A Comparative Study of YOLOv8 Segmentation and Detection Models for Urban Object Detection & Classification in SAR Images

Arya Ahadi* Mohammadreza Hassannejad Bibalan†

Abstract

A comparative study of two distinct approaches—detection and segmentation using YOLOv8—for object detection and classification in Synthetic Aperture Radar (SAR) images was conducted using a real-world dataset of Milwaukee from Capella's open data. This study highlights the significance of object detection and classification in urban development and land use, emphasizing their applications in city planning, optimizing land allocation, and ensuring sustainable urban growth. Key challenges in processing SAR images stem from their grayscale nature and speckle noise, which complicate analysis and require advanced algorithms. Results indicate that the segmentation model consistently outperforms the detection model in classification tasks due to its ability to capture detailed spatial information through pixel-wise masks. While the detection model offers speed and efficiency, it struggles with cluttered backgrounds, leading to lower precision. Conversely, the segmentation model, despite requiring more computational resources, provides better localization and accuracy, making it better suited for the complexities of SAR images and crucial for urban development applications.

Keywords: Synthetic Aperture Radar (SAR), YOLOv8, Detection, Segmentation, Classification, Pixel-wise Masks, Urban Development, Grayscale Images, Speckle Noise, Localization, Accuracy

1 Introduction

Synthetic Aperture Radar (SAR) aerial imaging offers several advantages over traditional methods, such as optical or infrared imaging. One key advantage is that SAR operates in the microwave region of the electromagnetic spectrum, allowing it to capture images regardless of weather conditions or lighting. This all-weather, day-and-night capability makes SAR particularly useful in areas with frequent cloud cover or during night operations, where optical systems might fail. Additionally, SAR can penetrate certain materials like vegetation, ice, and even dry soil, providing more detailed information about underlying surfaces [11].

These valuable features make SAR an ideal candidate for advanced image processing techniques, including Artificial Intelligence (AI), Computer Vision, Machine Learning (ML), Deep Learning, and other methods to find and extract valuable and important information from objects in these images.

Today, object detection on SAR images has various potential applications across diverse fields. There is a considerable amount of research on SAR images related to natural disasters and ecosystem monitoring, predicting the impacts on human life, environmental pollution, and devastating economic and social consequences [14] [9] [1]. Another widely studied area in processing these images is ship tracking in the event of a disaster. Quickly identifying the position and status of vessels is vital for rescue teams to deploy efficiently in disaster areas. When responding to emergencies or natural disasters, ship tracking technology plays a critical role in supporting emergency rescue operations and improving the overall resilience of the maritime transportation system [13].

But less attention has been given to urban development applications, and fewer studies have been carried out in this field. Land use and object classification using SAR images are useful for city planning, optimizing land allocation, and ensuring sustainable urban growth. SAR imagery can also be used for the automated detection of urban elements like buildings, roads, and other man-made structures. It can help update maps, monitor construction projects, and detect illegal developments.

Even though there are many applicable features and advantages of SAR images, there are marked barriers to achieving those goals. The main challenges in processing SAR images arise from their grayscale nature and speckle noise [7] [10], which make SAR images complex to analyze, often necessitating advanced algorithms and careful model selection. The aim of this work is to compare two different approaches to identifying urban objects from SAR images: one based on a bounding-box detection model and the other on a segmentation model.

*Department of Electrical Engineering, Imam Khomeini International University, arya.ahadi@hotmail.com
†Department of Electrical Engineering, Imam Khomeini International University, mhbibalan@eng.ikiu.ac.ir

2 Experiment

2.1 Data-set and Labeling

The method was tested on a real-world SAR dataset of Milwaukee, United States, collected by Capella's open data [2]. One of the primary challenges when working with SAR images is their large size and high resolution. For example, the dataset included a TIFF image with dimensions of 25,460x25,596 pixels, which was too large for direct processing. To address this, the image was divided into smaller tiles of 2048x2048 pixels, as shown in Fig. 1, creating a more manageable dataset for training. This method necessitates accurate labeling, which was achieved using the Computer Vision Annotation Tool (CVAT). Given the difficulty of distinguishing object boundaries due to the complexity of the objects, paths and roads were labeled using the polyline labeling method, ensuring high-quality training data, as shown in Fig. 2.

