Sensors and Materials, Vol. 38, No. 11 6581–6594
MYU Tokyo

S & M 4696

Real-time Glass Fiber Cross-section Instance Segmentation Using Improved You Only Look Once Segmentation Model Network

Siyuan Zhao,[1] Lei Dong,[1] Wei Ren,[2]
Hsiung-Cheng Lin,[3*] Zhongmin Wang,[1] and Hongchao Zhuang[1]

[1]School of Mechanical Engineering, Tianjin University of Technology and Education, Tianjin 300401, China
[2]Beijing Readir Technology Co., Ltd., Beijing 100031, China
[3]Department of Electronic Engineering, National Chin-Yi University of Technology, Taichung 41170, Taiwan

(Received August 6, 2026; accepted September 14, 2026)

Keywords: CMOS vision sensor. glass fiber cross-section inspection, instance segmentation. YOLO26-seg, industrial defect detection

Glass fiber cross-section inspection plays a critical role in quality control during precision manufacturing and composite material production. However, existing vision-based inspection methods still suffer from limited segmentation accuracy owing to ambiguous boundaries, degraded image contrast, large scale variations, and complex background interference. To address these challenges, in this paper, we propose an improved end-to-end instance segmentation network based on You Only Look Once 26 Segmentation (YOLO26-seg) for accurate glass fiber cross-section inspection. Three complementary modules are introduced to enhance feature representation and boundary localization. First, a dual-stream context attention block (DCAB) is embedded into the backbone to jointly model high-frequency spatial details and low-frequency contextual information, thereby suppressing background interference while preserving discriminative edge features. Second, a bidirectional feature pyramid network concatenation (BiFPN-Concat) module with adaptive dynamic weighting is incorporated to improve multiscale feature aggregation and strengthen the representation of objects with large scale variations. Finally, a high-frequency enhancer (HFE) is used to recover fine boundary structures and alleviate mask bleeding caused by geometric deformation. Experiments on an industrial dataset demonstrate that the proposed method achieves 97.97% mAP@0.5 and 79.88% mAP@0.5:0.95 with a real-time inference speed of 79.5 frames per second. Compared with the YOLO26-seg baseline, the proposed model improves mAP@0.5:0.95 by 12.28 percentage points and also outperforms representative models such as YOLO11-seg, Mask R-CNN, and Mask2Former. These results demonstrate strong potential for real-time sensor-based industrial quality inspection.

*Corresponding author: e-mail: hclin@ncut.edu.tw
https://doi.org/10.18494/SAM6565

ISSN 0914-4935 © MYU K.K.
https://myukk.org/

1. Introduction

With the rapid development of precision manufacturing and advanced composite materials, increasingly stringent quality requirements have been imposed on microscopic defect inspection.[1,2] Glass fiber cross sections are essential indicators for evaluating manufacturing quality, as defects such as cracks, incomplete sections, and irregular boundaries may significantly affect the mechanical performance and service life of composite materials.[3,4] Consequently, accurate and real-time inspection of glass fiber cross sections has become an important task in modern intelligent manufacturing systems.

Recently, CMOS vision sensors have been widely adopted in industrial inspection because of their high spatial resolution, rapid image acquisition, low power consumption, and noncontact measurement capability.[5,6] Combined with machine vision algorithms, CMOS-based inspection systems enable efficient and automated quality evaluation without interrupting the production process.[7,8] However, images captured in practical manufacturing environments are often affected by nonuniform illumination, strong surface reflections, low foreground–background contrast, and complex texture interference.[9] These factors considerably increase the difficulty of accurately segmenting glass fiber cross sections, particularly when the targets exhibit irregular shapes or large scale variations.[10] Therefore, the robust segmentation of images acquired by vision sensors is essential for reliable automated quality inspection.

Traditional machine vision approaches mainly rely on handcrafted features, geometric constraints, or threshold-based segmentation methods.[11] Although these techniques perform satisfactorily under controlled laboratory conditions, their robustness is limited when applied to practical industrial scenarios. Variations in illumination, background texture, and object morphology frequently result in inaccurate contour extraction, missed detections, and false positives. Moreover, many conventional algorithms require predefined geometric assumptions or manually designed parameters, making them difficult to be generalized to complex production environments.

Recent advances in deep learning have significantly improved instance segmentation performance.[12] Representative methods include CNN-based frameworks such as Mask R-CNN,[13] Transformer-based models such as Mask2Former,[14] and lightweight one-stage networks represented by the You Only Look Once (YOLO) family.[15,16] Mask R-CNN provides accurate object localization but suffers from the loss of fine structural information caused by repeated region sampling. Transformer-based methods achieve stronger global contextual modeling but require considerable computational resources, limiting their deployment on real-time industrial platforms. Although YOLO-based instance segmentation networks provide an attractive balance between accuracy and efficiency, they still face challenges in preserving high-frequency boundary information, suppressing background interference, and effectively aggregating multiscale features under complex industrial conditions.[17–19]

2. Proposed Network Architecture

2.1 Overall network architecture

The overall architecture of the proposed network is illustrated in Fig. 1. It is developed on the basis of the YOLO26-seg instance segmentation architecture and consists of three major components, namely, the backbone, the neck, and the segmentation head. A dual-stream context attention block (DCAB) is embedded into the C3k2 bottleneck blocks. The neck adopts an adaptive multiscale feature fusion strategy based on BiFPN-Concat. Before the segmentation head, a lightweight high-frequency enhancer is introduced to refine boundary representations.

