Check for updates

High Power Laser Science and Engineering, (2026), Vol. 14, e79, 15 pages.
doi:10.1017/hpl.2026.10167

HIGH POWER LASER SCIENCE AND ENGINEERING

RESEARCH ARTICLE

A spot analysis system for high-power laser facilities based on an improved YOLOv8-seg model

Zhenghao Song[1,2], Peiying Zeng[1], Panzheng Zhang[1], and Jianqiang Zhu[1]

[1]*National Laboratory on High Power Laser and Physics, Shanghai Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, Shanghai, China*

[2]*University of Chinese Academy of Sciences, Beijing, China*

(Received 26 December 2025 UTC; revised 24 April 2026 UTC; accepted 26 May 2026 UTC)

Abstract

A spot analysis system based on an improved YOLOv8-seg model was developed for high-power laser facilities to meet the urgent need for fast and precise positional control of optical components throughout the entire laser chain. The system addresses the inefficiency of traditional manual or semi-automatic methods when aligning numerous beamlines. An enhanced segmentation model (mean average precision@0.5 = 0.901) was constructed to recognize and extract various spot types. A multi-reference center detection mechanism was designed, achieving a positioning error of 1.9 pixels and a success rate of 98.7%, thereby improving centering accuracy and robustness in complex imaging scenarios. The system automates spot center location, morphological analysis and offset calculation with efficient processing speed to satisfy the real-time alignment demand of the facility. It outputs pixel/μm-level offset values, adjustment step size, step count and direction as structured decision parameters, providing actionable inputs for the closed-loop control of motorized actuators in rapid laser alignment.

Keywords: adaptive multi-scale fusion; high-power laser facility; intelligent detection of laser spots

1. Introduction

Inertial confinement fusion (ICF) research stands as one of the foremost scientific challenges in the world today, and its successful implementation holds profound significance for national energy security and fundamental scientific exploration[1]. Large-scale laser facilities serve as the critical platforms for such research, with representative examples being the National Ignition Facility (NIF) in the United States[2–5] and the Laser Megajoule (LMJ) facility in France[6,7], where the NIF has already achieved an energy gain greater than one. For these massive laser facilities, the precise alignment of thousands of optical components in the beamlines to micrometer or even sub-micrometer accuracy is paramount to the final output performance[8].

The core of achieving high-precision alignment lies in the real-time and accurate quantitative analysis of the laser spot images propagating through the entire amplification chain. The morphology and center position of these spots are the most direct information carriers reflecting beam quality and spatial alignment status[9]. From a technical perspective, traditional automatic spot image analysis methods face significant challenges. Despite their widespread use, classical automatic spot analysis methods face fundamental limitations that severely constrain their effectiveness in modern high-power laser facilities. Threshold segmentation methods[10–13] assume bimodal grayscale distributions and perform poorly when spots exhibit low contrast, nonuniform illumination or overlapping boundaries – conditions prevalent in high-power laser beamlines due to thermal effects and optical damage. Edge detection algorithms[14–16] rely on gradient-based feature extraction and are highly sensitive to noise and stray light, often producing fragmented or spurious contours that compromise subsequent morphological analysis. Centroid algorithms[17,18], while computationally efficient, are inherently global statistical measures that cannot distinguish between multiple overlapping spots or adapt to irregular morphologies, and their accuracy degrades significantly under partial occlusion or asymmetric intensity distributions.

Furthermore, these classical methods typically serve as isolated preprocessing modules within conventional proportional–integral–derivative (PID)-based closed-loop control systems, where the segmentation, feature extraction

Correspondence to: P. Zeng, National Laboratory on High Power Laser and Physics, Shanghai Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, Shanghai 201800, China. Email: zengpeiying@siom.ac.cn

© Shanghai Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, 2026. Published by Cambridge University Press in association with Chinese Laser Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/hpl.2026.10167 Published online by Cambridge University Press

and control decision-making stages are decoupled. This modular separation prevents end-to-end optimization: errors propagate cumulatively from segmentation to positioning, the system cannot autonomously adapt to new spot morphologies without manual parameter retuning and the lack of instance-level discrimination makes parallel multi-beamline alignment infeasible. Consequently, classical approaches require extensive human intervention for parameter adjustment and validation, rendering them inadequate for the unattended, high-throughput alignment demands of next-generation laser facilities such as Shenguang-II (SG-II). While methods based on traditional machine learning show some improvement, their overall performance remains limited by the representational capacity of handcrafted features.

In recent years, deep learning technology, particularly the convolutional neural network (CNN), has achieved a series of groundbreaking advances in the field of computer vision[19–21], offering novel research ideas and technical pathways for spot analysis. These methods can automatically learn deep, abstract features highly relevant to the task from raw image data, demonstrating strong generalization capabilities against complex conditions such as illumination variations and noise interference. Nevertheless, most current research efforts still focus on the preliminary detection and rough localization of spots. Key issues such as sub-pixel-level fine contour segmentation, automatic quantitative extraction of multi-dimensional morphological parameters and the robust, adaptive determination of reference centers under varying operational conditions still lack systematic solutions. Deep learning is adopted here because classical methods cannot meet three key requirements: pixel-level instance discrimination for overlapping spots; adaptive feature representation for variable spot morphologies across amplification stages; and an end-to-end pipeline integrating image interpretation and control parameter generation. Unlike generic computer vision tasks, laser spot analysis has domain-specific challenges (weak contrast, micrometer-scale positioning, real-time constraints), justifying specialized architectural adaptations instead of off-the-shelf models.

To address the requirement for intelligent, fully automated and rapid alignment of the SG-II laser facility, this paper presents the development of a deep learning-based spot analysis system. The core of this method is an improved YOLOv8-seg instance segmentation model, which enables accurate recognition and pixel-level contour segmentation of typical spots during alignment operations. Furthermore, a multi-mode alignment reference center detection framework is designed. This framework integrates a deep learning-driven cross-mark localization algorithm with a geometry-constrained four-dot recognition algorithm to achieve rapid detection and identification of the spatial reference benchmarks throughout the facility's alignment chain. The results obtained from the aforementioned methods are subsequently fed into a feature quantification module to automatically calculate spot centroids, morphological parameters and their offset vectors relative to the reference centers. The system achieves a network inference time of 55 ± 2 ms per frame, measured as the model forward-pass duration on an NVIDIA RTX 4060 GPU, averaged over three independent runs across the full validation set. The complete processing workflow encompasses image preprocessing, instance segmentation, multi-mode reference center detection and morphological feature quantification. With a network inference latency of only 55 ms and moderate additional overhead from the aforementioned stages, the total processing time remains comfortably within the 1 s cycle budget of the SG-II facility's alignment system, thereby meeting the real-time requirement for automated closed-loop alignment. The system outputs structured adjustment commands to drive motorized optical components.

2. System architecture and core algorithms

The system developed in this study is designed for the alignment of the SG-II high-power laser facility. The upgraded SG-II facility comprises eight nanosecond laser beams and a ninth beam capable of delivering either nanosecond or picosecond pulses[22]. Its alignment beamline is illustrated in Figure 1, where two sets of imaging systems are deployed at key alignment nodes of the beamline. The key specifications of the imaging systems are as follows: the system at position (1) in Figure 1 has an image resolution of 2040×2048 with a pixel size of 7.4 μm; the system at position (2) in Figure 1 has an image resolution of 782×582 with a pixel size of 8.3 μm.

In this paper, we propose three types of reference centers: the four-dot pattern, the cross-mark and the image geometric center. They are deployed at position (1) in Figure 1, position (2) in Figure 1 and the far-field position before main amplifier injection, respectively. Example images of these three types of reference centers are presented in Figure 2. The proposed spot analysis system is uniformly applied to all nine beamlines of the SG-II facility, and all experimental data used in this study are real alignment images captured from routine physical experiments on this operational large-scale high-power laser facility.

These reference centers are all physical benchmarks fixed during the facility installation and precision alignment stage, rather than digitally superimposed markers. The cross-mark is a physical reticle placed at the confocal plane of the rod-amplifier spatial filter, serving as the absolute far-field reference. The four-dot pattern consists of four physical pinholes arranged in a square at the focal plane of the cavity spatial filter (CSF), whose geometric center is defined as the absolute alignment reference. Both mechanical references are precisely calibrated to coincide with the theoretical optical axis, guaranteeing the stability and repeatability of alignment benchmarks.

https://doi.org/10.1017/hpl.2026.10167 Published online by Cambridge University Press

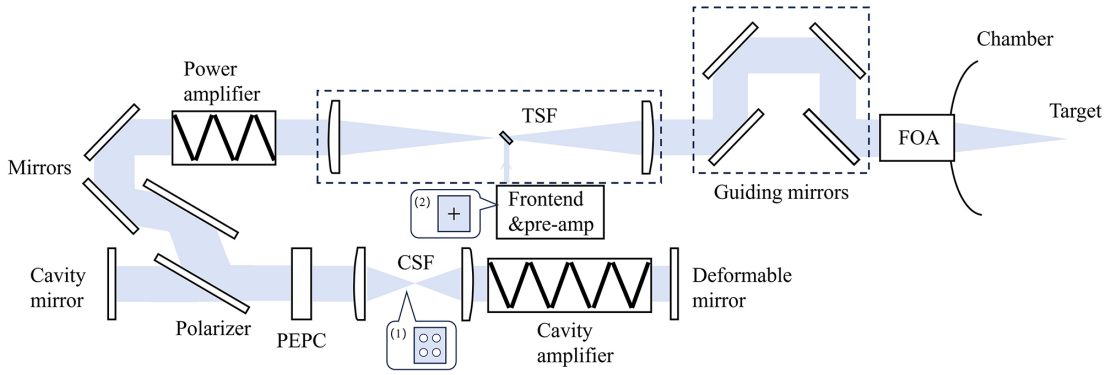


Figure 1. SG-II alignment beamline. TSF, transport spatial filter; CSF, cavity spatial filter; PEPC, plasma electrode Pockels cell.

