SITEKNIK
Sistem Informasi, Teknik dan Teknologi Terapan
E-ISSN: 3032-3991 P-ISSN: 3090-1626
Vol. 3, No. 3 July 2026. Pages. 108-115

Fine-Tuning Panoptic FPN with ResNet-50 for Maritime Obstacle Detection on the LaRS Dataset

Istifa Shania Putri[1], Sugih Ahmad Fauzan[2], Mega Fitri Yani[3], Cindy Muhdiantini[4]
istifashaniaputri@telkomuniversity.ac.id[1], 23523038@std.stei.itb.ac.id[2],
megafy@telkomuniversity.ac.id[3], cindymuhdiantinicm@telkomuniversity.ac.id[4]
[134] Information System, Telkom University, Indonesia
[2] School of Electrical and Informatics Engineering, Bandung Institute of Technology, Indonesia

Keywords: Panoptic Segmentation, Fine-Tuning, Mask R-CNN, Feature Pyramid Networks, Maritime Obstacle Detection, LaRS Dataset

Submitted: 29/03/2026
Revised: 13/06/2026
Accepted: 20/06/2026

Abstract

Maritime obstacle detection is a critical challenge for Unmanned Surface Vehicles (USVs) operating in complex and dynamic environments. This study investigates the effectiveness of fine-tuning Panoptic FPN, a Mask R-CNN-based architecture augmented with Feature Pyramid Networks, for panoptic segmentation on the LaRS (Lake, River, Seas) dataset. Unlike prior work that explored model comparisons broadly, this research focuses specifically on the impact of hyperparameter tuning and backbone selection on maritime panoptic segmentation performance. Through systematic ablation studies, we demonstrate that adjusting the learning rate to 0.002 and the gamma decay factor to 0.2 yields significant improvements. Our fine-tuned Panoptic FPN with a ResNet-50 backbone achieves a Panoptic Quality (PQ) of 45.31%, surpassing the previous state-of-the-art Mask2Former Swin-B (41.7%) by 3.61 percentage points. Notably, ResNet-50 outperforms the deeper ResNet-101 backbone (36.47% PQ), suggesting that heavier architectures may overfit on domain-specific maritime datasets. Furthermore, Panoptic FPN requires only 8 hours of training compared to approximately 2 days for Mask2Former Swin-L, demonstrating superior computational efficiency. These findings highlight that targeted fine-tuning of lightweight architectures can outperform larger transformer-based models in maritime panoptic segmentation tasks.

Corresponding Author:
Istifa Shania Putri
Information System, School of Industrial Engineering
Gedung Bangkit Telkom University Jl. Telekomunikasi , Terusan Buah Batu Indonesia 40257, Bandung, Indonesia
Email: istifashaniaputi@telkomuniversity.ac.id

INTRODUCTION

Maritime accidents are frequently attributed to human factors, underscoring the importance of improving detection and awareness systems in maritime operations (Sánchez-Beaskoetxea et al., 2021). Unmanned Surface Vehicles (USVs) have emerged as a promising solution for automating maritime coordination tasks, but their development hinges on robust obstacle detection capabilities (Jin et al., 2024). Maritime environments present unique challenges for computer vision systems: the water surface produces strong reflections and sun glare, lighting conditions change rapidly, and dynamic obstacles vary greatly in size and appearance. Studies have shown that segmentation accuracy can drop by as much as 66% under strong reflection conditions compared to moderate conditions (Žust et al., 2023). These factors make comprehensive and reliable scene understanding a fundamental prerequisite for safe autonomous maritime navigation.

Panoptic segmentation, which unifies semantic segmentation (category-level pixel labeling) and instance segmentation (individual object identification), offers a compelling framework for maritime obstacle detection (Li, X., & Chen, 2022). Panoptic segmentation produces a single coherent scene understanding, classifying both background regions such as water and sky and foreground objects such as ships and debris at the pixel level (Cheng et al., 2020). Research has demonstrated that panoptic segmentation provides more comprehensive scene understanding than using separate segmentation models (Jaus et al., 2023), and exhibits greater robustness under varying maritime conditions (Žust et al., 2023) (Dou et al., 2024).