2.2 Dual-stream context attention block

The accurate extraction of weak edge features remains a major challenge in industrial inspection because glass fiber images are often affected by specular reflections, uneven illumination, and complex background textures. Although conventional attention mechanisms improve contextual modeling, they usually introduce considerable computational overhead, making them unsuitable for lightweight real-time inspection networks. To address these limitations, a lightweight DCAB is proposed.[20,21] As illustrated in Fig. 2, the module employs two parallel feature extraction branches to capture complementary information from different perspectives.

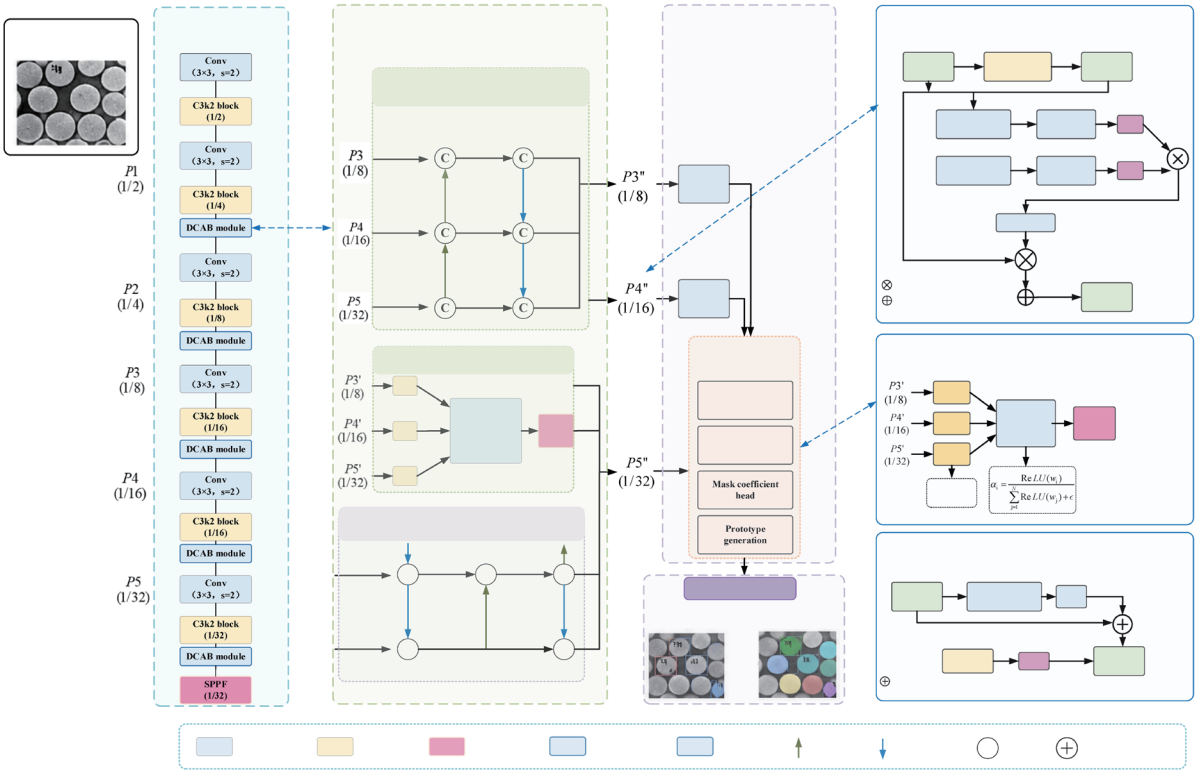

Fig. 1. (Color online) Overall architecture of the proposed improved YOLO26-seg network for instance segmentation.

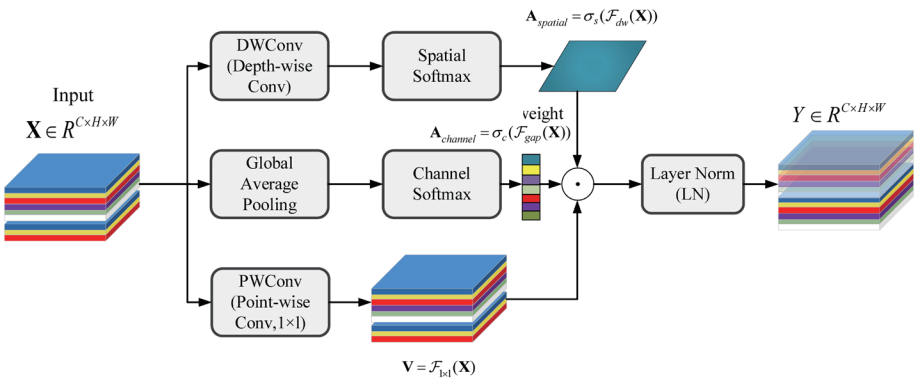

Fig. 2. (Color online) Detailed architecture of the proposed DCAB.