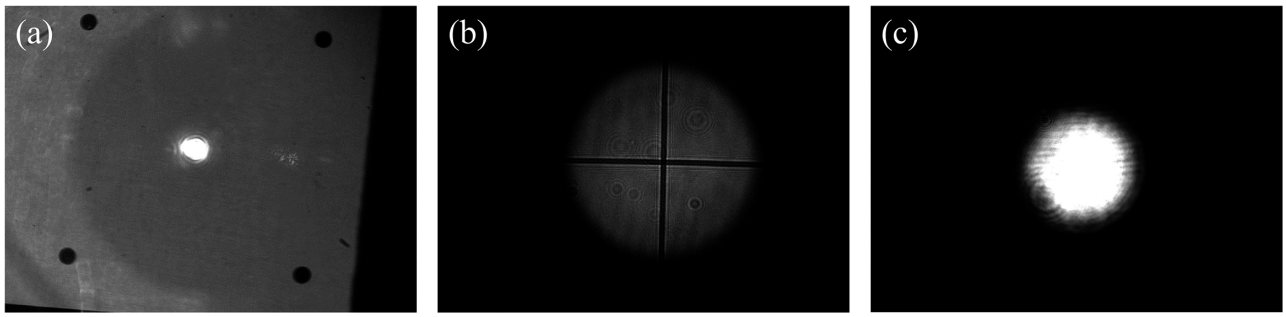


Figure 2. Three types of reference center images: (a) four-dot pattern, serving as the far-field reference from the grating in the main amplifier CSF cavity; (b) cross-mark, used as the near-field reference before main amplifier injection; (c) image geometric center, provided by the charge-coupled device (CCD) sensor plane as the far-field reference before main amplifier injection.

2.1. Overall system architecture

The entire processing system, as illustrated in Figure 3, consists of three modules: a raw image preprocessing module, an improved YOLOv8-seg module and a feature quantification module. The system takes raw image data $I \in \mathbb{R}^{H\times W\times C}$ captured by an alignment camera as its input. This input, a raw spot image with dimensions $H\times W\times C$ (where H is height, W is width and C is the number of channels), is first uniformly resized through preprocessing. It is then processed by the improved YOLOv8-seg model, which outputs an instance mask set $M = \{M_1, M_2, \cdots, M_N\}$ (where $M_i \in \{0,1\}^{H\times W}$ is the binary mask matrix for the ith spot (in which N is the total number of detected spots) and the reference center coordinates $C_{\text{ref}} = (x_{\text{ref}}, y_{\text{ref}})$. Subsequently, M and C_{ref} are fed into the feature quantification module. Through comprehensive computation, this module derives the geometric, morphological and spatial offset parameters for each spot, forming a feature set $F = \{f_1, f_2, \cdots, f_N\}$, thereby providing complete data support for alignment decision-making.

The uniform resizing to 800 × 800 pixels is performed to meet the YOLOv8-seg input requirements while using letterbox padding to maintain the aspect ratio and avoid geometric distortion. This preprocessing also ensures consistent annotation formats and enables efficient batch processing.

The calculated offset vectors and geometric parameters are further converted into adjustment direction, step size and step count for motorized optical components. These structured parameters support closed-loop alignment control with a processing latency within 1 s, matching the 1 Hz data acquisition rate of the facility's alignment system.

2.2. Improved YOLOv8-seg instance segmentation model

2.2.1. Loss function reconstruction and weight optimization

The general total loss function of the YOLOv8-seg model is composed of three weighted components: bounding box regression loss $\mathcal{L}_{\text{box}}$, classification loss $\mathcal{L}_{\text{cls}}$ and segmentation mask loss $\mathcal{L}_{\text{mask}}$. Among these, $\mathcal{L}_{\text{box}}$ employs complete intersection over union (CIoU) loss function, while $\mathcal{L}_{\text{cls}}$ and $\mathcal{L}_{\text{mask}}$ use binary cross-entropy (BCE) loss.

Spot analysis in this context involves three categories – circle, square and irregular – with circular targets constituting a high proportion in the segmentation task. To address this characteristic, this paper introduces a shape constraint term to reconstruct the loss function.

Considering that the circularity of the circle category spots is closest to the ideal value, a circularity constraint weight λ_{circle} is added to the mask loss to guide the predictions for circular spots closer to a geometric true circle. The corresponding loss function is as follows:

$$\mathcal{L}_{\text{circle}} = \sum_{(x,y)\in M_i} |\kappa_i - \kappa_{\text{ideal}}|, \tag{1}$$

https://doi.org/10.1017/hpl.2026.10167 Published online by Cambridge University Press

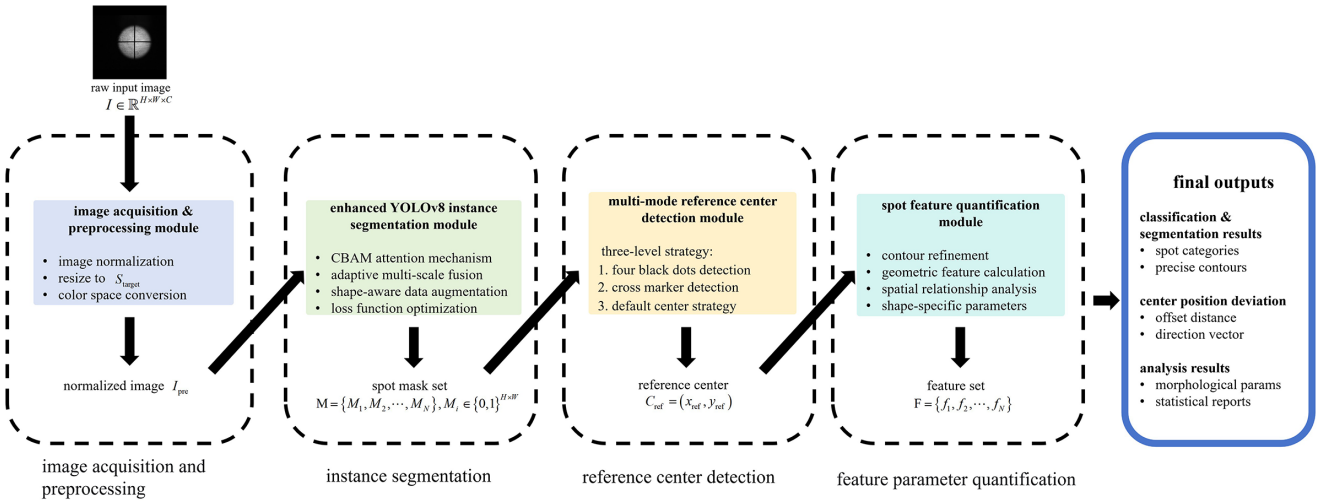


Figure 3. Architecture of the laser spot analysis system.

where κ represents the circularity, with a value of 1 denoting a perfect circle. This constraint encourages the predictions for circular spots to better approximate the ideal geometry.

The new, reconstructed total loss function is expressed as follows:

$$\mathcal{L}_{\text{total}} = \lambda_{\text{box}}\mathcal{L}_{\text{box}} + \lambda_{\text{cls}}\mathcal{L}_{\text{cls}} + \lambda_{\text{mask}}\mathcal{L}_{\text{mask}} + \lambda_{\text{circle}}\mathcal{L}_{\text{circle}}. \quad (2)$$

In the equation, λ_{box}, λ_{cls}, λ_{mask} and λ_{circle} are the weight coefficients for each loss term, and $\mathcal{L}_{\text{circle}}$ is the circularity constraint loss term. The values of the weight coefficients are set according to the configuration file, with the mask loss weight $\lambda_{\text{mask}} = 2.0$ and the circularity constraint weight $\lambda_{\text{circle}} = 1.5$, ensuring a balance among all loss components during the training process.

