

Dice-Tversky-Based Instance Segmentation of Small Defects in Automotive Sealing Strips

Ye Zhu

ShanDong University of Science and Technology, Qingdao 266590, China

Abstract

Small surface defects in automotive sealing strips, including cracks, indentations and local surface anomalies, usually occupy only a small foreground area and have fragmented boundaries. Conventional mask losses can therefore be dominated by background pixels, leading to missed small regions or insufficient boundary fitting. To address this issue, this study improves YOLOv8s-seg by introducing Dice loss and Tversky loss into the original BCE mask loss, forming a Dice-Tversky mask overlap constraint. The Dice term strengthens region-level overlap supervision, whereas the Tversky term adjusts the relative costs of false positives and false negatives. Experiments were conducted on the automotive sealing strip surface defect dataset ASSD using four settings: baseline, Dice-only, Tversky-only and Dice-Tversky. Compared with the baseline, Dice-Tversky improved Mask P, Mask R, Mask50 and Mask50-95 by 2.1, 1.3, 2.8 and 10.1 percentage points, respectively. The largest gain appeared in Mask50-95, indicating that the proposed constraint mainly improves high-IoU mask fitting for slender and small-area defects.

Keywords

Automotive Sealing Strip; Small-defect Segmentation; Instance Segmentation; YOLOv8s-seg; Dice-Tversky Loss.

1. Introduction

Automotive sealing strips are key components for body sealing, vibration isolation and noise reduction. Surface cracks, indentations and local anomalies may affect assembly quality and service reliability. In production images, these defects are often small, irregular and affected by texture, reflection and local background variation[1-3]. Bounding-box detection can locate suspicious regions, but it cannot describe defect area, contour or boundary detail[4,5]. Instance segmentation is therefore better suited to sealing-strip inspection when defect size and shape are important for downstream review[6,7].

A major difficulty in this task is the imbalance between a sparse defect foreground and a large background region. Many industrial defect segmentation models rely mainly on pixel-wise classification losses. When the defect occupies only a tiny area, the model may learn a background-biased solution, which appears as missed fine cracks, incomplete indentation boundaries or weak fitting of local anomalies. Dice loss directly constrains the overlap between predicted and ground-truth regions, whereas Tversky loss further separates the penalties for false positives and false negatives[8,9]. This study uses these two losses to redesign the mask supervision objective of YOLOv8s-seg for automotive sealing strip defects.

To keep the manuscript focused, the revised study uses only the ASSD dataset. The research question is therefore direct: whether a Dice-Tversky mask constraint can improve instance segmentation of small and fine-boundary defects in automotive sealing strips under the same training protocol.

2. Dataset and Data Preparation

The ASSD dataset consists of automotive sealing strip surface defect images. It includes three categories, namely crack, indentation and surface_defect. The dataset uses the YOLOv8 segmentation annotation format, in which each defect instance is represented by a class index and normalized polygon vertices. Compared with bounding boxes, polygon annotations preserve defect contours and boundary morphology, making them more suitable for evaluating mask quality in small-defect segmentation.

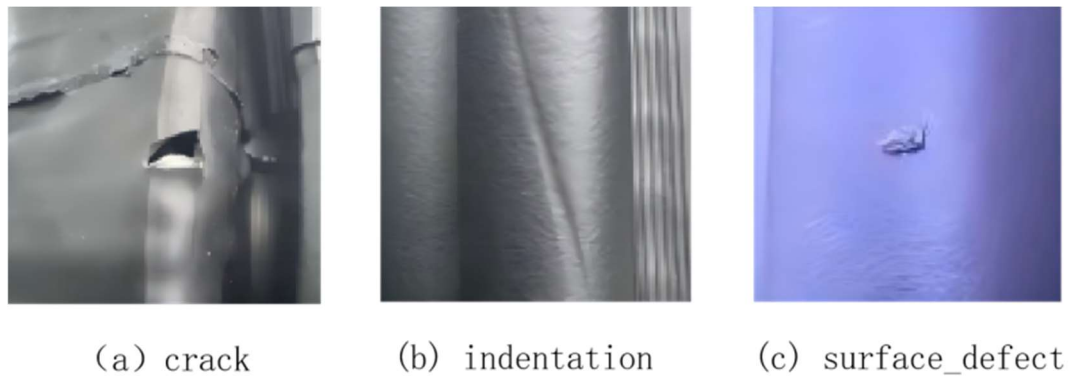


Fig. 1 Examples from the ASSD dataset

During data export, EXIF orientation correction and 640 x 640 size adaptation were performed. The training set contains 204 images and the validation set contains 143 images. In the training set, the numbers of crack, indentation and surface_defect instances are 130, 111 and 98, respectively. In the validation set, the corresponding numbers are 84, 73 and 56. These values show that the task is not mainly limited by severe class-count imbalance. Its main difficulty comes from sparse foreground regions, fragmented boundaries and weak local texture contrast, especially because the median mask area proportion of crack and surface_defect is below 0.1%.