The first branch focuses on local spatial information. A depth-wise convolution is employed to extract high-frequency responses associated with object boundaries and fine textures, followed by spatial Softmax normalization to generate a spatial attention map. This branch emphasizes regions with significant local structural variations while reducing redundant background responses. The purpose of the second branch is to capture global contextual information through global average pooling (GAP). The pooled feature descriptor is normalized using channel-wise Softmax to produce channel attention weights, allowing the network to adaptively highlight feature channels that contribute most significantly to target representation.[22]

Given an input feature map $X \in \mathbb{R}^{C\times H\times W}$, where C denotes the number of channels and H and W represent the spatial height and width, respectively, the proposed DCAB performs feature refinement through three parallel computational branches. To capture weak geometric boundary information of glass fiber cross sections, the local spatial branch employs a 3 × 3 depth-wise separable convolution (DWConv) to extract high-frequency spatial responses. Subsequently, a spatial Softmax normalization operation is applied to generate the spatial attention weight map $A_{spatial}$:

$$A_{spatial} = \sigma_s\left(\mathcal{F}_{dw}(X)\right), \tag{1}$$

where $\mathcal{F}_{dw}(\cdot)$ represents a depthwise convolution operation with the number of groups equal to C, and σ_s denotes the Softmax normalization calculated along the spatial dimension ($H \times W$). This operation enables the network to selectively enhance regions with significant local gradient variations and preserve fine boundary structures.

To suppress false responses caused by illumination variations and complex background textures, the global context branch employs GAP to compress spatial information and obtain a global contextual representation. The channel attention vector $A_{channel}$ is calculated as follows:

$$A_{channel} = \sigma_c\left(\mathcal{F}_{gap}(X)\right), \tag{2}$$

where $\mathcal{F}_{gap}(\cdot)$ maps the input feature map into a global context descriptor with the dimension of $\mathbb{R}^{C\times1\times1}$, and σ_c represents the channel-wise Softmax normalization. The generated channel weights enable the network to adaptively emphasize feature channels that contribute more significantly to defect representation.

To preserve the original semantic information while enhancing discriminative features, a 1 × 1 point-wise convolution is first applied to linearly transform the input feature map $\boldsymbol{X}$, producing the feature representation $\boldsymbol{V}$.

$$\boldsymbol{V} = \mathcal{F}_{1\times1}(\boldsymbol{X}) \tag{3}$$

Then, the spatial attention map $A_{spatial}$, channel attention vector $A_{channel}$, and feature representation V are fused through element-wise multiplication. The final output feature map Y is reconstructed using layer normalization (LN) and a residual connection.

$$Y = X + LN(V \odot A_{spatial} \odot A_{channel}) \tag{4}$$

The spatial attention map and the channel attention vector are broadcast along their corresponding dimensions before element-wise multiplication to ensure dimensional consistency during feature fusion. In Eq. (4), $\odot$ denotes element-wise multiplication between feature maps and attention weights. The proposed cross-dimensional decoupling mechanism enables the network to simultaneously focus on fine-grained high-frequency structures and global low-frequency contextual information with negligible additional computational cost. By integrating spatial and channel-wise attention responses, DCAB effectively reduces the influence of specular reflections and complex backgrounds, thereby improving the representation capability of weak defect features.

2.3 Bidirectional feature pyramid network concatenation

Industrial defects often produce significant scale variations and irregular geometric distributions, making multiscale feature representation essential for accurate instance segmentation. Conventional feature pyramid networks (FPNs) usually employ direct concatenation or equal-weight feature aggregation, which treats features from different pyramid levels equally.[23] To overcome these limitations, an adaptive BiFPN-Concat module is introduced into the neck network, as illustrated in Fig. 3. On the basis of the adaptive feature pyramid network (AFPN), the proposed module performs dynamic feature fusion by assigning learnable importance weights to feature maps from different pyramid levels. This strategy enables the network to adaptively adjust the contribution of each feature in accordance with the input image while introducing only negligible computational overhead.[24]

Given a set of multi-scale feature maps extracted from different receptive field levels of the backbone, the proposed BiFPN-Concat module first resizes all feature maps to the same spatial resolution through upsampling or downsampling operations before feature fusion. Given the

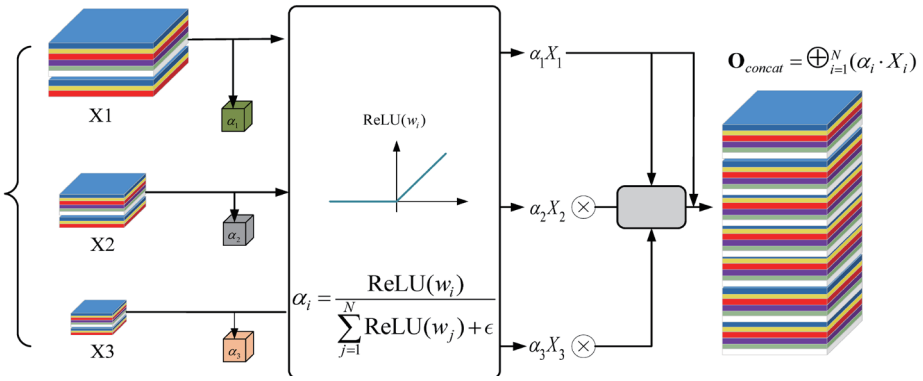

Fig. 3. (Color online) Structure of the BiFPN-Concat module with adaptive dynamic weighting strategy.

aligned multiscale feature set $\mathcal{X} = \{X_1, X_2, \ldots, X_N\}$, where N denotes the number of feature branches participating in the current fusion stage, a learnable scalar weight is assigned to each feature level $\boldsymbol{w} = \{w_1, w_2, \ldots, w_N\}$.