2.2.2. *Network architecture improvements*

In recent years, instance segmentation models have evolved into two dominant technical paradigms: (1) CNN-based detection-segmentation frameworks, including two-stage models such as Mask R-CNN and single-stage models such as the YOLO segmentation series, which excel in real-time inference and deployment-friendliness; (2) transformer-based end-to-end frameworks, pioneered by a detection transformer (DETR)[23] and hierarchical vision backbones such as Swin Transformer[24], which achieve strong global context modeling but often come with a high computational overhead and slow inference speed. Considering the real-time detection requirement and limited computing resources of the laser spot alignment scenario, this paper adopts the YOLOv8-seg model as the baseline and proposes targeted structural improvements. The improved model consists of a three-stage architecture: a backbone, a neck for feature fusion and a head, as shown in Figure 4. The detailed optimizations for the three core components are described below.

To clarify the actual processing effects of the proposed network on laser spot images, the corresponding visualization results are presented in Figure 5, which are strictly aligned with the core modules of the network architecture shown in Figure 4.

Firstly, attention mechanisms have become a core module to boost the feature representation capability of CNN models for dense prediction tasks. The pioneering squeeze-and-excitation networks (SENets)[25] first introduced channel-wise attention to adaptively recalibrate feature weights, while the efficient channel attention network (ECA-Net)[26] and subsequent visual attention frameworks[27] further optimized this paradigm with a lightweight local cross-channel interaction strategy, achieving a better trade-off between performance and computational cost. However, both the SENet and ECA-Net only model channel attention, ignoring spatial information that is critical for detecting weak boundary spots. Therefore, we embed the convolutional block attention module (CBAM)[28], which comprises channel and spatial attention sub-modules, as shown in Figure 6, into the C2f modules of the backbone. This enables the network to adaptively calibrate feature weights, thereby enhancing its perception of spots with low contrast and weak boundaries. Specifically, the input feature maps undergo channel attention learning (which learns weights for C channels, where C is the channel number of the raw image $I \in \mathbb{R}^{H\times W\times C}$) and spatial attention learning (which learns $H\times W$ spatial weights, where H and W are the height and width of $I \in \mathbb{R}^{H\times W\times C}$). The channel and spatial weights are then dynamically adjusted and combined to output an enhanced feature set, improving the capture capability for boundaries with low grayscale differences.

Secondly, within the neck, the path aggregation network (PANet) architecture was improved by introducing a dynamic weight allocation mechanism. The fusion weights for low-, medium- and high-level feature maps were changed

https://doi.org/10.1017/hpl.2026.10167 Published online by Cambridge University Press

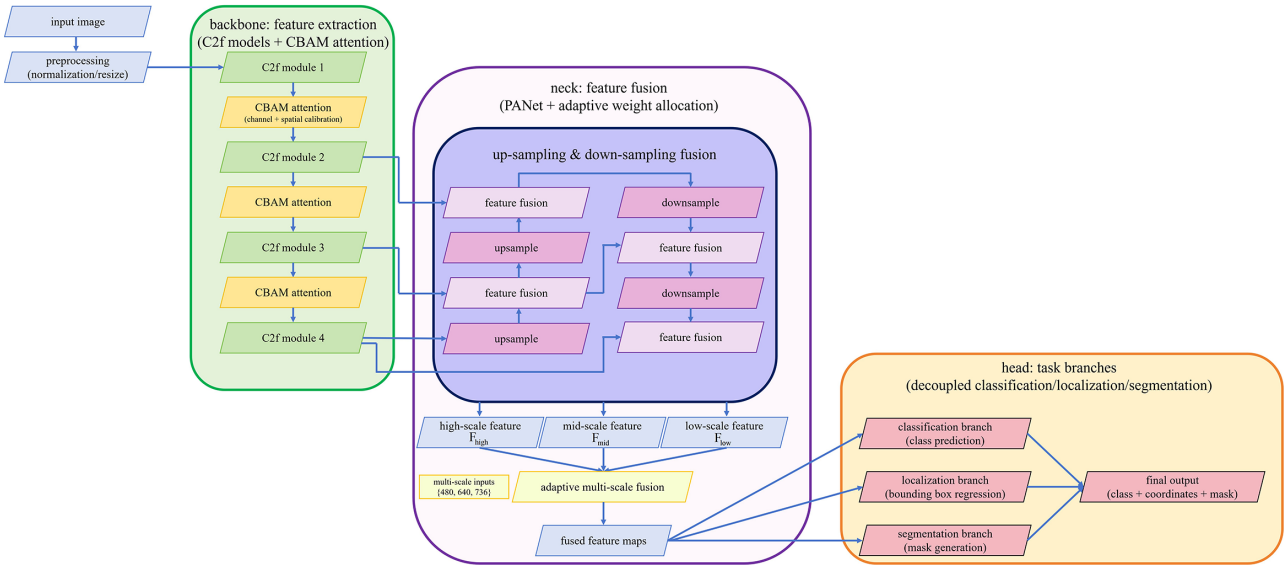


Figure 4. Improved YOLOv8-seg network architecture diagram.

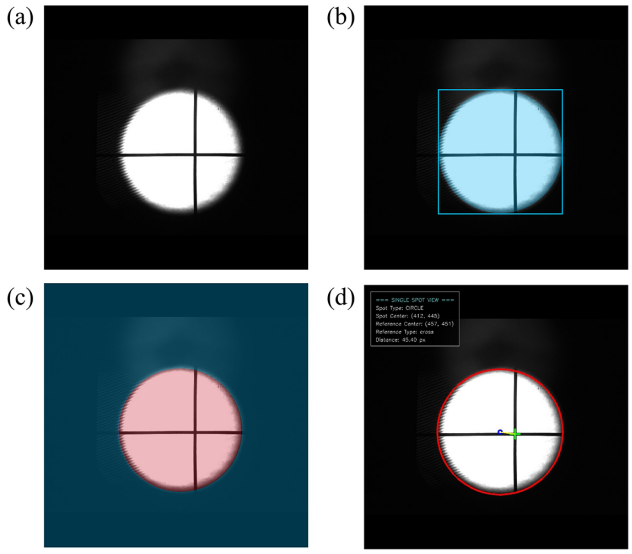


Figure 5. Workflow of the proposed laser spot analysis system: (a) raw input spot image with cross-mark reference, acquired from the SG-II facility; (b) laser spot detection result with a bounding box, output by the improved YOLOv8-seg model; (c) pixel-level instance segmentation mask from the segmentation branch, providing an accurate contour for morphological parameter calculation; (d) final center localization and quantitative result, including the spot center, reference center and offset distance.

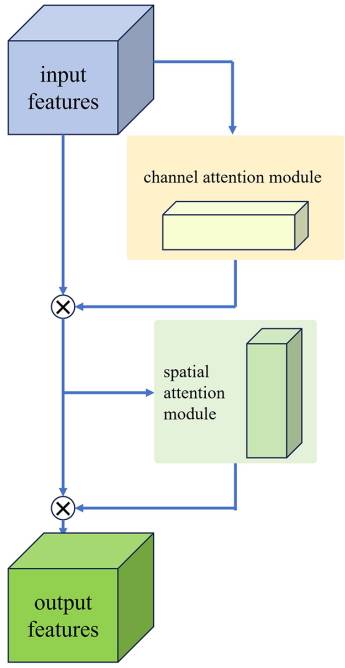


Figure 6. CBAM network structure diagram, where ⊗ represents convolution.

from fixed values to a ratio generated in real-time by a lightweight attention module. This allows the neck to adaptively adjust its fusion strategy based on the scale distribution of the input spots, avoiding the feature dilution problem associated with fixed weights.

Finally, the head contains three parallel branches for classification, localization and mask prediction, combined with a prototype network (ProtoNet) to generate the final segmentation masks. During inference, multi-scale testing augmentation is employed to ensure detection stability for spots ranging from 5 to 200 pixels in size.

During training, a comprehensive shape-aware data augmentation pipeline is applied to introduce controlled geometric and photometric variations that simulate real-world operational perturbations encountered in high-power laser beamlines, including thermal drift, optical component wear and illumination fluctuations. These perturbations are dynamically generated on-the-fly from the fixed training set during each iteration to enhance the model's robustness and

https://doi.org/10.1017/hpl.2026.10167 Published online by Cambridge University Press

Table 1. Ablation study on frozen backbone layer settings (network component analysis)[a].