Figure 1: 2048x2048 tiles obtained from original SAR image

2.2 The Computer Vision Model

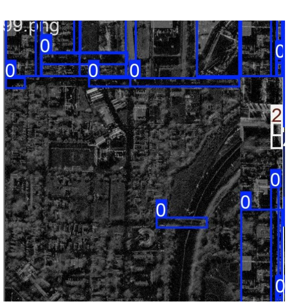
YOLO (You Only Look Once) is an efficient model for object detection and segmentation [15]. YOLO's ability to perform detection and segmentation in real time makes it well-suited for SAR data, which often requires quick and accurate analysis. This speed is crucial when dealing with large-scale SAR datasets, as it reduces computational overhead. Additionally, YOLO's grid-based detection system allows it to detect multiple objects in a single pass, making it effective for handling the cluttered or noisy backgrounds typical of SAR images [3]. The adaptability of YOLO to segment objects in challenging environments enhances its utility in SAR image analysis.

Figure 2: Labeling data-set in CVAT

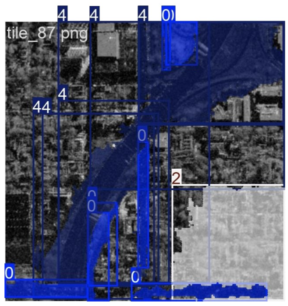
To overcome these challenges, a robust YOLOv8 model with high recall and precision was required. The YOLOv8x model (for bounding-box detection) and the YOLOv8x-seg model (for segmentation) were selected from the various YOLOv8 models. YOLOv8x-seg, the extra-large variant, is known for its high accuracy, with 54.7 mAPbox, 43.8 mAPmask, 62.1 million parameters, and 319 GFLOPS. This model achieves great precision due to its complexity and ability to capture intricate patterns, but it requires significantly high computational resources. Two different approaches were utilized: one using a bounding-box detection model and the other employing a segmentation model. Both models were trained on the gathered dataset tiles for 100 epochs to obtain the results shown in Fig. 3. The obtained tiles of SAR images included about 50 tiles, with 40 used for training and 10 for validation.

3 Results

Both the detection and segmentation models exhibit reductions in training losses, but the segmentation model demonstrates greater stability in validation losses, especially in box loss. However, classification loss in the segmentation model fluctuates more, particularly in the early training epochs. One reason for this fluctuation in the segmentation model is its finer pixel-level investigation, which leads to the extraction of richer features. Notably, the authors of this article also achieved promising results using the segmentation technique [5]. The ability to learn finer pixel-level details is particularly beneficial in SAR images, where object shapes, textures,