To guarantee nonnegative feature weights while reducing the computational overhead introduced by the Softmax operation, the proposed BiFPN-Concat adopts the fast normalized fusion strategy.[25] The normalized adaptive fusion coefficient α_i is calculated as

$$\alpha_i = \frac{\mathrm{ReLU}(w_i)}{\sum_{j=1}^{N} \mathrm{ReLU}(w_j) + \varepsilon}, \tag{5}$$

where ReLU(·) ensures that all fusion weights remain nonnegative, and $\varepsilon = 10^{-4}$ is a small constant introduced to maintain numerical stability during optimization.

After obtaining the adaptive fusion coefficients, each feature map is weighted through scalar multiplication and progressively aggregated using channel-wise concatenation. The fused output feature map is expressed as

$$O_{concat} = \oplus_{i=1}^{N} \left(\alpha_i \cdot X_i \right), \tag{6}$$

where $\oplus$ denotes the channel-wise concatenation operation.

BiFPN-Concat can dynamically determine the contribution of each pyramid level during training. More informative feature maps are assigned larger weights, whereas less informative features are automatically suppressed. Consequently, shallow geometric information and deep semantic representations are effectively integrated, improving the representation of objects with substantial scale variations and irregular shapes. This adaptive fusion strategy enhances segmentation performance while maintaining the lightweight characteristics required for real-time industrial inspection.

2.4 High-frequency enhancer

In conventional instance segmentation networks, the segmentation head primarily relies on deep semantic features to generate instance masks.[26] Accurate boundary localization remains challenging in instance segmentation because repeated downsampling gradually weakens high-frequency spatial information, resulting in blurred contours and incomplete mask prediction for irregular objects. This problem is particularly evident in glass fiber cross sections, which often exhibit nonconvex shapes, and complex boundary variations.[27,28]

To address this issue, a lightweight high-frequency enhancer is introduced before the segmentation head, as illustrated in Fig. 4. Rather than relying on predefined geometric constraints, the proposed module directly enhances local gradient responses in the feature space, thereby improving boundary representation before mask prediction.

Given the fused feature map generated by the neck network, $F_{in} \in \mathbb{R}^{C\times H\times W}$, the proposed HFE module first applies a 3 × 3 grouped convolution as a learnable spatial high-pass filter to extract local gradient variations and boundary information. The resulting feature map is then normalized by batch normalization (BN) to obtain the high-frequency edge representation:

$$F_{edge} = \mathcal{B}\left(\mathcal{F}_{gconv}(F_{in})\right), \tag{7}$$

where $\mathcal{F}_{gconv}(\cdot)$ denotes a grouped convolution with the number of groups equal to the channel number C. This operation suppresses interchannel interference while preserving channel-specific geometric gradient information. $B(x)$ represents the batch normalization operation. The extracted high-frequency feature map is subsequently fused with the original semantic feature through an element-wise residual addition:

$$\boldsymbol{F}_{aug} = \boldsymbol{F}_{in} \oplus \boldsymbol{F}_{edge}, \tag{8}$$

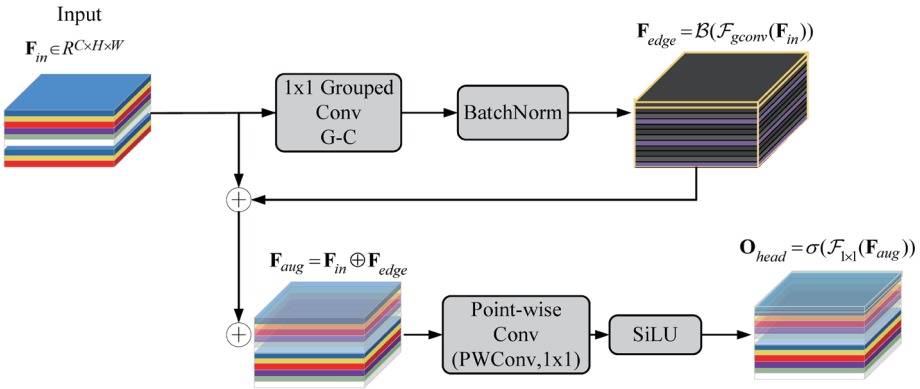

Fig. 4. (Color online) Schematic diagram of the proposed high-frequency enhancer module.

where the residual connection effectively preserves the original semantic representation while enhancing fine boundary details.

Finally, a 1 × 1 point-wise convolution is employed to reorganize channel-wise features, followed by the SiLU activation function to generate the refined feature representation for the segmentation head:

$$\boldsymbol{O}_{head} = \sigma\left(\mathcal{F}_{1\times1}(\boldsymbol{F}_{aug})\right), \tag{9}$$

where $\sigma(x)$ denotes the SiLU activation function.

By enhancing high-frequency boundary information before mask prediction, the proposed HFE module effectively improves contour localization and alleviates mask bleeding around irregular object boundaries. The lightweight design introduces only minimal computational overhead while producing more accurate and continuous instance masks, making it well suited to real-time industrial inspection.

3. Performance Results and Analysis

3.1 Dataset preparation

To support real-world glass fiber cross-section inspection, a dedicated dataset was established using images collected from an actual production line. The acquired samples mainly contain normal cross sections together with typical defects, including incomplete sections, cracks, and edge damage, as shown in Fig. 5. Image acquisition was performed using an industrial camera integrated with a metallurgical microscope under different operating conditions, including uneven illumination, specular reflections, and complex background environments. A total of 1158 original images were collected. All images were resized to 640 × 640 pixels before training. After augmentation, the dataset contained 3545 annotated images. All samples were manually annotated using Labelme, where each glass fiber cross section was labeled with a polygonal mask to accurately describe its contour. The complete dataset was randomly divided into training, validation, and test sets at a ratio of 7:2:1.