Frozen layers	mAP@0.5	mAP@0.5:0.95	Convergence epoch	Final loss
0 (full fine-tuning)	0.898 ± 0.003	0.782 ± 0.004	85	0.234
5	0.899 ± 0.002	0.785 ± 0.003	78	0.228
10	0.901 ± 0.002	0.788 ± 0.003	72	0.221
15	0.894 ± 0.004	0.779 ± 0.005	68	0.235
22 (head only)	0.887 ± 0.005	0.771 ± 0.006	65	0.248

[a]Freezing 10 layers achieves optimal balance between convergence speed and accuracy. Freezing fewer layers risks overfitting; freezing more limits adaptation to laser spot characteristics. The term 'ablation study' here refers to machine learning-based network component contribution analysis, and does not refer to physical laser ablation.

generalization capability under varying imaging conditions, thereby improving its adaptability to diverse spot morphologies across different amplification stages. The data augmentation pipeline includes two categories of transformations: geometric transformations and HSV color space perturbations. The geometric transformations consist of random rotation (±5°), random isotropic scaling (0.7×–1.3×, uniform sampling), random perspective transformation and random horizontal flipping (applied with a probability of 0.5). For color perturbations, we apply random hue shift (±0.1, hue channel normalized to [0,1]), random saturation scaling (0.8×–1.2×, uniform sampling) and random value (brightness) scaling (0.8×–1.2×, uniform sampling).

2.2.3. Training strategy optimization

To effectively prevent overfitting and enhance training efficiency, this study adopts a transfer learning-based model initialization and training strategy. Specifically, the model is first initialized with weights pre-trained on a general-purpose dataset. Subsequently, the parameters of the first 10 layers in the backbone network are frozen, and only the detection head is fine-tuned.

Ablation study on frozen layer depth (0, 5, 10, 15 layers), as shown in Table 1, showed that freezing 10 layers achieves optimal balance: preserving low-level feature extractors while allowing sufficient adaptation to laser spot characteristics. Freezing fewer layers risks overfitting; freezing more limits adaptation.

The model is initialized with weights pre-trained on the Common Objects in Context (COCO) dataset (YOLOv8-seg official weights), which contains over 200,000 images across 80 categories. The first 10 layers of the backbone are frozen during fine-tuning to preserve low-level feature extractors while adapting higher layers to laser spot characteristics.

For the optimization strategy, a cosine annealing learning rate scheduler with a warm-up phase is employed. Given a total training epoch of E, a batch size of B and a learning rate η, the dynamic adjustment of the learning rate with respect to the current epoch $t \in [1, E]$ follows the rule

$$\eta(t) = \begin{cases} \eta_0 \cdot \frac{t}{T_{\text{warmup}}}, & 1 \le t < T_{\text{warmup}}, \\ \eta_{\min} + \frac{\eta_0 - \eta_{\min}}{2}\left[1 + \cos\frac{(t - T_{\text{warmup}})\pi}{E - T_{\text{warmup}}}\right], & T_{\text{warmup}} \le t \le E, \end{cases} \tag{3}$$

where $\eta_0 = 5 \times 10^{-4}$ is the initial learning rate, representing the maximum step size for parameter updates; $\eta_{\min} = 1 \times 10^{-5}$ is the minimum learning rate, preventing training stagnation due to the learning rate decaying to zero; and T_{warmup} is the number of warm-up epochs, during which the learning rate is linearly increased at the beginning of training to stabilize gradients. In addition, an early stopping mechanism is applied with a patience parameter set to 30 epochs, effectively avoiding the waste of computational resources.

2.3. Multi-mode reference center detection

To address the application scenarios involving multiple types of reference centers across the entire facility chain, this paper designs a three-level progressive reference center detection framework, as illustrated in Figure 7. This framework establishes a clear detection priority: four-dot pattern → cross-mark → image geometric center. Through redundant design, it ensures the continuous availability of the spatial reference frame under operational conditions such as occlusion, wear and strong glare. This design achieves a balance between detection robustness and accuracy through the priority mechanism, demonstrating higher reliability under various lighting and occlusion conditions compared to single-reference detection methods.

2.3.1. Four-dot pattern detection

The detection of the four-dot pattern employs a traditional image processing-based algorithm, with its core workflow detailed below.

Firstly, adaptive threshold segmentation is applied to the grayscale image I_g (as shown in Figure 8(a)). The local threshold for each pixel (x, y) is calculated as follows:

$$I_b(x,y) = \begin{cases} 255, & I_g(x,y) < \mu_{\text{local}}(x,y) - \beta \cdot \sigma_{\text{local}}(x,y), \\ 0, & \text{otherwise}, \end{cases} \tag{4}$$

where μ_{local} and σ_{local} represent the local mean and standard deviation, respectively, within a 31×31 neighborhood centered at (x, y), and β is an adjustment coefficient. Pixels

https://doi.org/10.1017/hpl.2026.10167 Published online by Cambridge University Press

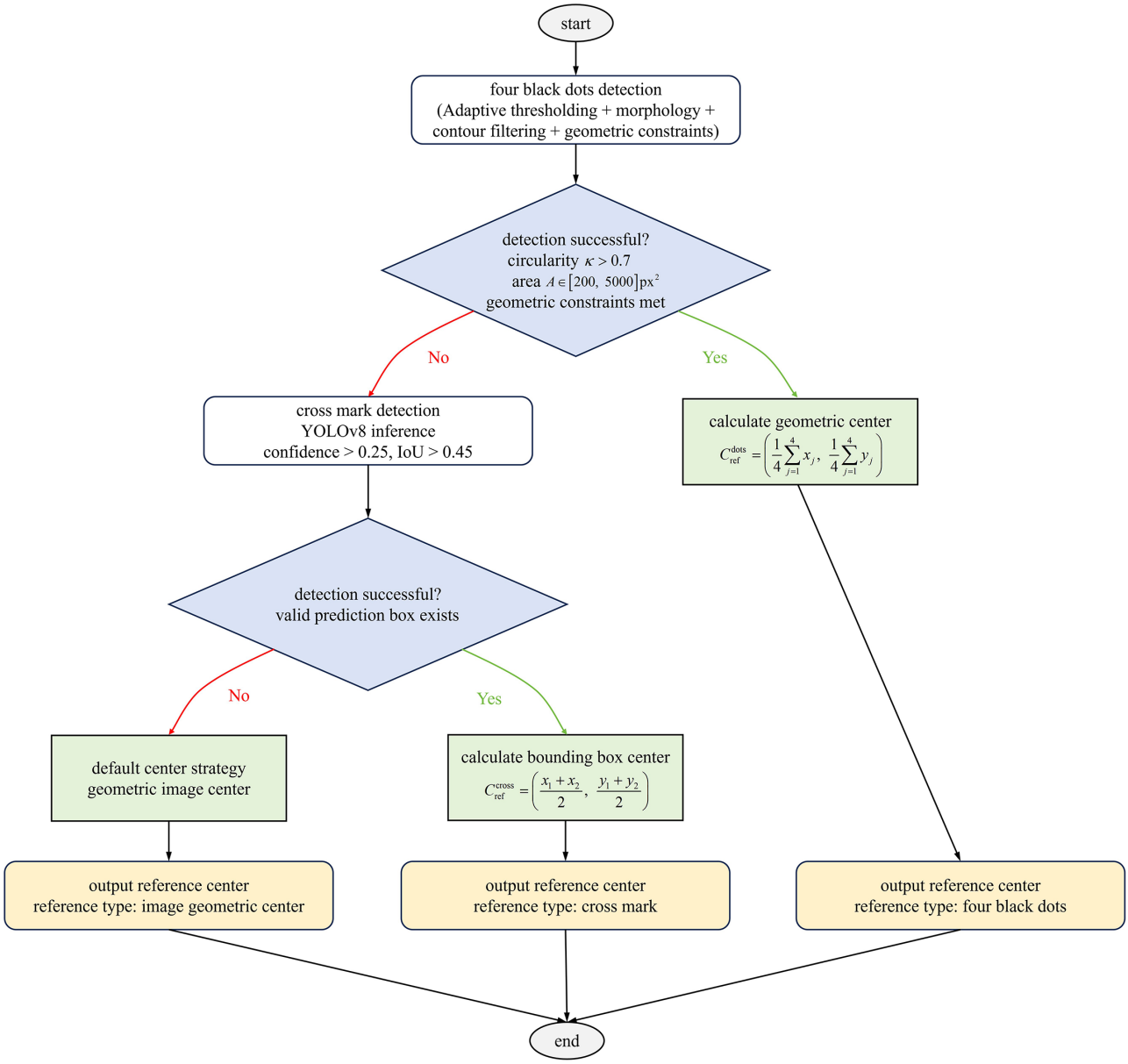


Figure 7. Three-level progressive reference center detection flowchart.

with grayscale values below this threshold are marked as foreground, generating a binary image I_{b} (Figure 8(b)).

Next, morphological filtering is performed on the binary image. An opening operation removes isolated noise points, and a closing operation fills holes within the targets, both using a 3×3 rectangular structuring element. The processed result is shown in Figure 8(c), from which all contours are extracted (Figure 8(d)). Candidate regions are then screened through contour analysis by calculating the area A_j, perimeter P_j and circularity $\kappa_j = \frac{4\pi A_j}{P_j^2}$ for each contour. Here, A_j is the total number of pixels within the contour region, P_j is the total length of the contour boundary and κ_j is a dimensionless parameter measuring how closely the contour approximates an ideal circle. Only contours with a circularity greater than 0.7 and an area within the range of $[200, 5000]$ pixels are retained, as shown in Figure 8(e).