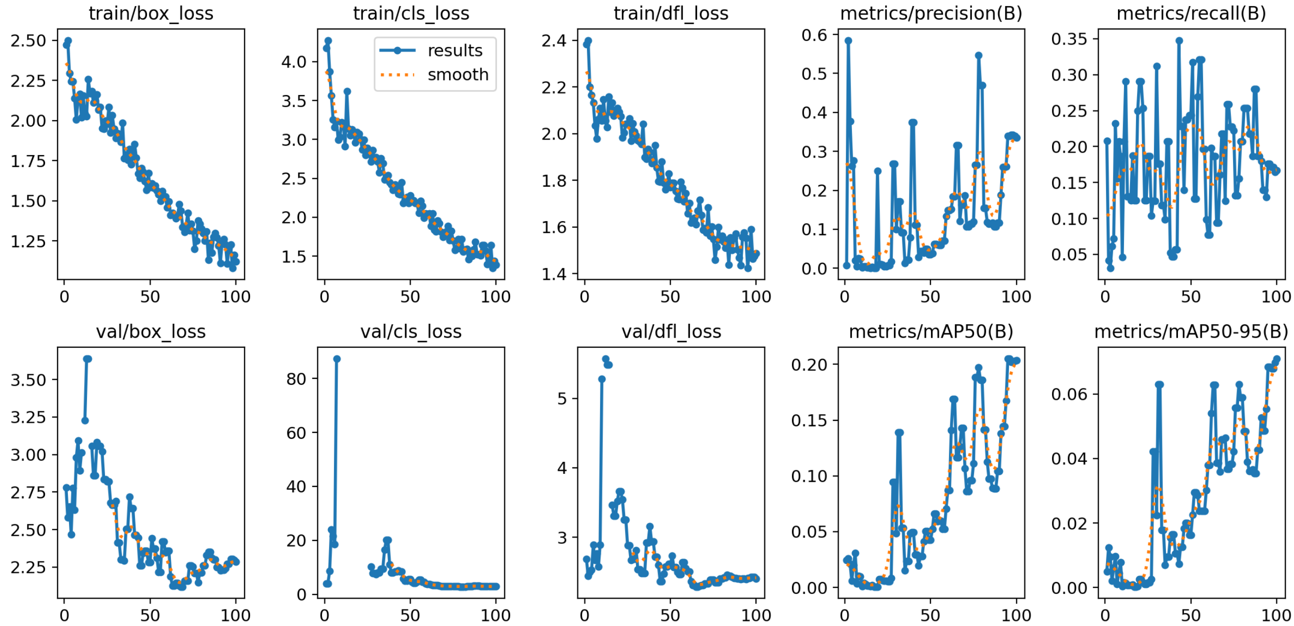


(a) Detection on bounding-box model

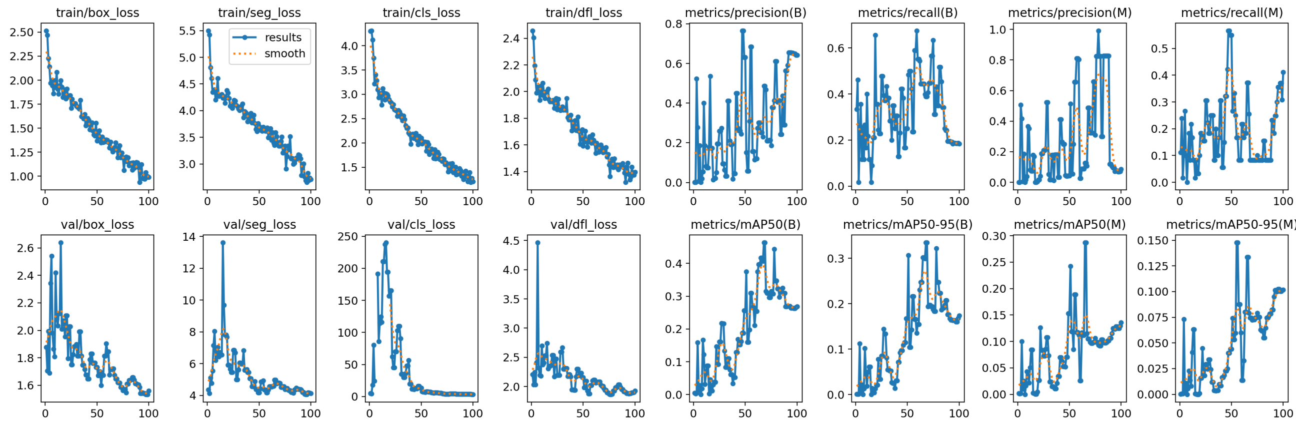


(b) Detection on segmentation model

Figure 3: Detected object during validation process on both models



(a) Detection model metrics results



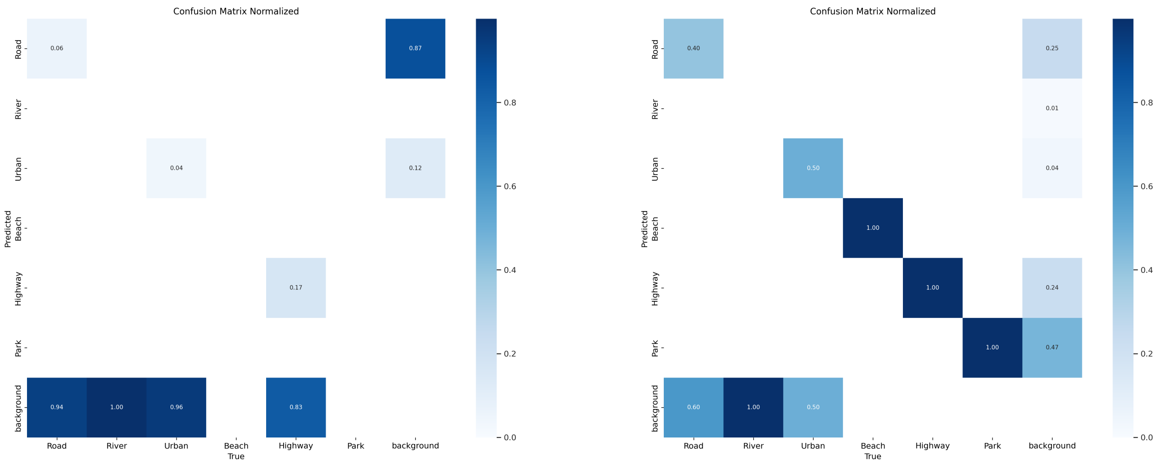
(b) Segmentation model metrics results

Figure 4: Detection and Segmentation model training metrics results

and boundaries can be challenging to identify due to noise and reflections [6].

The segmentation loss in the segmentation model is relatively high and noisy in validation during the first epochs. However, this loss still decreases over time, suggesting progress despite the noise. The segmentation model focuses on predicting object boundaries more precisely, which is particularly challenging in SAR images where boundaries may not be clearly defined—a challenge also encountered by Huang *et al.* [6]. The constant refinement of these boundaries leads to fluctuations in loss as the model improves its understanding of pixel-level features.

Precision and recall were some of the indices used to assess the identification of relevant vs. irrelevant items. The segmentation model consistently outperforms the detection model in precision and recall, particularly in later epochs. This is especially important for SAR data, where identifying objects with high precision is challenging due to cluttered backgrounds and noisy data. As Marmanis *et al.* also demonstrated in their boundary detection work using a segmentation model, this approach provides better localization and delineation of object boundaries due to its pixel-level attention [8]. Although the precision and recall of the segmentation model decrease to some extent in the latest epochs, they



(a) Detection model confusion matrix

(b) Segmentation model confusion matrix

Figure 5: Confusion matrices of both models

still remain higher than those of the detection model. This decrease in precision and recall may be due to overfitting [12], which can be mitigated by various methods, such as increasing the dataset size.