3. Method

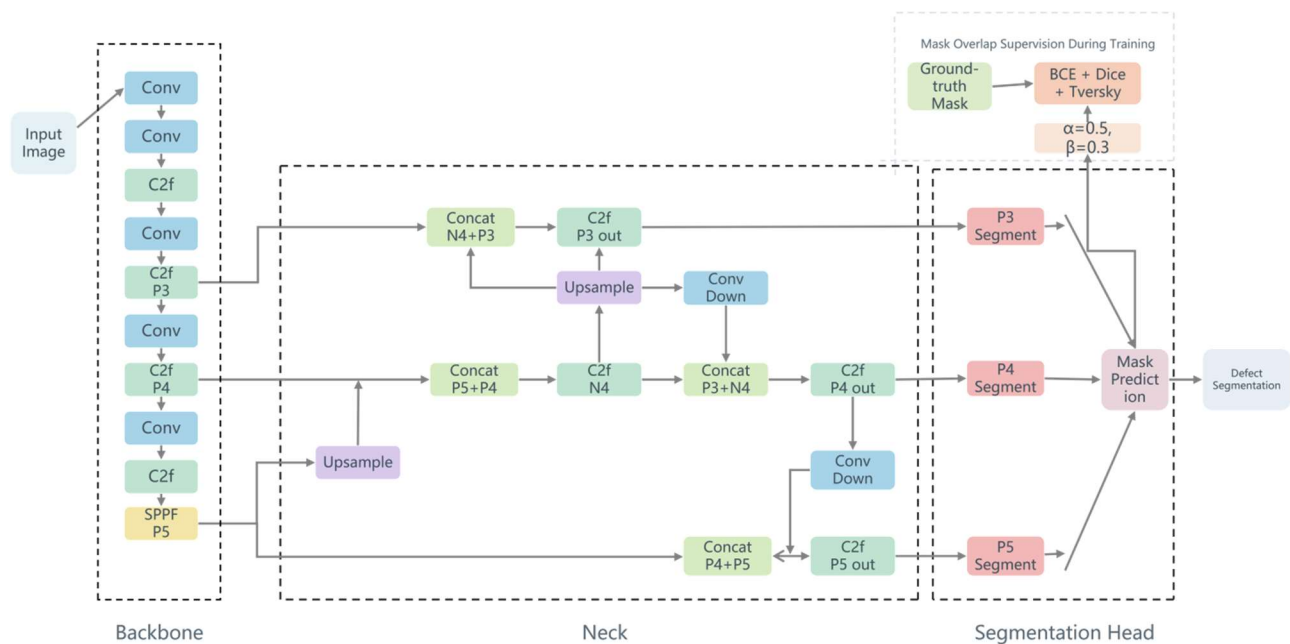


Fig. 2 Training structure of the improved YOLOv8s-seg model

YOLOv8s-seg consists of a Backbone, a Neck and a Segment Head. The Backbone extracts multi-scale features, the Neck fuses features at different resolutions, and the Segment Head predicts bounding boxes, categories and instance mask coefficients. For small sealing-strip defects, shallow detail is important for boundary localization, while the mask loss directly determines how well the predicted mask fits the true defect region.

The original YOLOv8s-seg mask branch mainly uses BCE loss within the cropped target region. BCE provides pixel-level foreground supervision, but it can be weakened when the background occupies most pixels. This study adds Dice and Tversky terms to the original mask loss. The final mask loss is written as $L_{\text{mask}} = L_{\text{BCE}} + \lambda_D L_D + \lambda_T L_T$. Here, L_D improves region-level overlap, and L_T adjusts the weights of FP and FN through alpha and beta, allowing the model to balance false detections and missed detections more explicitly.

Dice-Tversky mask loss. The segmentation branch of YOLOv8s-seg generates each target instance mask by combining prototype masks with mask coefficients. Let a_k denote the predicted mask coefficient vector of the k -th target and P_j the j -th prototype mask. The predicted mask of the target is expressed as:

$$\hat{M}_k = \sigma(\sum_{j=1}^{C_p} a_{(k,j)} P_j) \quad (1)$$

where C_p denotes the number of prototype mask channels, $\sigma(\cdot)$ is the sigmoid function, and $\hat{M}_k$ denotes the predicted probability mask of the k -th target. Industrial defects typically have small foreground areas, fragmented boundaries, and a high background proportion. If the overlap loss is calculated over the entire image, the large number of background pixels can weaken the model's attention to small defects. Therefore, the Dice and Tversky auxiliary losses are calculated within the cropped bounding-box region to reduce interference from irrelevant background areas in the mask-overlap calculation.