3.2 Setup of performance environment

Performance evaluation was performed on a workstation equipped with an Intel Xeon Platinum 8350C processor, 32 GB of system memory, and an NVIDIA RTX 3090 Ti graphics card featuring 24 GB of video memory. The implementation of all compared methods was based on the PyTorch 2.0.0 framework running on the Windows 11 operating system. The corresponding hardware specifications are listed in Table 1.

Identical hyperparameter settings were adopted for all comparative models. The input image size was 640 × 640, the optimizer was SGD, the initial learning rate was 0.01, the momentum was 0.937, the weight decay was 0.0005, and all models were trained for 200 epochs. The complete training configuration is listed in Table 2.

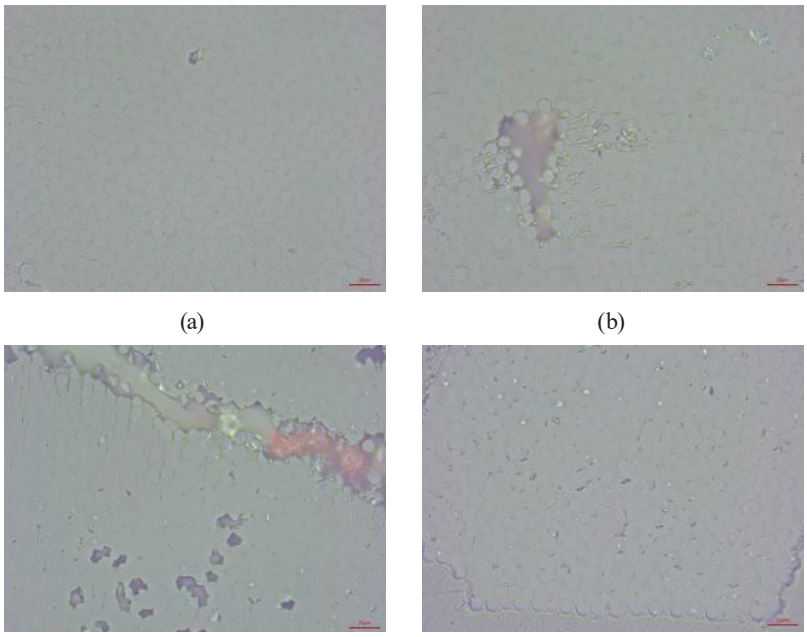

Fig. 5. (Color online) Typical samples and defect categories in the glass fiber cross-section dataset. (a) Normal, (b) hole, (c) crack, and (d) edge.

Table 1
Configuration of the experimental platform.

Component name	Component details
CPU	Intel Xeon Platinum 8350C @ 2.60 GHz
RAM	32 GB
GPU	RTX 3090 Ti (24 GB)
Operating system	Windows 11
PyTorch	2.0.0
CUDA version	11.8

Table 2
Hyperparameter settings of the YOLO Model.

Training parameters	Parameter value
Input image resolution	640 × 640
Number of training epochs	200
Learning rate	0.01
Weight decay	0.0005
Momentum	0.937
Optimizer	SGD

3.3 Performance results and discussion

The proposed method was evaluated using precision, recall, mAP@0.5, mAP@0.5:0.95, model parameters (Params), and inference speed (FPS). In this study, comprehensive comparisons were conducted with representative instance segmentation models, including

YOLO series (YOLOv5-seg, YOLOv8-seg, YOLOv9-seg, YOLO11-seg, and YOLO26-seg), the two-stage CNN-based model Mask R-CNN, and the Transformer-based model Mask2Former.[29] The quantitative comparison results are summarized in Table 3. As can be seen, the proposed method achieves the best overall segmentation performance among all models. For example, it obtains 97.97% mAP@0.5 and 79.88% mAP@0.5:0.95, representing the highest segmentation accuracy while maintaining real-time inference capability. Compared with the baseline YOLO26-seg, the proposed method improves mAP@0.5:0.95 from 67.60 to 79.88%, corresponding to an absolute improvement of 12.28%.

A series of ablation tests was conducted using YOLO26-seg as the baseline. DCAB, BiFPN-Concat, and HFE were incrementally incorporated into the baseline network, and the quantitative results are summarized in Table 4.

First, incorporating DCAB alone (Variant A) produces only a limited improvement in overall segmentation accuracy but noticeably increases precision. This suggests that DCAB primarily suppresses background interference and reduces false-positive predictions by jointly modeling local spatial information and global contextual features. When only the BiFPN-Concat module is introduced (Variant B), mAP@0.5:0.95 increases from 67.60 to 70.65%, demonstrating that adaptive feature weighting effectively enhances multiscale feature representation and preserves fine structural information. When only the HFE module is used (Variant C), mAP@0.5:0.95 further increases to 71.96%, corresponding to an improvement of 4.36% over the baseline. This indicates that reinforcing high-frequency boundary information effectively compensates for the loss of spatial details.

Table 3
Quantitative comparison with representative instance segmentation models.

Model configuration	Precision	Recall	mAP@0.5:0.95	mAP@0.5	Params (M)	FPS
YOLOv5	0.92206	0.93628	0.70224	0.95796	3.311	89.5
YOLOv8-seg	0.94877	0.87318	0.67407	0.93545	3.199	104.6
YOLOv9-seg	0.95375	0.92658	0.67294	0.93089	3.332	76.3
YOLO11-seg	0.94733	0.92508	0.69488	0.96781	3.173	74.6
YOLO26-seg	0.93129	0.89803	0.67596	0.94892	3.311	89.5
Mask R-CNN	0.96176	0.97762	0.69830	0.95833	44.38	23.7
Mask2Former	0.96158	0.96347	0.74097	0.94042	42.11	15.3
Proposed model	0.95997	0.94286	0.79880	0.97965	7.874	79.5

Table 4
Ablation study table.