The LaRS (Lake, River, Seas) dataset (Žust et al., 2023) is the most comprehensive publicly available benchmark for maritime panoptic segmentation, providing panoptic annotations for 4,006 key frames across diverse maritime environments. Recent benchmark studies on LaRS have evaluated several architectures including Panoptic FPN (Kirillov, A., et al., 2019) and Mask2Former (Cheng, B., et al., 2022), establishing Mask2Former Swin-B as the previous state-of-the-art with a Panoptic Quality (PQ) of 41.7%. Prior work on panoptic segmentation for maritime autonomous vehicles has also demonstrated its potential in related domains (Qiao, D., et al., 2022). However, these studies have not fully explored whether systematic fine-tuning of hyperparameters and careful backbone selection can unlock further performance gains from existing architectures.

Panoptic FPN (Kirillov, A., et al., 2019) integrates Mask R-CNN with a Feature Pyramid Network (FPN), enabling simultaneous instance and semantic segmentation by leveraging multi-scale feature representations. The FPN architecture is particularly well-suited for handling objects of diverse sizes, a key requirement in maritime settings where obstacles range from large vessels to small buoys (Lin, T.-Y., et al., 2016). Despite its effectiveness on general-purpose benchmarks, the potential of fine-tuning Panoptic FPN specifically for maritime datasets remains underexplored. In particular, the effects of learning rate scheduling, gamma decay, and backbone depth on maritime panoptic segmentation quality have not been systematically investigated.

This study addresses this gap by presenting a focused experimental investigation into fine-tuning Panoptic FPN for maritime obstacle segmentation on the LaRS dataset. The central research questions are: (1) Can fine-tuning of hyperparameters, particularly learning rate and gamma, improve Panoptic FPN performance on the LaRS dataset? (2) How does backbone choice, specifically ResNet-50 versus ResNet-101, affect segmentation quality in this domain? (3) Does a lighter, well-tuned model offer a favorable trade-off between accuracy and computational efficiency compared to larger transformer-based models? Through systematic ablation studies, we show that our fine-tuned Panoptic FPN with ResNet-50 achieves a PQ of 45.31%, surpassing the previous state-of-the-art by 3.61 percentage points while requiring significantly less training time and computational resources.

RESEARCH METHODS

Dataset

Experiments are conducted on the LaRS (Lake, River, Seas) dataset (Žust et al., 2023), a publicly available panoptic segmentation benchmark specifically designed for maritime obstacle detection. Among existing maritime obstacle datasets (Bovcon, B., et al., 2022), LaRS is the most comprehensive in terms of panoptic annotations. The dataset consists of 4,006 key frames drawn from diverse maritime environments including lakes, rivers, and open seas. Each frame is annotated at the pixel level with 11 semantic categories: 8 dynamic obstacle instance classes (boat/ship, row boat, paddle board, buoy, swimmer, animal, float, other) and 3 stuff classes (water, sky, coastline/static obstacle). Additionally, each frame is labeled with 19 global scene attributes describing environmental conditions such as lighting, reflection intensity, weather, and surface texture. Images have a resolution of 2,208 x 1,242 pixels. The dataset is split into 2,605 training images, 1,023 test images, and 198 validation images. The LaRS dataset is particularly valuable due to its panoptic annotations, scene diversity, and rigorous annotation quality assurance, including manual review and correction of approximately 3,600 labeling errors.

Model Architecture: Panoptic FPN

The proposed framework is based on Panoptic FPN (Kirillov, A., et al., 2019), which extends the Mask R-CNN architecture with a Feature Pyramid Network to perform simultaneous instance and semantic segmentation within a unified framework. The model comprises four main components. First, the backbone (ResNet-50 or ResNet-101) extracts hierarchical feature maps from input images across multiple resolution levels (Ashrafi, S. S., et al., 2023). Second, the FPN constructs a multi-scale feature pyramid by combining top-down and bottom-up pathways with lateral connections (Lin, T.-Y., et al., 2016), enabling the model to detect objects of diverse scales. Third, the Region Proposal Network (RPN) generates candidate object regions using anchor boxes of various scales and aspect ratios; proposals are then processed via RoI Align to preserve spatial precision. Fourth, function branches include classification (SoftMax), bounding box regression, and binary mask prediction heads for instance segmentation, alongside a semantic segmentation head operating on FPN levels 0 through 4 using Group Normalization and Cross-Entropy Loss.