Dice loss measures the overall overlap between the predicted mask and the ground-truth mask and is defined as:

$$L_D^k = 1 - (2TP + \epsilon) / (2TP + FP + FN + \epsilon) \quad (2)$$

where TP denotes the intersection between the predicted and ground-truth masks, FP denotes pixels predicted as defects but belonging to the background, FN denotes defect pixels predicted as background, and ϵ is a smoothing term used to prevent division by zero. The Dice term increases the model's attention to the overall coverage of the defect foreground, but penalizes FP and FN relatively symmetrically.

To further adjust the effects of false detections and missed detections, Tversky loss is introduced as:

$$L_T^k = 1 - (TP + \epsilon) / (TP + \alpha FP + \beta FN + \epsilon) \quad (3)$$

where α and β control the weights assigned to FP and FN, respectively. A larger β makes the model more sensitive to missed regions, whereas a larger α more strongly suppresses false-positive regions. In industrial defect segmentation, defect regions are usually small, and missed detections can directly affect subsequent defect recognition and quality assessment. The Tversky term therefore strengthens the model's ability to learn small foreground regions.

On this basis, Dice loss and Tversky loss are used as auxiliary mask-overlap constraints and combined with the original BCE mask loss to form the final mask loss:

$$L_{\text{mask}}^k = L_{\text{BCE}}^k + \lambda_D L_D^k + \lambda_T L_T^k \quad (4)$$

where L_{BCE}^k denotes the original binary cross-entropy mask loss of the k -th target, and λ_D and λ_T denote the weights of Dice loss and Tversky loss, respectively. Four comparative settings, baseline, Dice-only, Tversky-only, and Dice-Tversky, were used to evaluate the effects of different overlap constraints on segmentation performance.

Four experimental settings were designed. The baseline used the original YOLOv8s-seg mask loss. Dice-only added only the Dice term, Tversky-only added only the Tversky term, and Dice-Tversky added both terms. All settings used the same training protocol: YOLOv8s-seg, imgsz=800, batch=8, SGD optimizer, random seed 0 and 200 training epochs. Performance was evaluated by Mask P, Mask R, Mask50 and Mask50-95, which reflect false-detection control, missed-detection control, average precision at IoU 0.5 and stricter mask quality across IoU thresholds[10].

Table 1. Mask loss settings used in the ablation study

Setting	Mask loss	Role in the experiment
baseline	BCE	Original pixel-level mask supervision
Dice-only	BCE + Dice	Improves overall mask overlap
Tversky-only	BCE + Tversky	Adjusts FP and FN penalties
Dice-Tversky	BCE + Dice + Tversky	Combines overlap constraint with missed-region control

4. Results and Analysis

The training curves show that the model still converged normally after the Dice-Tversky constraint was introduced. ASSD contains slender cracks, indentations and local surface anomalies[11-13]. Because these defects are small and have irregular boundaries, Mask50-95 is more informative than Mask50 for judging whether the predicted mask truly fits the defect contour.

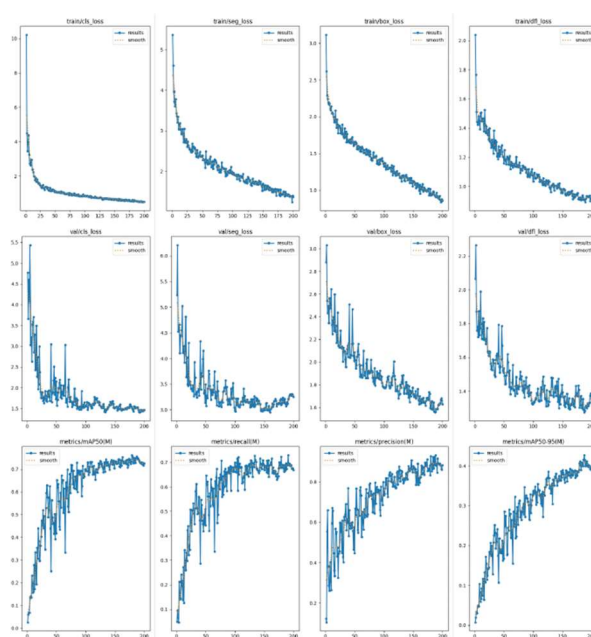


Fig. 3 Training and validation curves of the Dice-Tversky model

The ablation results show that Dice-only and Tversky-only did not improve all metrics consistently. Dice-only increased Mask50 from 0.754 to 0.762, but Mask50-95 decreased from 0.399 to 0.392, indicating that an overall overlap constraint alone did not reliably improve fine-boundary fitting under stricter IoU thresholds. Tversky-only increased Mask50-95 to 0.421, suggesting that asymmetric error weighting helped missed regions and boundary detail, but it did not further improve Mask P. Thus, using either term alone was insufficient to balance overall coverage and error asymmetry.