Model configuration	DCAB	BiFPN-Concat	HFE	Precision	Recall	mAP@0.5:0.95	mAP@0.5	Params (M)	FPS
Baseline	—	—		0.93129	0.89803	0.67596	0.94892	3.311	89.5
Variant A	√	—		0.95204	0.89617	0.67407	0.94856	3.199	104.6
Variant B	—	√		0.95824	0.92737	0.70648	0.96974	3.332	76.3
Variant C			√	0.95757	0.93286	0.71961	0.96584	9.173	74.6
Variant D	√	√		0.91758	0.88339	0.66021	0.93341	3.22	99.1
Variant E		√	√	0.96096	0.93031	0.70916	0.96349	9.193	80.5
Variant F	√		√	0.94975	0.91642	0.70013	0.95887	9.06	93.3
Proposed model	√	√	√	0.95997	0.94286	0.79880	0.97965	7.874	79.5

Interestingly, several dual-module combinations (Variants D–F) do not consistently outperform the corresponding single-module models. This implies that directly combining two enhancement strategies does not necessarily achieve optimal feature collaboration. Instead, each module is designed to address a different stage of feature processing, including feature extraction, multiscale fusion, and boundary refinement. Consequently, their individual advantages can only be fully exploited when all three modules are integrated into a unified framework.

When DCAB, BiFPN-Concat, and HFE are incorporated simultaneously, the proposed network achieves the best overall performance, obtaining 97.97% mAP@0.5 and 79.88% mAP@0.5:0.95 while maintaining a real-time inference speed of 79.5 FPS. Compared with the baseline YOLO26-seg, the complete model improves mAP@0.5:0.95 by 12.28%, confirming that the three modules complement one another and jointly enhance feature representation, multiscale information fusion, and boundary localization.

To further evaluate the detection capability of the proposed method under different confidence thresholds, the precision–recall (PR) curves and F1-score curves of representative instance segmentation models are presented in Figs. 6 and 7, respectively.

As shown in Fig. 6, the proposed method consistently maintains a PR curve closer to the upper-right corner than the compared models, indicating superior PR characteristics over a wide range of confidence thresholds. The larger area under the PR curve is consistent with the highest mAP values reported in Table 3 and indicates more reliable instance segmentation under complex industrial conditions. The corresponding F1-score curves are shown in Fig. 7. The proposed method attains the highest peak F1-score among all compared models and maintains relatively stable performance over a broad confidence interval.

Qualitative visualization on densely distributed glass fiber cross sections was also conducted to compare the proposed method with existing approaches,[30] as shown in Fig. 8. In Figs. 8(a)–8(c), YOLO26-seg, Mask R-CNN, and Mask2Former, respectively, produce incomplete masks or

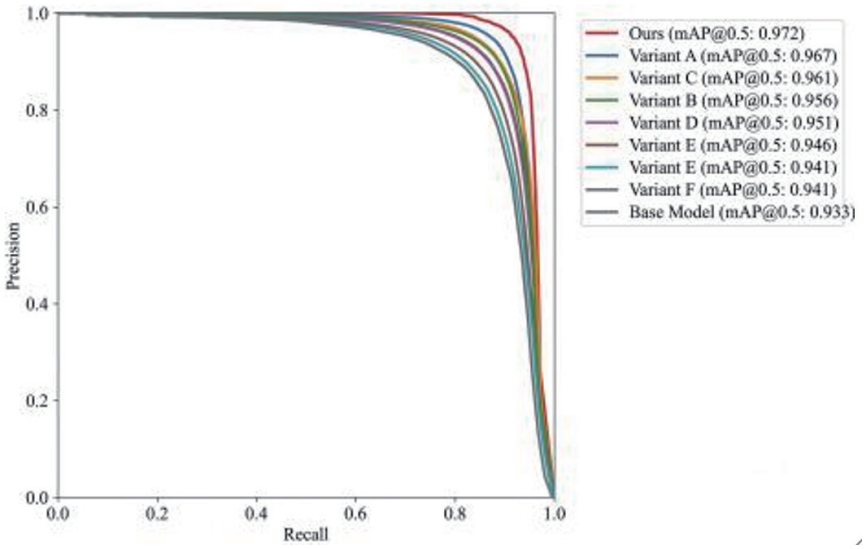

Fig. 6. (Color online) PR curves for the baseline model, various ablation variants, and our proposed method.

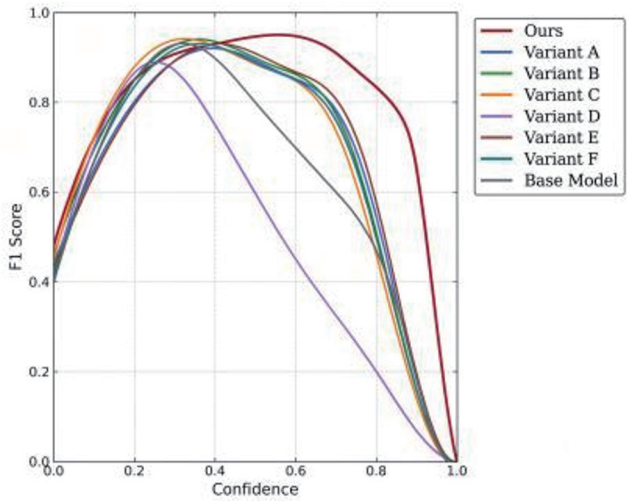

Fig. 7. (Color online) F1-Confidence curves of ablation variants and the proposed model under different confidence thresholds.