Finally, geometric constraint verification is applied to the center points of the filtered contours. Given the coordinates of four candidate points (x_1, y_1), (x_2, y_2), (x_3, y_3) and (x_4, y_4), the aspect ratio, opposite-side length differences and diagonal length differences of the quadrilateral they form are calculated. The four-dot pattern is validated against specific geometric constraints: the aspect ratio must fall within 0.9–1.1 ($\pm 10\%$), the opposite-side length difference and diagonal length difference must not exceed 5%, the circularity must exceed 0.7 and the area must lie within [100, 5000] pixels. Only when all these geometric deviations remain below their respective preset thresholds are the four

https://doi.org/10.1017/hpl.2026.10167 Published online by Cambridge University Press

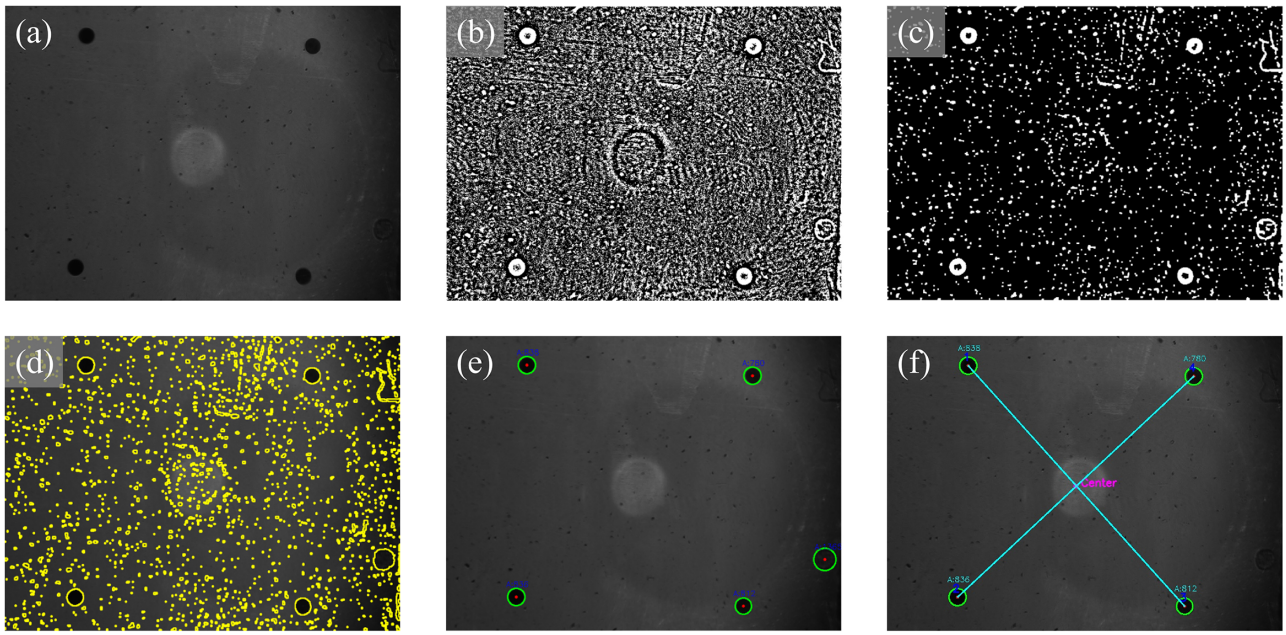


Figure 8. Schematic diagram of the four black dots center detection process: (a) original image; (b) after binarization; (c) after morphological processing; (d) all contours; (e) filtered contours; (f) reference center obtained after geometric constraints.

points recognized as constituting a valid square reference pattern. The reference center is then determined as the geometric center of these four points (Figure 8(f)), calculated by the following:

$$C_{\text{ref}}^{\text{dots}} = \left(\frac{x_1 + x_2 + x_3 + x_4}{4}, \frac{y_1 + y_2 + y_3 + y_4}{4} \right). \tag{5}$$

2.3.2. *Cross-mark detection*

For cross-mark detection, we adopt a dedicated lightweight YOLOv8 sub-network, with core design considerations as follows. As the key near-field reference for laser alignment, cross-mark imaging is frequently disturbed by partial occlusion, stray light and contrast fluctuation in practical beamline operation, where traditional template-matching methods have insufficient robustness to meet the high-reliability requirements of unattended closed-loop alignment. Meanwhile, this model can accurately extract the structural features of cross-marks to achieve stable high-precision localization; its unified architecture with the main improved YOLOv8-seg model also simplifies system deployment, and the ultra-low computational overhead fully guarantees the real-time performance of the system.

Specifically, the detector takes an input resolution of 640×640 pixels and relies on the CNN to encode deep visual patterns, such as the geometric structure of the cross, linear edge features and symmetry. Compared to traditional template-matching methods, this learning-based detection approach exhibits superior adaptability to disturbances such as partial occlusion and contrast fluctuations. Once the cross-mark is located, its reference center is defined as the center point of the predicted bounding box:

$$C_{\text{ref}}^{\text{cross}} = \left(\frac{x_1 + x_2}{2}, \frac{y_1 + y_2}{2} \right), \tag{6}$$

where (x_1, y_1) are the pixel coordinates of the top-left corner of the predicted box and (x_2, y_2) are the coordinates of the bottom-right corner. For model inference, post-processing is performed using a confidence threshold $\tau_{\text{conf}} = 0.25$ and an intersection over union (IoU) threshold $\tau_{\text{iou}} = 0.45$. This parameter combination, determined through receiver operating characteristic (ROC) analysis to reside at the optimal operating point, achieves a good balance between recall and precision.

2.3.3. *Fallback center strategy*

When both aforementioned detection levels fail, the system automatically adopts a protective fallback strategy that uses the calculated image geometric center as the reference benchmark, defined as follows:

$$C_{\text{ref}}^{\text{default}} = \left(\frac{W}{2}, \frac{H}{2} \right), \tag{7}$$

where W and H represent the width and height of the image in pixels, respectively.

2.4. *Spot feature extraction and post-processing optimization*

2.4.1. *Contour feature refinement*

Upon obtaining the segmentation masks output by the head, the spot contour is first extracted from the target image.

https://doi.org/10.1017/hpl.2026.10167 Published online by Cambridge University Press

The Douglas–Peucker algorithm[29] is then employed for polygon approximation to reduce the number of contour points. In this study, the algorithm's threshold is set to 0.5% of the contour perimeter. This threshold achieves a favorable balance between accuracy and efficiency on the validation set – it reduces the average number of contour points by 72% while introducing an increase in centroid coordinate error of less than 0.3 pixels, which has a negligible impact on the final sub-pixel analysis accuracy.

2.4.2. Geometric feature quantification

The system constructs a three-dimensional feature parameter system encompassing geometry, shape and spatial relationships, which include the following: area $A_i = \sum_{(x,y)} M_i(x,y)$, defined as the total number of foreground pixels of the ith spot; perimeter $P_i = \sum_j \left\| p_{j+1} - p_j \right\|$, the total length of the refined contour chain code; centroid coordinates $\left(c_{x_i}, c_{y_i}\right)$, satisfying $c_{x_i} = \frac{M_{10}}{M_{00}}, c_{y_i} = \frac{M_{10}}{M_{00}}$, where $M_{pq} = \sum_x \sum_y x^p y^q M_i(x,y)$ is the $(p+q)$th-order geometric moment, physically representing the weighted sum of pixel coordinates and grayscale values; circularity κ_i, a dimensionless parameter calculated as $\kappa_i = \frac{4\pi A_i}{P_i^2}$; equivalent radius $r_{\mathrm{eq},i} = \sqrt{\frac{A_i}{\pi}}$; offset distance $d_i = \sqrt{\left(c_{x_i} - x_{\mathrm{ref}}\right)^2 + \left(c_{y_i} - y_{\mathrm{ref}}\right)^2}$, the Euclidean distance from the spot's centroid to the reference center; and azimuth angle $\theta_i = \arctan \frac{c_{y_i} - y_{\mathrm{ref}}}{c_{x_i} - x_{\mathrm{ref}}}$, the direction angle relative to the reference center.

3. Results and discussion

To comprehensively evaluate the overall processing capability of the proposed system, multiple sets of experiments were designed and conducted. The experimental design covers various aspects including overall model performance, ablation analysis, comparative experiments, convergence studies and reference detection accuracy, aiming to provide a holistic assessment of the method's comprehensive efficacy and practical engineering applicability. All comparative experiments adhere to a consistent training–validation–test dataset split to ensure the comparability of the results. The final reported values are averaged over three independent runs using different random seeds.

3.1. Dataset and experimental environment

3.1.1. Dataset construction

All data for the validation experiments were sourced from laser spot images acquired during the operation of the SG-II facility from 2024 to 2025. The dataset comprises 1214 training images and 304 validation images, encompassing three typical spot morphologies: 'circle', 'square' and 'irregular'.