Another metric used to measure the performance of object detection is mAP. The mAP50(B) and mAP50-95(B) values for the detection model fluctuate significantly during training. Initially, performance is low (mAP50 starts at 0.05) but gradually improves to 0.2 for mAP50-95 by the end of training. The segmentation model achieves higher overall mAP(B) values. By the end of training, the mAP50(B) reaches close to 0.4, and the mAP50-95(B) approaches 0.3. This suggests that the segmentation model captures more accurate predictions across various IoU thresholds, making it better suited for the complex nature of SAR images due to richer supervision at the pixel level [4]. Richer supervision helps improve overall accuracy and enhances the model's ability to detect objects at multiple IoU thresholds, leading to higher mAP scores.

Figure 5 shows the confusion matrices of both models. As the figure depicts, the segmentation model has overall better performance compared to the detection model. The segmentation model predicted more objects correctly in each class. Its confusion matrix has more diagonal values compared to the detection model, which represent correctly classified instances for each class, especially for Beach, Highway, and Park, with a detection rate of 1.00.

4 Conclusion

This study compares two distinct approaches—detection and segmentation—for object detection and classification in SAR images using YOLOv8 models. The findings reveal that segmentation models consistently outperform detection models in classification tasks, largely due to their ability to capture more detailed spatial information and context. Unlike detection models, which rely on bounding boxes, segmentation models generate pixel-wise masks, allowing for a more nuanced understanding of object boundaries and interactions.

References

[1] K. Arai, Y. Nakaoka, and H. Okumura Method for Disaster Area Detection with Just One SAR Data Acquired on the Day After Earthquake Based on YOLOv8. *International Journal of Advanced Computer Science and Applications (IJACSA)*, 15(3), 2024. `http://dx.doi.org/10.14569/IJACSA.2024.0150344`.

