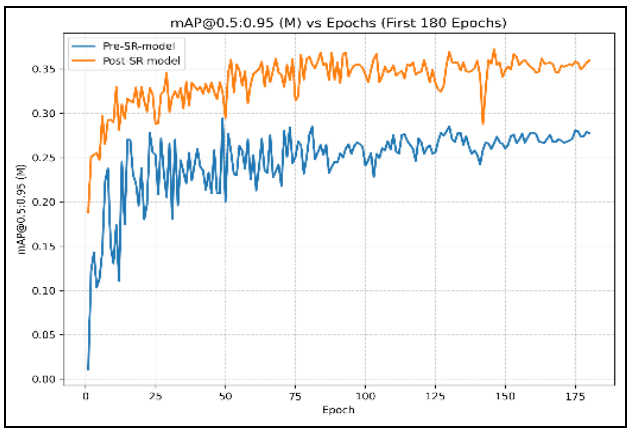


Figure 6. mAP@0.5:0.95 (M) vs. Epoch

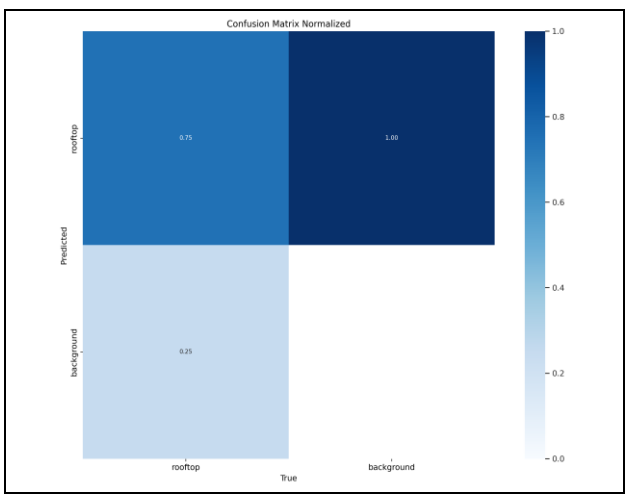


Figure 7. Confusion matrix of the Pre-SR model

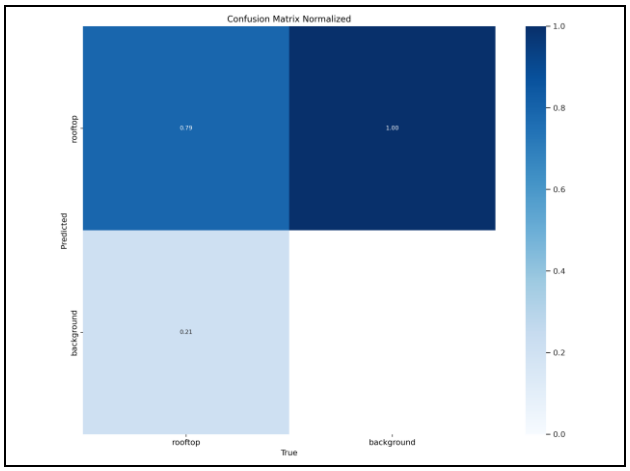


Figure 8. Confusion matrix of the Post-SR model

The comparison between the Pre-SR and Post-SR YOLOv11 models show a clear improvement when using the super-resolution (SR) pre-processing technique. Figure 3 (Precision) illustrates that the Post-SR model makes more accurate predictions, reducing false positives and improving detection confidence. Figure 4 (Recall) shows identification of greater number of true objects. In Figure 5 (mAP@0.5), the Post-SR model consistently achieves better accuracy in detecting rooftops, which indicates that super resolution data helps in better model learning. Figure 6 (mAP@0.5:0.95) demonstrates that even at stricter evaluation thresholds, the Post-SR model outperforms the Pre-SR model, showing better object localization and segmentation precision. Figures 7 and 8 show the confusion matrices for YOLOv11, providing a visual summary of the Pre-SR and Post-SR model performance in rooftop detection.

Metric (mean)	Pre-SR-model	Post-SR-model
Precision(B)	0.5569	0.6340
Recall(B)	0.4985	0.5898
mAP50(B)	0.4918	0.6044
mAP50-95(B)	0.2234	0.3286
Precision(M)	0.5214	0.6545
Recall(M)	0.4776	0.5829
mAP50(M)	0.4419	0.6031
mAP50-95(M)	0.1634	0.2849

Table 1. Validation performance metrics

Table 1 shows mean scores which represent the average values from all epoch-wise validation performance instead of single checkpoint selection. Averaging over the training period removes short-term variations, providing a clearer and more consistent representation of the model performance. The Post-SR-model achieves higher scores than the Pre-SR-model in all evaluation metrics for bounding box detection and segmentation mask tasks. Detection precision increased from 0.5569 to 0.6340 and recall improved from 0.4985 to 0.5898 resulting in better mAP50 from 0.4918 to 0.6044 and mAP50–95 from 0.2234 to 0.3286. Segmentation precision improved from 0.5214 to 0.6545 while recall increased from 0.4776 to 0.5829 and mAP50 rises from 0.4419 to 0.6031 and mAP50–95 from 0.1634 to 0.2849. The consistent improvement in all evaluation metrics at higher IoU thresholds proved super-resolution preprocessing creates better rooftop feature resolution which leads to more robust training generalization.