The panoptic fusion head merges instance segmentation outputs with semantic segmentation predictions to produce the final panoptic output (Elharrouss, O., et al., 2021). During inference, non-maximum suppression (NMS) is applied with an IoU threshold of 0.5 and a score threshold of 0.6, limiting detections to 100 per image. For panoptic fusion, an overlap threshold of 0.5 and a minimum stuff area of 4,096 pixels are applied.

Input preprocessing includes normalization using ImageNet mean and standard deviation, BGR-to-RGB conversion, and padding to ensure image dimensions are divisible by 32. The model was implemented using the MMDetection library with PyTorch 2.2.0+cu118, and experiments were conducted on a server equipped with an Intel Xeon Silver 4208 processor, an NVIDIA RTX A5000 GPU (24 GB VRAM), and 256 GB DDR4 RAM.

Evaluation Metrics

Model performance is evaluated using the Panoptic Quality (PQ) metric and its decomposed components, Segmentation Quality (SQ) and Recognition Quality (RQ), as defined in standard panoptic segmentation benchmarks (Milioto, A., et al., 2020) (Zendel, O., et al., 2022). These metrics are widely adopted in the panoptic segmentation literature (Li, X., & Chen, 2022). PQ is computed as the product of SQ and RQ across all matched segments, where a segment pair is considered matched if their Intersection-over-Union (IoU) exceeds 0.5. These metrics are reported separately for all classes (All), thing classes (Th), and stuff classes (St). PQ captures both segmentation accuracy and recognition accuracy in a single unified metric, making it appropriate for evaluating the combined panoptic output.

Experimental Design: Hyperparameter Fine-Tuning

The baseline Panoptic FPN configuration used in prior LaRS benchmark studies employs ResNet-50 with a learning rate of 0.001 and a gamma decay factor of 0.1 (Žust et al., 2023). Building on this baseline, we conducted ablation experiments to investigate the impact of two key hyperparameters: learning rate and gamma. The learning rate was increased to 0.002 and the gamma value was adjusted to 0.2. The MultiStepLR learning rate scheduler was used, which reduces the learning rate by a multiplicative factor of gamma at predefined milestone epochs (Sellat, Q., et al., 2022). This scheduling strategy allows the model to make larger parameter updates early in training while progressively fine-tuning as training progresses. We additionally compared the ResNet-50 and ResNet-101 backbones to assess whether deeper feature extractors provide performance benefits on the LaRS dataset.

RESULTS AND DISCUSSION

Benchmarking Results

Table 1 presents the panoptic segmentation performance of our fine-tuned Panoptic FPN models alongside baseline and comparison architectures on the LaRS test set. The fine-tuned Panoptic FPN with ResNet-50 achieves the highest overall PQ of 45.31%, representing an improvement of 3.61 percentage points over the previous state-of-the-art Mask2Former Swin-B (41.7%) reported by Žust et al. (Žust et al., 2023). This result is particularly significant because Panoptic FPN with ResNet-50 is a substantially lighter and more computationally efficient architecture. Training required approximately 8 hours, compared to approximately 2 days for Mask2Former Swin-L. In terms of model complexity, Panoptic FPN with ResNet-50 contains approximately 46 million parameters and executes 0.288 TFLOPS, whereas Mask2Former Swin-L contains approximately 216 million parameters and 0.913 TFLOPS, roughly five times more complex.

Table 1. Panoptic Quality Comparison on LaRS Dataset

Architecture	Backbone	PQ (%)			SQ (%)			RQ (%)		
		All	Th	St	All	Th	St	All	Th	St
MaX-DeepLab (Wang, H., et al., 2021)	MaX-S	31.9	9.5	91.7	71.3	62.5	73.7	36.1	13.4	96.6
Panoptic Deeplab (Cheng et al., 2020)	ResNet-50	34.7	13.4	91.4	69.5	60.0	71.3	40.3	19.3	96.2
Panoptic FPN (Žust et al., 2023)	ResNet-50	40.1	21.7	89.3	73.5	66.1	93.1	46.9	28.6	95.8
Panoptic FPN (Žust et al., 2023)	ResNet-101	38.7	19.7	89.4	73.6	66.1	93.5	45.0	26.1	95.5
Mask2Former (Žust et al., 2023)	ResNet-50	37.6	17.0	92.4	71.3	62.4	95.0	43.7	23.6	97.3
Mask2Former (Žust et al., 2023)	ResNet-101	37.2	16.3	92.8	71.4	62.3	95.5	43.0	22.7	97.1
Mask2Former (Žust et al., 2023)	Swin-T	39.2	18.8	93.7	72.2	63.5	95.4	45.5	25.8	98.1