[2] Capella Space, Capella Space Open Data Gallery, `https://www.capellaspace.com/earth-observation/gallery`.

[3] Y. Ding, Y. Li, and W. Yu, Learning from label proportions for SAR image classification, *EURASIP Journal on Advances in Signal Processing*, 2017(41), 2017. `https://doi.org/10.1186/s13634-017-0478-8`.

[4] L. Dhruthi, K. Praveen, Megharaj, P. Pranav, K. Niharika, K. Asha Rani, and P. G. S., State-of-the-Art Object Detection: An Overview of YOLO Variants and their Performance, in *Proceedings of the International Conference on Smart Electronics and Communication (ICOSEC)*, pp. 1018-1024, 2023. `https://doi.org/10.1109/icosec58147.2023.10276030`.

[5] P. Gamba and M. Aldrighi, "SAR Data Classification of Urban Areas by Means of Segmentation Techniques and Ancillary Optical Data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, no. 4, pp. 1140-1148, Aug. 2012. `https://doi.org/10.1109/JSTARS.2012.2195774`.

[6] Z. Huang, M. Datcu, Z. Pan, and B. Lei, Deep SAR-Net: Learning objects from signals, *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 161, pp. 179-193, 2020. https://doi.org/10.1016/j.isprsjprs.2020.01.016.

[7] S. Liu, Z. Cao, and H. Yang, Information Theory-Based Target Detection for High-Resolution SAR Image, *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 3, pp. 404-408, March 2016. https://doi.org/10.1109/LGRS.2016.2516535.

[8] D. Marmanis, K. Schindler, J. D. Wegner, S. Galliani, M. Datcu, and U. Stilla, "Classification with an edge: Improving semantic image segmentation with boundary detection," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 135, pp. 158-172, 2018. https://doi.org/10.1016/j.isprsjprs.2017.11.009.

[9] A. Nasraoui, T. Selmi, and Z. Hajaiej, Integration of Deep Learning Object Detection Techniques and Drone Technology for Disaster Monitoring and Detection, in *Proceedings of the International Conference on Advanced Systems and Emergent Technologies (IC_ASET)*, 2024. https://doi.org/10.1109/ic_aset61847.2024.10596153.

[10] S. Tebaldini, M. Manzoni, D. Tagliaferri, M. Rizzi, A. V. Monti-Guarnieri, C. M. Prati, U. Spagnolini, M. Nicoli, I. Russo, and C. Mazzucco, Sensing the Urban Environment by Automotive SAR Imaging: Potentials and Challenges, *Remote Sensing*, vol. 14, p. 3602, 2022. https://doi.org/10.3390/rs14153602.

[11] A. Tsokas, M. Rysz, P. M. Pardalos, and K. Dipple, SAR data applications in earth observation: An overview, *Expert Systems with Applications*, vol. 205, 2022, 117342. https://doi.org/10.1016/j.eswa.2022.117342.

[12] Y. Toledo, T. Almeida, F. Bernardini, and E. de Oliveira Andrade, A case study about overfitting in multiclass classifiers using Convolutional Neural Networks, in *Proceedings of the Brazilian Symposium on Artificial Intelligence (ENIAC)*, 2019. https://doi.org/10.5753/ENIAC.2019.9335.

[13] M. Yasir, S. Liu, M. Xu, S. Hui, N. Shah, X. Zhang, T. Arife, and C. Isiacik, YOLOv8-BYTE: Ship tracking algorithm using short-time sequence SAR images for disaster response leveraging GeoAI, *ITC Journal*, 2024. https://doi.org/10.1016/j.jag.2024.103771.

[14] Y. J. Yang, S. Singha, and R. Mayerle, A deep learning based oil spill detector using Sentinel-1 SAR imagery, *International Journal of Remote Sensing*, vol. 43, no. 11, pp. 4287–4314, 2022. https://doi.org/10.1080/01431161.2022.2109445.

[15] G. Jocher, A. Chaurasia, and J. Qiu, Ultralytics YOLOv8, version 8.0.0, 2023. https://github.com/ultralytics/ultralytics, ORCID: 0000-0001-5950-6979, 0000-0002-7603-6750, 0000-0003-3783-7069. Licensed under AGPL-3.0.