3. Experimental Setup

Model training was carried out on the CDAC PARAM Siddhi-AI HPC system with four NVIDIA A100 SXM4 GPUs (40GB). Using Ultralytics Segmentation Framework (v8.3.133), which ran on Python 3.10 with PyTorch 2.6.0+cu118. We applied 50% dropout, 0.0005 weight decay, 0.9 momentum and SGD optimiser with cosine learning rate scheduler.

4. Result and Discussion

The model trained on the data pre-super-resolution (Pre-SR-model) and post-super-resolution (Post-SR-model) took 12 and 23 minutes respectively for 200 epochs with the batch size of 32. The Table 2 summarizes segmentation test results of Pre-SR-model and Post-SR-model.

Metrics	Pre-SR-model	Post-SR-model
IoU	36.20%	61.15%
Loss	0.485	0.251
Precision	76.42%	80.92%
Recall	40.64%	71.17%
F1 Score	51.49%	74.88%

Table 2. Performance metrics on test data

The Post-SR-model delivered superior segmentation results, as its mean IoU reached 61.15%, outperforming the Pre-SR-model at 36.20%. It achieved better predictions with a lower loss value of 0.251 compared to the Pre-SR-model 0.485. Precision was maintained well by both models, yet the Post-SR-model performed better with a precision score of 80.92% versus 76.42%, leading to fewer false positives. The Post-SR-model also produced substantially better recall results, reaching 71.17% compared to the Pre-SR-model 40.64%. This showed that the Post-SR-model identified more true positives while reducing missed segmentation targets. It achieved strong segmentation performance with an F1 Score of 74.88%, whereas the Pre-SR-model reached only 51.49%. The study showed that training data pre-processing with super-resolution methods significantly enhanced segmentation outcomes across multiple evaluation metrics. The Post-SR-model demonstrated superior ability in both identifying and precisely outlining target areas in test images, highlighting the practical advantage of using higher-resolution data for segmentation applications. The inference outputs (Figure 9) displayed the Post-SR-model enhanced segmentation results, showing better precision and more complete outputs compared to the Pre-SR-model.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-5-W2-2025-537-2025