Table 2. Distribution of the laser spot dataset.

Category	Training samples	Validation samples	Proportion/%
Circle	607	152	50.0
Square	364	91	29.9
Irregular	243	61	20.1
Total	1214	304	100

These three morphologies coexist simultaneously during facility operation and correspond to spot patterns at different optical path positions. Specifically, the circular spot is the near-field morphology of the rod amplifier before main amplification injection, the square spot is the near-field morphology at the main amplification output and the irregular spot is the far-field morphology at the main amplification output.

Standard examples of these three categories are shown in Figure 9. All images were uniformly resized to 800×800 pixels during preprocessing. Pixel-level annotation was performed using the LabelMe tool and subsequently converted to the YOLO format.

While the dataset size (1214 training + 304 validation images) is modest for deep learning, strong pre-training on the COCO database and data augmentation effectively mitigate overfitting risks. Three independent runs show stable performance, indicating good generalization.

The detailed category distribution of the dataset is presented in Table 2. It can be observed that the 'circle' category constitutes approximately 50% of the dataset, which aligns with the most common physical characteristic of far-field spots in the actual operation of the SG-II facility.

The dataset is mainly derived from normal SG-II facility alignment scenarios, with approximately 20% low signal-to-noise ratio (SNR) (<15 dB) samples included, while systematic coverage of extreme anomalies (beam missing, strong saturation, severe misalignment, out-of-reference spots) is limited due to their extremely low occurrence frequency and insufficient annotated samples. For real-time alignment, a graceful degradation mechanism is adopted: the system raises detection failure flags for abnormal cases instead of outputting erroneous estimates, triggering manual intervention to ensure operational safety.

3.1.2. Experimental environment

The experiments were conducted on a computer running the Windows 11 operating system. The hardware configuration included an Intel Core i9-14900HX processor (with a base frequency of 2.20 GHz) and an NVIDIA GeForce RTX 4060 Laptop GPU (with 16 GB of VRAM). The software environment consisted of Python 3.13, PyTorch 2.8.0 and CUDA 12.9. The training was implemented based on the Ultralytics YOLOv8 framework.

https://doi.org/10.1017/hpl.2026.10167 Published online by Cambridge University Press

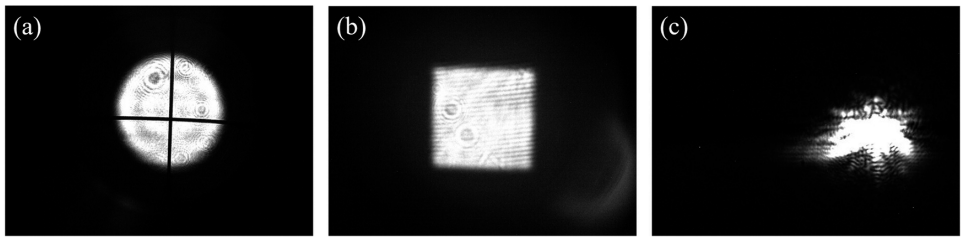


Figure 9. Examples of three types of laser spots: (a) circle; (b) square; (c) irregular.

3.2. Evaluation metrics

This study employs a three-tier evaluation system encompassing segmentation accuracy, localization accuracy and reference center accuracy.

Segmentation performance is primarily evaluated using the mean average precision (mAP), which is calculated strictly following the standard COCO evaluation protocol for instance segmentation tasks. The core calculation rules are defined as follows: average precision (AP) is computed as the area under the precision–recall curve for each spot category, and mAP is the arithmetic mean of AP across all three spot categories. The IoU between the predicted segmentation mask and the ground-truth mask is used as the core criterion to distinguish true positive and false positive predictions. Two mAP variants are adopted in this study:

(1) mAP@0.5 uses a fixed IoU threshold of 0.5, and the final value is the mean of AP for all categories under this single threshold;

(2) mAP@0.5:0.95 averages the AP results over 10 evenly spaced IoU thresholds from 0.5 to 0.95 with a step size of 0.05, serving as the primary metric for comprehensive segmentation accuracy evaluation.

Localization accuracy is quantified by the Euclidean distance between the predicted centroid and the annotated ground-truth centroid. Reference center accuracy is comprehensively assessed based on the localization error and the detection success rate, the latter defined as the proportion of samples where the localization error is less than 5 pixels.

3.3. Ablation study (network component analysis) and module contribution validation

3.3.1. Impact of backbone network improvements

Progressive ablation experiments demonstrate that the baseline YOLOv8-seg model achieves a baseline mAP@0.5 performance of 0.872 ± 0.003. The introduction of the CBAM module increases this score to 0.891 ± 0.002, an improvement of 2.18%, validating its effectiveness in extracting features from low-contrast spots. Although the dynamic multi-scale fusion strategy causes a slight overall decrease in mAP@0.5 to 0.885 ± 0.004, it increases the recall rate for the irregular category by 4.1%, indicating enhanced adaptability to deformed targets. The circularity constraint loss function raises the average circularity of circle-type spots from 0.832 to 0.896.

The full model, integrating all improvements, achieves an mAP@0.5 of 0.901±0.002, representing a 3.33% improvement over the baseline, with only a 10 ms increase in inference time. It maintains real-time processing capability at 55 ms per frame on an RTX 4060 GPU. The detailed results are listed in Table 3.

To further analyze the model's performance across different spot types, we present the confusion matrix and per-class metrics in Figure 10. The results demonstrate that the model achieves consistent performance across all three categories.

The model achieves consistent and competitive detection performance across all three spot categories, with a mean AP@0.5 of 0.901. Notably, the circle class achieves the highest AP@0.5 (0.912), with a precision of 0.926 and recall of 0.908, benefiting from its distinct circular shape and abundant training samples (50% of the dataset). The irregular class, despite having the smallest sample size (20% of the dataset), maintains excellent performance with an AP@0.5 of 0.893 and the highest F1-score of 0.918, demonstrating the model's outstanding robustness to shape variations. The square class also delivers balanced detection capability with a high recall of 0.901, ensuring reliable and low-miss-rate detection in practical laser alignment scenarios, as shown in Table 4.

3.3.2. Impact of post-processing optimization on contour quality

The post-processing module takes the raw segmentation mask as input and sequentially performs morphological filtering and shape constraint optimization. Without post-processing, the radius error for circle-type spots was 2.35 ± 0.15 pixels, and the circularity error was 0.068 ± 0.008. After introducing the morphological filtering layer (with a 1×1 opening kernel and a 4×4 closing kernel), the radius error decreased to 2.01 ± 0.11 pixels (an improvement of 14.5%, $p < 0.01$), the circularity error decreased to 0.051 ± 0.006, and the processing time increased by 3 ms. Upon adding the circularity constraint layer (which blends the predicted mask with an ideal circle in a 3:7 weight ratio), the radius error was further reduced to 1.92 ± 0.09 pixels (a cumulative improvement of 18.3%, $p < 0.01$) and the

https://doi.org/10.1017/hpl.2026.10167 Published online by Cambridge University Press

Table 3. Results of ablation experiments on backbone network improvements (network component analysis)[a].

Model configuration	mAP@0.5	mAP@0.5:0.95	Inference time/ms	Training GPU memory/GB
YOLOv8-seg (baseline)	0.872 ± 0.003	0.756 ± 0.004	45 ± 1	8.7 ± 0.2
+ CBAM attention	0.891 ± 0.002*	0.774 ± 0.003*	48 ± 1	8.9 ± 0.2
+ Dynamic multi-scale fusion	0.885 ± 0.004	0.781 ± 0.002	52 ± 1	9.1 ± 0.2
+ Circle constraint loss	0.888 ± 0.003	0.771 ± 0.005	47 ± 1	8.8 ± 0.2
Full model	0.901 ± 0.002**	0.788 ± 0.003**	55 ± 2	9.2 ± 0.2

[a]The term 'ablation experiments' here refers to network component validation in machine learning, not the physical process of laser ablation. * and ** indicate statistically significant improvement compared to baseline (*$p < 0.01$, **$p < 0.005$).

Table 4. Per-class performance metrics.