The joint Dice-Tversky setting produced the most stable improvement. Compared with the baseline, Mask P increased from 0.900 to 0.921, Mask R from 0.690 to 0.703, Mask50 from 0.754 to 0.782, and Mask50-95 from 0.399 to 0.500. The 10.1 percentage-point improvement in Mask50-95 indicates that the proposed loss mainly improves high-IoU mask fitting. In this task, the Dice term improves overall coverage of the defect region, whereas the Tversky term reduces the missed segmentation of small foreground areas. Their combination is therefore more suitable for slender and small-area defects in automotive sealing strips.

The metric behavior is consistent with the design of the loss. Mask50 mainly checks whether the predicted mask overlaps the target at a relatively loose threshold. Mask50-95 averages performance across stricter IoU thresholds, so it is more sensitive to boundary displacement, broken contours and small missing regions. For the ASSD task, a thin crack or a small indentation may occupy only a few pixels after resizing. A slight contour shift can therefore cause a large relative change in IoU. The larger improvement in Mask50-95 indicates that Dice-Tversky did not merely increase coarse localization accuracy, but improved the geometric agreement between predicted masks and annotated defect regions.

Table 2. Ablation results on the ASSD dataset

YOLOv8s-seg	Dice	Tversky	Precision	Recall	mAP(0.5)	mAP(0.5:0.95)
√	—	—	0.9	0.69	0.754	0.399
√	√	—	0.889	0.692	0.762	0.392
√	—	√	0.889	0.698	0.756	0.421
√	√	√	0.921	0.703	0.782	0.5

5. Conclusion

This study improves YOLOv8s-seg for small surface defect segmentation in automotive sealing strips by adding a Dice-Tversky mask overlap constraint to the mask supervision objective. Without changing the main model architecture, the proposed loss combines pixel-level BCE supervision with region-overlap and asymmetric-error constraints. Experiments on ASSD show that Dice-Tversky improves Mask P, Mask R, Mask50 and Mask50-95 compared with the baseline, with the largest gain on Mask50-95. These results indicate that the method is useful for improving high-IoU mask fitting of slender and small-area defects. The conclusion is currently limited to the ASSD dataset and a fixed training protocol. Future work should test the method on additional sealing-strip production scenarios, multiple random seeds and lightweight deployment settings.

References

- [1] Yang, H. Y. (2025). Design of a tire X-ray image defect detection system based on YOLOv8 and PyQt5. *Computer and Network*, 51(3), 265–274.
- [2] Zheng, M. J., Pan, Q., & Xu, H. Y. (2025). Design of chip packaging defect detection system based on deep learning. *Computer Measurement & Control*, 33(8), 45–53.
- [3] Zhao, Y. H., Li, Z., Jin, S., et al. (2026). Rapid detection of photovoltaic defects based on improved YOLOv8. *Acta Energaie Solaris Sinica*, 47(3), 584–593.

- [4] Ren, S. Q., He, K. M., Girshick, R., et al. (2017). Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6), 1137–1149. <https://doi.org/10.1109/TPAMI.2016.2577031>
- [5] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: unified, real-time object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 779–788). IEEE.
- [6] He, K. M., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask R-CNN. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 2961–2969). IEEE.
- [7] Božič, J., Tabernik, D., & Skočaj, D. (2021). Mixed supervision for surface-defect detection: from weakly to fully supervised learning. *Computers in Industry*, 129, 103459. <https://doi.org/10.1016/j.compind.2021.103459>
- [8] Milletari, F., Navab, N., & Ahmadi, S. A. (2016). V-Net: fully convolutional neural networks for volumetric medical image segmentation. In *2016 Fourth International Conference on 3D Vision* (pp. 565–571). IEEE.
- [9] Zhang, H. C., Zheng, S., Zhu, Y. H., et al. (2023). Surface defect detection of wooden boards based on multi-modal fusion. *IEEE*.
- [10] Lin, T. Y., Maire, M., Belongie, S., et al. (2014). Microsoft COCO: common objects in context. In *European Conference on Computer Vision* (pp. 740–755). Springer.
- [11] Huang, J. J., Jiang, X. J., & Liu, Y. (2024). CAGK-YOLO: an efficient detection model for micro-scale and irregular surface defects based on deep learning. *IEEE*.
- [12] Sun, H. Y., An, J. X., Zhang, Z. S., et al. (2025). Based on YOLOv11 steel surface defect detection. *IEEE*.
- [13] Liu, K. H., Chen, S. J., & Liu, T. J. (2020). Unsupervised UNet for fabric defect detection. *IEEE*.