Table 1. Panoptic Quality Comparison on LaRS Dataset (Continued)

Mask2Former (Žust et al., 2023) (prev. SotA)	Swin-B	41.7	21.8	94.7	78.2	71.5	96.2	48.5	29.7	98.5
Mask2Former (ours)	Swin-L	40.42	20.38	93.86	70.81	61.36	96.00	47.10	28.12	97.72
Panoptic FPN (ours)	ResNet-101	36.47	16.42	89.95	71.79	63.68	93.43	43.03	23.12	96.14
Panoptic FPN (ours, best)	ResNet-50	45.31	28.40	90.36	72.57	64.69	93.59	54.25	38.41	96.49

Th = Thing classes; St = Stuff classes; SotA = State-of-the-Art

Analyzing the class-level results, the best-performing model achieves a PQ of 90.36% on stuff classes (water, sky, coastline), indicating strong performance in delineating static background regions. However, the PQ for thing classes (dynamic obstacles) is 28.40%, and the RQ for thing classes is 38.41%, highlighting that detecting and accurately segmenting dynamic obstacles remains a significant challenge. This asymmetry between stuff and thing performance is consistent with the inherent difficulty of maritime obstacle detection, where dynamic objects are visually diverse, frequently small, and appear against cluttered reflective backgrounds (Žust et al., 2023).

Ablation Study: Backbone and Hyperparameter Effects

Table 2 presents the ablation study results comparing backbone architectures and hyperparameter configurations. Contrary to the common intuition that deeper architectures yield better performance, the ResNet-101 backbone underperforms ResNet-50 under the same fine-tuned hyperparameter setting, achieving a PQ of only 36.47% compared to 45.31% for ResNet-50. This can be attributed to the tendency of deeper models to overfit when trained on domain-specific datasets of moderate size. ResNet-101's additional layers enable it to capture more complex pattern representations (Ashrafi, S. S., et al., 2023), but this capacity becomes a liability when the training distribution is relatively constrained, as in the LaRS dataset with only 2,605 training images.

Table 2. Ablation Study: Backbone and Hyperparameter Configurations on LaRS Dataset

Backbone	LR	Gamma	PQ All	PQ Th	PQ St	SQ All	RQ All	Train Time	Params (M)
ResNet-50 (baseline (Žust et al., 2023))	0.001	0.1	40.1	21.7	89.3	73.5	46.9	~8 hrs	46
ResNet-101 (ours)	0.002	0.2	36.47	16.42	89.95	71.79	43.03	~8 hrs	65
ResNet-50 (ours — best)	0.002	0.2	45.31	28.40	90.36	72.57	54.25	~8 hrs	46

LR = Learning Rate; Params = Model Parameters (millions)

The effect of hyperparameter tuning is substantial. Increasing the learning rate from 0.001 to 0.002 and the gamma from 0.1 to 0.2 resulted in a PQ improvement of 5.21 percentage points for the ResNet-50 backbone (from 40.1% to 45.31%). This suggests that the original learning rate was too conservative for the maritime domain. The higher gamma value allows the learning rate to remain relatively elevated for longer during training, enabling more thorough exploration of the parameter space before convergence (Sellat, Q., et al., 2022). This is consistent with findings in the Ada-Segment study, where adaptive loss and learning rate adjustments yielded a 2.7% PQ improvement on COCO (Gengwei, Z., et al., 2021), confirming that domain-specific hyperparameter tuning is a reliable strategy for improving panoptic segmentation models.

Comparison with Transformer-Based Models

Our experiments also evaluate Mask2Former Swin-L, a transformer-based architecture utilizing the Swin-L backbone with 192 hidden channels in the first stage (Liu, Z., et al., 2021). Despite its substantially larger parameter count (~216 million) and longer training time (~2 days), Mask2Former Swin-L achieves a PQ of only 40.42%, lower than our fine-tuned Panoptic FPN with ResNet-50. This result challenges the assumption that larger transformer-based models automatically outperform CNN-based approaches in domain-specific settings. The Swin-L model's performance also falls short of the smaller Mask2Former Swin-B (41.7%), suggesting that the larger model may introduce overfitting on the relatively small LaRS training set.