Class	AP@0.5	AP@0.5:0.95	Precision	Recall	F1
Circle	0.912	0.798	0.926	0.908	0.917
Square	0.898	0.782	0.918	0.918	0.918
Irregular	0.893	0.784	0.872	0.901	0.886
Mean	0.901	0.788	0.905	0.909	0.907

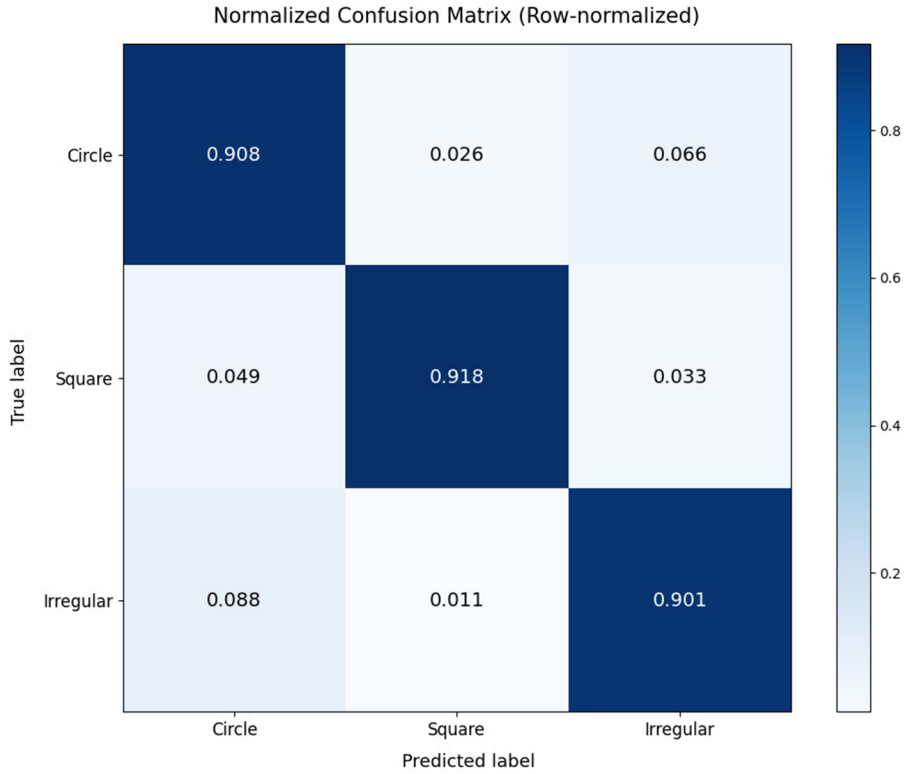


Figure 10. Confusion matrix for three spot types at the IoU threshold of 0.5.

circularity error decreased to 0.041 ± 0.005, with a total processing time increase of 8 ms. The effects of post-processing optimization are summarized in Table 5.

3.3.3. Training dynamics analysis and performance evaluation

This study provides a detailed analysis of the model's performance throughout the training process, with a specific focus on the evolution of the segmentation loss and the mAP.

Figure 11 illustrates the variation of the segmentation loss during training and validation. After a brief peak at epoch 56, the segmentation loss continued to decline, indicating stable learning convergence. The brief validation loss peak at epoch 56 coincides with the transition from learning rate warm-up to cosine annealing schedule. This transient instability quickly resolves as the optimizer adapts to the new learning rate regime, demonstrating good training robustness.

Figure 12 shows the progression of the segmentation mAP metrics across training epochs. It can be observed that both mAP@0.5 and mAP@0.5:0.95 increased rapidly during the initial phase before stabilizing. The final mAP@0.5 approached 0.9, while mAP@0.5:0.95 stabilized around

https://doi.org/10.1017/hpl.2026.10167 Published online by Cambridge University Press

Table 5. Summary of post-processing optimization effects[a].

Post-processing configuration	Radius error/pixel	Circularity error	Processing time/ms
No post-processing	2.35 ± 0.15	0.068 ± 0.008	Baseline
+ Morphological filtering	2.01 ± 0.11	0.051 ± 0.006	+3
+ Circle constraint	1.92 ± 0.09*	0.041 ± 0.005*	+5
Complete post-processing	1.92 ± 0.09*	0.041 ± 0.005*	+8

[a]* indicates statistically significant improvement ($p < 0.01$).

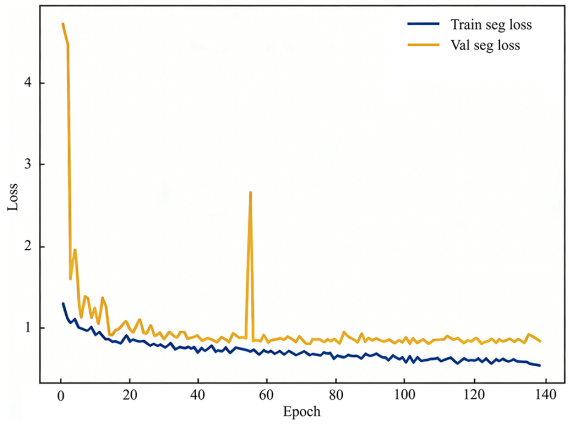


Figure 11. Segmentation loss trend.

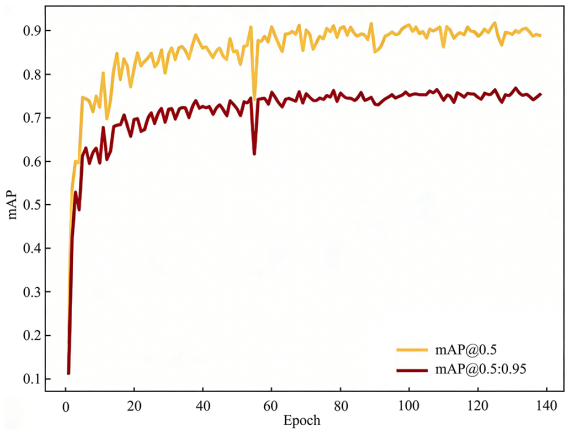


Figure 12. Segmentation mAP metrics trend.

0.75, demonstrating a steady improvement in segmentation performance.

The loss variation curve indicates that the model can stably learn effective feature representations from the data. As training progresses, the loss value continuously decreases, reflecting a gradual improvement in the model's fit to the training samples and a corresponding reduction in prediction error. The transient peak in the mid-training phase is attributed to the learning rate strategy transition, rather than model overfitting or dataset bias. The model quickly recovers and continues to optimize thereafter, further validating its excellent training stability and robustness. Concurrently, the mAP metrics show a steady upward trend, with a particularly noticeable improvement during the initial training stage, indicating that the model's segmentation performance is substantively enhanced as training proceeds. Overall, the proposed model exhibits stable convergence characteristics and continuous performance improvement throughout the training process, and converges stably to high performance, providing a reliable foundation for its further optimization and practical deployment.

3.4. Comparative validation with mainstream methods

To validate the performance of the proposed system in practical applications and provide an objective comparison with current mainstream image segmentation models, this section conducts comparative experiments using a separately acquired standard test set against classical segmentation models, such as U-Net and Mask R-CNN. This test set is independent from the validation set (304 images) used in the ablation experiments in terms of acquisition time and sample distribution. It specifically includes approximately 20% of low SNR (<15 dB) samples to thoroughly evaluate the generalization capability of each model on unseen data, particularly under challenging conditions.

All comparison models use identical training/validation data splits and data augmentation strategies. Each model is trained for 100 epochs with its default optimizer configuration to ensure fair comparison at their respective optimal settings.

For comprehensive evaluation, we also include DeepLabv3+[30] and SegFormer[31], two representative semantic segmentation methods. DeepLabv3+ employs atrous spatial pyramid pooling (ASPP) for multi-scale feature extraction, while SegFormer leverages transformer architecture for global context modeling. These methods complement the comparison by representing different architectural paradigms: CNN-based encoder–decoder (U-Net, DeepLabv3+), two-stage detector (Mask R-CNN), transformer-based (SegFormer) and single-stage instance segmentation (YOLOv8-seg).

As shown in Table 6, the comparative results indicate that U-Net's performance is significantly limited in overlapping spot scenarios due to its lack of an instance discrimination mechanism, achieving an mAP@0.5 of only 0.853 ± 0.005 with an inference time of 56 ± 2 ms/frame, making it difficult to meet alignment requirements. Although Mask R-CNN achieves a precision of 0.879 ± 0.004, its two-stage

https://doi.org/10.1017/hpl.2026.10167 Published online by Cambridge University Press

Table 6. Performance comparison among different segmentation methods[a].

Method	mAP@0.5	mAP@0.5:0.95	Inference time/ms	Parameters/10^6	FLOPs/(10^9 Mac)	Inference memory usage/GB
U-Net	0.853 ± 0.005	0.732 ± 0.006	56 ± 2	31.0	18.2	2.8
Mask R-CNN	0.879 ± 0.004	0.761 ± 0.005	83 ± 3	44.2	42.5	4.1
DeepLabv3+	0.871 ± 0.004	0.748 ± 0.005	45 ± 2	35.6	24.8	3.2
SegFormer	0.884 ± 0.003	0.769 ± 0.004	67 ± 2	37.8	28.3	3.5
Proposed system	0.901 ± 0.002*	0.765 ± 0.003	55 ± 2*	26.5*	15.7*	2.1*

[a]Generalization to other laser facilities with different imaging parameters (pixel size, magnification, spot characteristics) remains to be validated. The proposed method framework can be adapted through fine-tuning on target facility data. It should be noted that the real-time inference performance reported in this paper is obtained from offline platform tests. * indicates statistically significant improvement ($p < 0.01$).