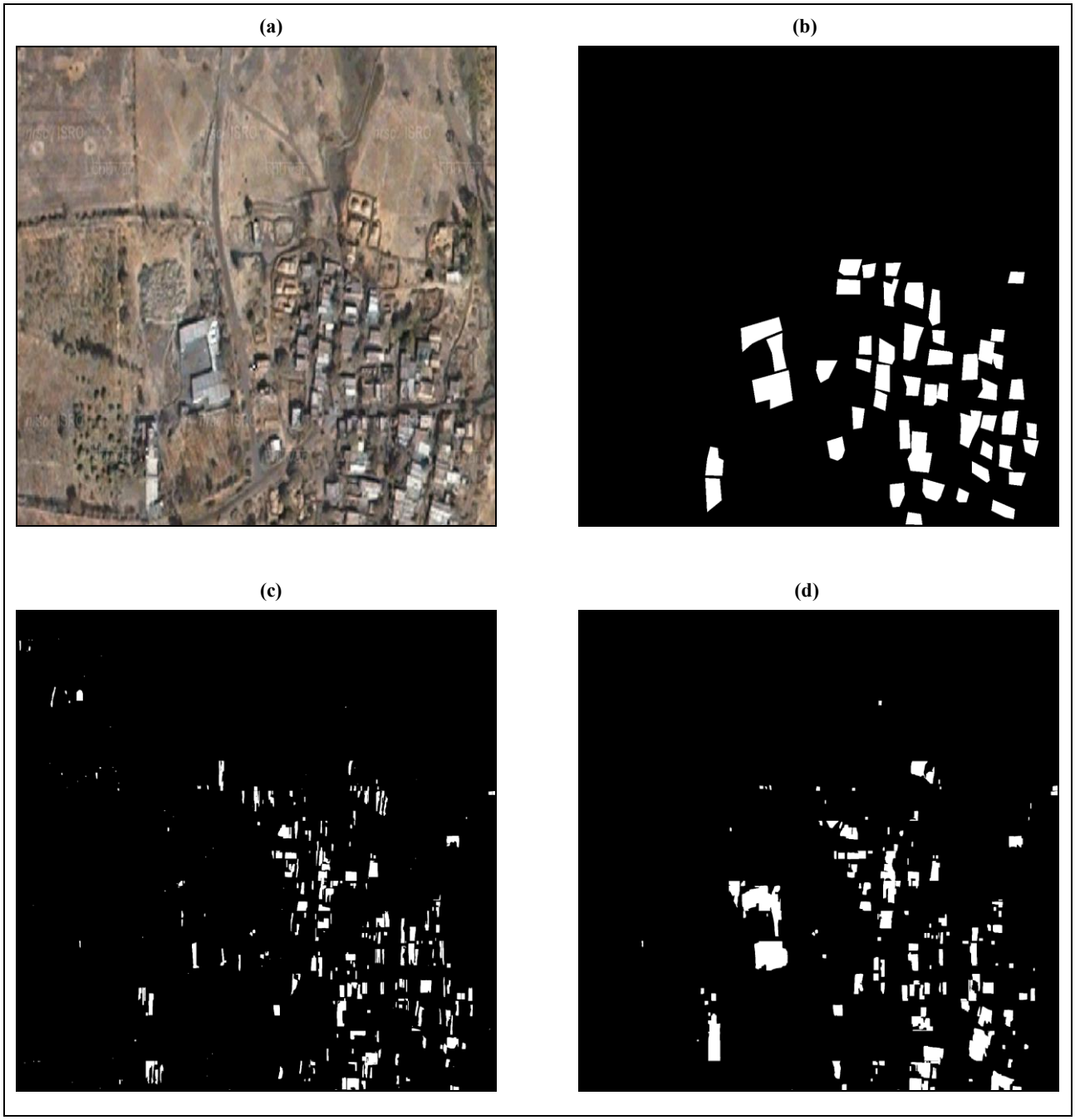


Figure 9. Inference results of Pre-SR-model and Post-SR-model.
(a) Input Image (b) Mask (c) Pre-SR-model output and (d) Post-SR-model output

5. Conclusion

The research presents a successful approach which utilizes deep learning-based segmentation for detecting rural Indian rooftops from satellite images. The study applies AI-based super-resolution to tackle the limited availability of high-resolution open-source data. The YOLOv11l model received better precision and mean Intersection over Union (IoU) results when trained on super-resolution imagery which demonstrated the importance of image enhancement for rooftop segmentation. The future research directions should include dataset expansion for multiple geographic terrains and seasonal changes, enhanced models, and address the issues like shadows and occlusions. The model demonstrates its potential to implement in national programs such as PMGSY National GIS.

Acknowledgment

The authors acknowledge the Centre for Development of Advanced Computing (C-DAC) for the support and funding for this R&D, and Indian Space Research Organisation (ISR0) - Bhuvan for sharing the data.

References

Amo-Boateng, M., Sey, N.E.N., Amproche, A.A. and Domfeh, M.K., 2022. Instance segmentation scheme for roofs in rural areas based on Mask R-CNN. *The Egyptian Journal of Remote Sensing and Space Science*, *25*(2), pp.569-577.

Chen, P., Wang, S., Wang, C., Wang, S., Huang, B., Huang, L. and Zang, Z., 2025. A GAN-enhanced deep learning framework for rooftop detection from historical aerial imagery. *International Journal of Remote Sensing*, pp.1-24.

Gella, G.W., Wendt, L., Lang, S., Tiede, D., Hofer, B., Gao, Y. and Braun, A., 2022. Mapping of dwellings in IDP/refugee settlements from very high-resolution satellite imagery using a mask region-based convolutional neural network. *Remote Sensing*, *14*(3), p.689.

Ghorbanzadeh, O., Crivellari, A., Tiede, D., Ghamisi, P. and Lang, S., 2022. Mapping dwellings in idp/refugee settlements using deep learning. *Remote Sensing*, *14*(24), p.6382.

Mei, Q., Bulatov, D. and Iwaszczuk, D., 2025. Polygonizing roof segments from high-resolution aerial images using YOLOv8-based edge detection. *arXiv preprint arXiv:2503.09187*.

Yang, J., Matsushita, B. and Zhang, H., 2023. Improving building rooftop segmentation accuracy through the optimization of UNet basic elements and image foreground-background balance. *ISPRS Journal of Photogrammetry and Remote Sensing*, *201*, pp.123-137.

Zhao, K., Kang, J., Jung, J. and Sohn, G., 2018. Building extraction from satellite images using mask R-CNN with building boundary regularization. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops* (pp. 247-251).