The superiority of Panoptic FPN over Mask2Former in this setting can be explained by several factors. The FPN architecture explicitly addresses multi-scale feature extraction (Lin, T.-Y., et al., 2016), which is well-suited to the diverse obstacle sizes in maritime scenes. Additionally, Panoptic FPN's joint training of instance and semantic heads may yield stronger regularization effects compared to transformer decoders trained with masked attention (Cheng, B., et al., 2022). The practical efficiency of Panoptic FPN also makes iterative hyperparameter search more tractable, enabling more effective domain adaptation. These findings are consistent with prior observations that adaptation strategies for panoptic models can yield improvements beyond what raw model size provides (Gengwei, Z., et al., 2021) (Darbyshire, M., et al., 2022).

CONCLUSIONS AND SUGGESTIONS

Conclusion

This study demonstrates that targeted fine-tuning of Panoptic FPN is a highly effective strategy for improving maritime obstacle panoptic segmentation on the LaRS dataset. By systematically adjusting the learning rate to 0.002 and the gamma decay factor to 0.2, we achieved a Panoptic Quality of 45.31% with a ResNet-50 backbone, surpassing the previous state-of-the-art Mask2Former Swin-B (Žust et al., 2023) by 3.61 percentage points while requiring only 8 hours of training compared to approximately 2 days for transformer-based competitors. These results establish a new benchmark on the LaRS dataset.

A key finding is that backbone depth does not linearly correlate with segmentation quality in domain-specific settings: the deeper ResNet-101 consistently underperforms ResNet-50, likely due to overfitting on the relatively small LaRS training set. This finding has important practical implications like practitioners should evaluate lighter architectures with careful hyperparameter tuning before resorting to larger models. The significant efficiency advantage of Panoptic FPN also makes it more viable for real-world USV deployment scenarios where computational resources and iteration speed are constrained (Jin, et al., 2024).

Suggestion

Despite the strong overall performance, the recognition quality of dynamic thing classes remains a notable limitation (RQ = 38.41% for the best model). Future research should focus on strategies to improve detection of dynamic maritime obstacles, particularly smaller and rarer categories such as swimmers, buoys, and marine animals. Potential directions include data augmentation strategies tailored to maritime imagery, class-balancing techniques, and auxiliary objectives emphasizing rare obstacle categories. The Ada-Segment approach of adaptive multi-loss weighting (Gengwei, Z., et al., 2021) and FPN improvements such as those demonstrated in brain tumor segmentation tasks (Sun, H., et al., 2023) could also be explored in the maritime domain.

Additionally, future work could investigate the integration of temporal video information to improve obstacle tracking consistency across frames, which is particularly relevant for USV operations (Jin et al., 2024). Exploring lightweight transformer architectures that can be fine-tuned efficiently on small domain-specific datasets, or combining CNN-based multi-scale features with attention mechanisms for improved contextual understanding, may yield further improvements. Evaluation under specific challenging scene attributes from the LaRS dataset, such as strong reflections, nighttime conditions, and rain, would provide more granular insight into model robustness in operationally critical scenarios (Žust et al., 2023).

ACKNOWLEDGMENTS

The authors would like to extend their sincere appreciation to the Server Riset Akademik Tambora at Institut Teknologi Bandung for providing the computational resources and infrastructure that enabled the training and evaluation experiments in this study. Their support was instrumental in the successful completion of this research.

REFERENCES

Sánchez-Beaskoetxea, J., Basterretxea-Iribar, I., Sotés, I., & Maruri Machado, M. de las M. (2021). Human error in marine accidents: Is the crew normally to blame? Maritime Transport Research, 2, 100016. https://doi.org/10.1016/j.martra.2021.100016

Jin, J., Liu, D., Li, F., Dai, Y., Li, L., & Ma, Y. (2024). Wide baseline stereovision based obstacle detection for unmanned surface vehicles. Signal, Image and Video Processing. https://doi.org/10.1007/s11760-024-03098-0

Žust, L., Perš, J., & Kristan, M. (2023). LaRS: A diverse panoptic maritime obstacle detection dataset and benchmark. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) (pp. 20247–20257). https://doi.org/10.1109/ICCV51070.2023.01857

Jaus, A., Yang, K., & Stiefelhagen, R. (2023). Panoramic panoptic segmentation: Insights into surrounding parsing for mobile agents via unsupervised contrastive learning. IEEE Transactions on Intelligent Transportation Systems, 24(4), 4438–4453. https://doi.org/10.1109/TITS.2022.3232897