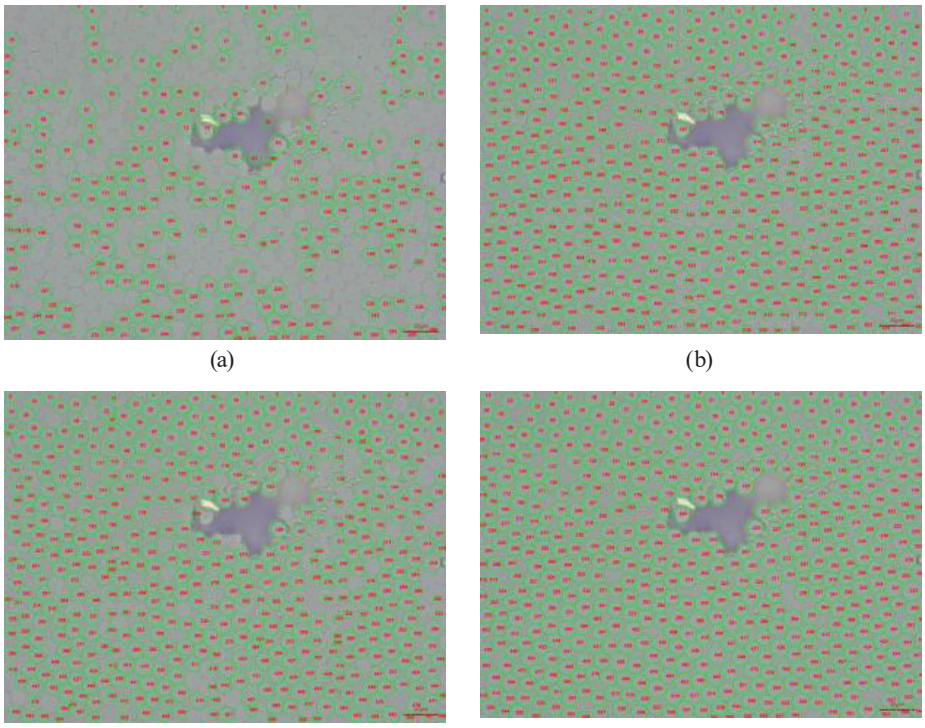

Fig. 8. (Color online) Qualitative results of glass fiber cross-section instance segmentation. (a) YOLO26-seg, (b) Mask R-CNN, (c) Mask2Former, and (d) Proposed model.

noticeable boundary leakage, miss adjacent instances, or merge neighboring objects into a single prediction. The most severe errors are seen in Fig. 8(a). In contrast, the proposed method separates individual fibers while preserving complete object contours, resulting in more accurate contour localization and smoother instance boundaries.

4. Conclusions

In practice, glass fiber cross-section inspection is challenged by large scale variations, blurred boundaries, and severe background interference. In this study, the three proposed modules, namely, DCAB, BiFPN-Concat, and HFE, complement one another throughout the feature extraction, multiscale feature fusion, and boundary refinement stages, enabling the more accurate segmentation of glass fiber cross sections with irregular shapes and complex backgrounds. The proposed model achieves 97.97% mAP@0.5 and 79.88% mAP@0.5:0.95. Compared with the YOLO26-seg baseline, it improves mAP@0.5:0.95 by 12.28 percentage points and also outperforms representative models such as Mask R-CNN and Mask2Former. The model maintains a lightweight architecture with only 7.87 M parameters and a real-time inference speed of 79.5 FPS. The results of PR and F1 analyses further verify that the proposed method makes instance masks more complete, continuous, and consistent. Also, the visualization results agree well with the quantitative evaluation. These findings demonstrate the suitability of the proposed method for processing vision-sensor images in real-time industrial quality inspection.

Considering practical situations, segmentation errors may still occur in challenging scenarios involving severe object adhesion, blurred boundaries, or strong imaging noise under the complexity of industrial environments. Future work will be focused on evaluating the method under more diverse industrial conditions, optimizing deployment on edge-computing platforms, and integrating three-dimensional measurement and multimodal sensing techniques to further improve segmentation robustness, geometric characterization, and intelligent quality inspection capability.

Acknowledgments

The Tianjin Education Commission Research Program Project [Project No. 2022ZD027] provided funding for this study.