Dou, Y., et al. (2024). PanopticUAV: Panoptic segmentation of UAV images for marine environment monitoring. Computer Modeling in Engineering & Sciences, 138(1), 1001–1014. https://doi.org/10.32604/cmes.2023.027764

Cheng, B., et al. (2020). Panoptic-DeepLab: A simple, strong, and fast baseline for bottom-up panoptic segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 12472–12482). https://doi.org/10.1109/CVPR42600.2020.01249

Kirillov, A., Girshick, R., He, K., & Dollár, P. (2019). Panoptic feature pyramid networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 6392–6401). https://doi.org/10.1109/CVPR.2019.00656

Qiao, D., Liu, G., Li, W., Lyu, T., & Zhang, J. (2022). Automated full scene parsing for marine ASVs using monocular vision. Journal of Intelligent & Robotic Systems, 104(2), 37. https://doi.org/10.1007/s10846-021-01543-7

Cheng, B., Misra, I., Schwing, A. G., Kirillov, A., & Girdhar, R. (2022). Masked-attention mask transformer for universal image segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 1280–1289). https://doi.org/10.1109/CVPR52688.2022.00135

Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2016). Feature pyramid networks for object detection. arXiv:1612.03144. https://arxiv.org/abs/1612.03144

Liu, Z., et al. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. arXiv:2103.14030. https://arxiv.org/abs/2103.14030

Li, X., & Chen, D. (2022). A survey on deep learning-based panoptic segmentation. Digital Signal Processing, 120, 103283. https://doi.org/10.1016/j.dsp.2021.103283

Gengwei, Z., Gao, Y., Xu, H., Zhang, H., Li, Z., & Liang, X. (2021). Ada-Segment: Automated multi-loss adaptation for panoptic segmentation. Proceedings of the AAAI Conference on Artificial Intelligence, 35, 3333–3341. https://doi.org/10.1609/aaai.v35i4.16445

Milioto, A., Behley, J., McCool, C., & Stachniss, C. (2020). LiDAR panoptic segmentation for autonomous driving. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (pp. 8505–8512). https://doi.org/10.1109/IROS45743.2020.9340837

Zendel, O., Schörghuber, M., Rainer, B., Murschitz, M., & Beleznai, C. (2022). Unifying panoptic segmentation for autonomous driving. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 21319–21328). https://doi.org/10.1109/CVPR52688.2022.02066

Wang, H., Zhu, Y., Adam, H., Yuille, A., & Chen, L.-C. (2021). MaX-DeepLab: End-to-end panoptic segmentation with mask transformers. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 5459–5470). https://doi.org/10.1109/CVPR46437.2021.00542

Ashrafi, S. S., Shokouhi, S. B., & Ayatollahi, A. (2023). Still image action recognition based on interactions between joints and objects. Multimedia Tools and Applications, 82(17), 25945–25971. https://doi.org/10.1007/s11042-023-14350-z

Elharrouss, O., Al-Maadeed, S. A., Subramanian, N., Ottakath, N., Almaadeed, N., & Himeur, Y. (2021). Panoptic segmentation: A review. arXiv:2111.10250. https://arxiv.org/abs/2111.10250

Sellat, Q., Bisoy, S. K., & Priyadarshini, R. (2022). Semantic segmentation for self-driving cars using deep learning: A survey. In Cognitive Big Data Intelligence with a Metaheuristic Approach (pp. 211–238). Academic Press. https://doi.org/10.1016/B978-0-323-85117-6.00002-9

Darbyshire, M., Sklar, E., & Parsons, S. (2023). Hierarchical Mask2Former: Panoptic segmentation of crops, weeds and leaves. In Proceedings of the IEEE International Conference on Robotics and Biomimetics (ROBIO).

Sun, H., et al. (2023). Brain tumor image segmentation based on improved FPN. BMC Medical Imaging, 23(1), 172. https://doi.org/10.1186/s12880-023-01131-1

Bovcon, B., Muhovič, J., Vranac, D., Mozetič, D., Perš, J., & Kristan, M. (2022). MODS: A USV-oriented object detection and obstacle segmentation benchmark. IEEE Transactions on Intelligent Transportation Systems, 23(8), 13403–13418. https://doi.org/10.1109/TITS.2021.3124192