References

1 N. K. T. Quah, X. Z. Liu, D. W. Y. Tay, M. Y. Li, M. J. Tan, T. N. Wong, and K. H. H. Li: Virtual Phys. Prototyping **20** (2025) 1120. https://doi.org/10.1080/17452759.2025.2544758
2 D. Zhang, Z. Meng and Z. Li: Compos. Struct. **387** (2026) 120341. https://doi.org/10.1016/J.COMPSTRUCT.2026.120341
3 K. Panasiuk, K. Dudzik and J. Kłosiński: Polish Maritime Res. **33** (2026) 128. https://doi.org/10.2478/POMR-2026-0012
4 J. Drummer, T. Luplow, L. Littner, R. Prots, S. Heimbs, M. Kreutzbruck, M. Gude, and B. Fiedler: Compos. Part C: Open Access **18** (2025) 100633. https://doi.org/10.1016/J.JCOMC.2025.100633
5 L. Wu, H. Y. Hao and Y. Song: Acta Autom. Sin. **50** (2024) 1261. https://doi.org/10.16383/j.aas.c230039
6 Z. Chen, J. Xu, D. Guan, X. Zheng, and Y. Li: Sensors **26** (2026) 3686. https://doi.org/10.3390/S26123686
7 J. Nakamura: Image Sensors and Signal Processing for Digital Still Cameras (CRC Press, Boca Raton, FL, 2006). https://doi.org/10.1201/9781420026856
8 J. Kim, Y. Lee and S. Park: Sens. Mater. **32** (2020) 4217. https://doi.org/10.18494/SAM.2020.3101
9 M. Gao, X. Gao, X. Feng, and D. Ye: Signal, Image and Video Process. **20** (2026) 202. https://doi.org/10.1007/S11760-026-05189-6

10 A. Yıldırım and R. Terzi: Visual Comput. **42** (2026) 145. https://doi.org/10.1007/S00371-025-04351-4
11 Y. Pei, H. Cao, Q. Liu, L. Chu, X. Chen, and X. Xu: J. Phys.: Conf. Series **3175** (2026) 012011. https://doi.org/10.1088/1742-6596/3175/1/012011
12 Y. Wang, Y. Zhang, Y. Yang, Y. Pu, A. Huang, and D. Xi: Smart Agric. Technol. **14** (2026) 102092. https://doi.org/10.1016/J.ATECH.2026.102092
13 K. He, G. Gkioxari, P. Dollár and R. Girshick: IEEE Trans. Pattern Anal. Mach. Intell. **42** (2020) 386. https://doi.org/10.1109/TPAMI.2018.2844175
14 B. Cheng, I. Misra, A. G. Schwing, A. Gupta and C. Charalambous: IEEE Trans. Pattern Anal. Mach. Intell. **46** (2024) 2154. https://doi.org/10.1109/TPAMI.2023.3317834
15 J. Redmon, S. Divvala, R. Girshick and A. Farhadi: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (2016) 779. https://doi.org/10.1109/CVPR.2016.91
16 C. Y. Wang, A. Bochkovskiy and H. Y. M. Liao: Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (2023) 7461. https://doi.org/10.1109/CVPR52729.2023.00729
17 G. Jocher, J. Qiu, M. Liu, S. Lyu, F. C. Akyon, and M. E. Kalfaoglu: arXiv:2606.03748 (2026). https://doi.org/10.48550/arXiv.2606.03748
18 M. Wang, B. Zhang, S. Zhan, L. Yang, L. Jiang, and Y. Wang: Appl. Sci. **15** (2025) 12192. https://doi.org/10.3390/APP152212192
19 X. Ling, Z. Tong, O. Li, S. Wang, P. Xia, C. Qin, Y. Huang, and C. Liu: Meas. Sci. Technol. **37** (2026) 276112. https://doi.org/10.1088/1361-6501/AE7C8E
20 L. Zhang, Q. Liu, X. Cao, Y. Zhang, J. Sun, M. Cui, X. Ren, and D. Wu: J. Electr. Comput. Eng. **2026** (2026) 2638110. https://doi.org/10.1155/JECE/2638110
21 L. Kang, H. Zhang, and Y. Lu: Computing **108** (2026) 118. https://doi.org/10.1007/S00607-026-01697-Y
22 C. Liu, J. Lou, G. Wen, G. Li, B. Sun, and X. Liang: Appl. Soft Comput. **202** (2026) 115819. https://doi.org/10.1016/J.ASOC.2026.115819
23 J. M. Hasan, W. M. H. S. W. Ahmad, A. F. M. Fauzi, F. S. Abas, J. T. H. Lee, S. Y. Khor, L. M. Looi, A. S. Khalid, T. B. Tang, and N. N. Razak: Biomed. Signal Process. Control **126** (2026) 110873. https://doi.org/10.1016/J.BSPC.2026.110873
24 J. Liu, M. Ding, X. Liu, S. Jiang, K. Zhang, and Z. Xu: The Int. J. Advanced Manuf. Technol. **2026** (2026) 1. https://doi.org/10.1007/S00170-026-18536-1
25 J. Liu, Z. Xia, Y. Chen, H. Sun, and S. Huang: Sci. Rep. **16** (2026) 54780. https://doi.org/10.1038/S41598-026-54780-X
26 C. Che, Z. Li, H. Wang, and H. Li: Proc. Inst. Mech. Eng. Part C **240** (2026) 5492. https://doi.org/10.1177/09544062261434389
27 P. Wang, X. Fan, Q. Yang, S. Tian, and L. Yu: Multimedia Syst. **31** (2025) 93. https://doi.org/10.1007/S00530-025-01687-8
28 Y. Liu, F. Zeng, H. Diao, J. Zhu, D. Ji, X. Liao, and Z. Zhao: Sensors **24** (2024) 4379. https://doi.org/10.3390/S24134379
29 S. Zhang, M. He, Z. Zhong, and D. Zhu: Measurement **221** (2023) 113433. https://doi.org/10.1016/J.MEASUREMENT.2023.113433
30 P. Zhang, X. Zhou, R. Liang, J. Li, K. Tang, and Y. Li: Composites Sci. Technol. **265** (2025) 111160. https://doi.org/10.1016/j.compscitech.2025.111160