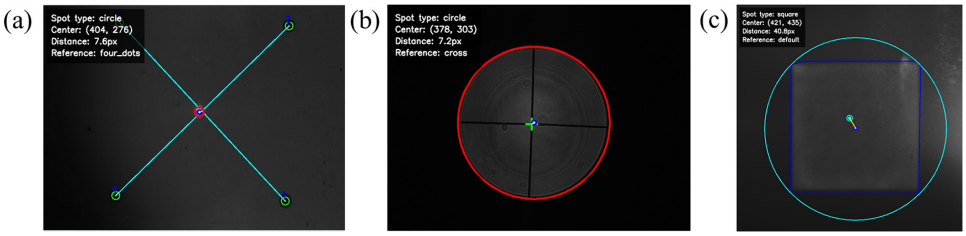


Figure 13. Effects of three types of reference center detection: (a) four-dot pattern; (b) cross-mark; (c) image geometric center.

detection pipeline results in a heavy computational burden, with an inference time of 83 ± 3 ms/frame, a parameter count of 4.42×10^7 and excessive computational overhead. Among the additional semantic segmentation methods included for comprehensive comparison, DeepLabv3+ achieves mAP@0.5 = 0.871 ± 0.004 with an inference time of 45 ± 2 ms/frame, outperforming U-Net but still limited by its semantic segmentation architecture in instance-level tasks, while SegFormer achieves mAP@0.5 = 0.884 ± 0.003 with an inference time of 67 ± 2 ms/frame, demonstrating the effectiveness of transformer architectures, but its higher computational cost (3.78×10^7 parameters, 2.83×10^{10} Mac FLOPs) and slower inference make it less suitable for real-time applications. All these methods lack explicit instance discrimination mechanisms, resulting in performance degradation when handling overlapping spots. In contrast, the proposed system achieves mAP@0.5 = 0.901 ± 0.002 and mAP@0.5:0.95 = 0.765 ± 0.003 on the standard test set, with an inference time of only 55 ± 2 ms/frame (comparable to U-Net at 56 ms but with 4.8% higher accuracy), more than 30% faster than Mask R-CNN (83 ms versus 55 ms) and 18% faster than SegFormer (67 ms versus 55 ms). Furthermore, it maintains the lowest resource consumption among all compared methods with only 2.65×10^7 parameters, 1.57×10^{10} Mac FLOPs and 2.1 GB memory usage. The proposed system attains the optimal balance of accuracy, speed and efficiency, with the 55 ms inference time being significantly lower than the 1 Hz cycle requirement of the SG-II laser facility, demonstrating significant advantages for practical deployment.

3.5. Evaluation of reference center detection accuracy

The three-level progressive reference detection framework proposed in this paper demonstrates robust performance on a 500-image test set, with recognition results for the three types of centers shown as examples in Figure 13. Performance evaluation on the standard test set ($n = 500$) shows that the four-dot pattern detection achieves a localization error of 1.8 ± 0.3 pixels with a success rate of 98.5%. This high precision primarily stems from its strict geometric constraint mechanism. The cross-mark detection strategy, implemented via a lightweight YOLOv8 model (with only 2.3×10^6 parameters), achieves robust localization under partial occlusion, with a detection error of 1.9 ± 0.4 pixels and a success rate of 96.2%. When both aforementioned detection levels fail, the system automatically defaults to using the image geometric center.

3.6. Visual analysis of experimental results

To intuitively validate the practical performance of the proposed method, this section presents the spot segmentation results and reference center identification outcomes under typical scenarios.

Figure 14 demonstrates the segmentation effects for the three spot types. The results in typical scenarios show that the system achieves high-precision segmentation and parameter extraction for circle-type spots (radius: 143.0 pixels, offset: 10.3 pixels), square-type spots (side length: 303.8 pixels,

https://doi.org/10.1017/hpl.2026.10167 Published online by Cambridge University Press

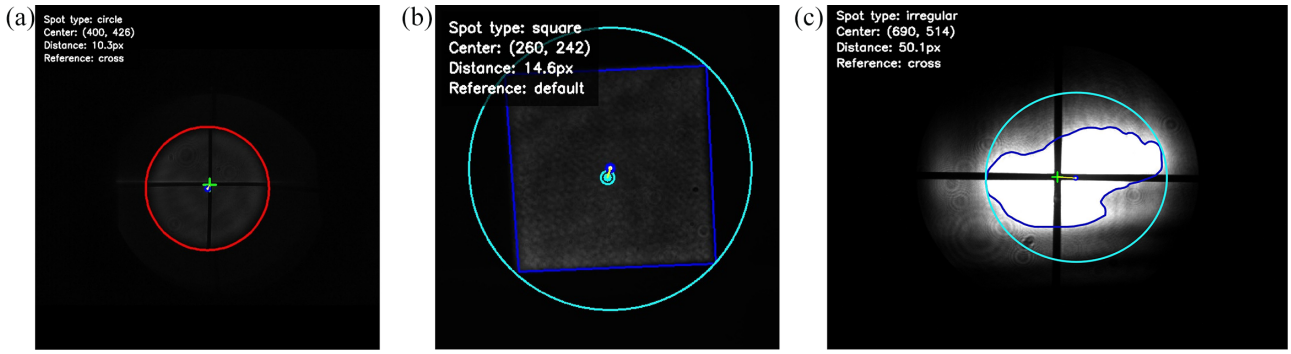


Figure 14. Segmentation results of the three spot types: (a) circle; (b) square; (c) irregular.

offset: 14.6 pixels, rotation angle: 3.38°) and irregular-type spots (offset: 50.1 pixels). These results verify the method's strong adaptability to diverse spot morphologies.

4. Conclusion

To address the fundamental limitations of classical alignment pipelines and the demand for automated alignment of the SG-II high-power laser facility, this study proposes a full-stack intelligent spot analysis system. The core breakthroughs of this work lie in three systematic methodological innovations for high-power laser alignment, which go far beyond incremental architectural optimizations to the instance segmentation model.

(1) An end-to-end perception-to-control alignment framework that directly maps raw spot images to executable motor control parameters, eliminating cumulative errors from multi-stage decoupled processing in classical pipelines.

(2) A three-level progressive reference center detection methodology with priority-based redundant design, which solves the long-standing reliability bottleneck of single-reference methods under complex imaging conditions.

(3) A multi-objective collaborative optimization methodology that balances segmentation accuracy, real-time inference performance and operational safety for engineering pre-deployment.

Guided by the above methodological framework, we further optimize the YOLOv8-seg instance segmentation model via circularity constraint loss, CBAM attention mechanism and dynamic multi-scale fusion. The optimized model achieves an mAP@0.5 of 0.901 with superior parameter efficiency compared to mainstream segmentation models. Offline validation fully confirms the effectiveness of the proposed methodological innovations: the three-level progressive reference detection framework achieves a localization error within 1.9 pixels and a 98.7% success rate, with significantly higher reliability than single-mode detection methods; the end-to-end alignment framework reduces full-process processing latency to 0.8 s, well below the 1 Hz cycle requirement of the facility; the multi-objective optimization methodology enables the system to achieve the optimal balance among accuracy, speed and resource consumption, with only 2.65×10^7 parameters and 2.1 GB memory usage.

Compared to current mainstream methods, the proposed system demonstrates excellent performance in segmentation accuracy, real-time capability and parameter efficiency.

At the current stage, the system has completed full offline test validation, and we are conducting performance testing with the latest alignment image data from the facility. After the processing speed and detection success rate meet the engineering requirements, the system will be deployed to the facility for online verification.

Subsequent research will focus on the practical deployment of the alignment system. This includes model lightweighting to adapt to multi-beamline parallel processing, exploring online incremental learning to counteract performance drift caused by optical component aging and integrating digital twin technology to achieve predictive control of the alignment process. These efforts aim to advance the facility's beamline alignment from a semi-automatic adjustment mode towards an intelligent and autonomous operational paradigm.

Acknowledgement

This work was supported by the Operation Project of the Shenguang-II Laser Facility, National Laboratory on High Power Laser and Physics, Shanghai Institute of Optics and Fine Mechanics, Chinese Academy of Sciences (Grant No. 86CL020101).

References

1. R. Betti and O. A. Hurricane, Nat. Phys. **12**, 435 (2016).
2. S. Burkhart, E. Bliss, P. Di Nicola, D. Kalantar, R. Lowe-Webb, T. McCarville, D. Nelson, T. Salmon, T. Schindler, J. Villanueva, and K. Wilhelmsen, Proc. SPIE **7797**, 77970M (2010).

https://doi.org/10.1017/hpl.2026.10167 Published online by Cambridge University Press

3. S. Burkhart, E. Bliss, P. Di Nicola, D. Kalantar, R. Lowe-Webb, T. McCarville, D. Nelson, T. Salmon, T. Schindler, and J. Villanueva, Appl. Opt. **50**, 1136 (2011).
4. M. L. Spaeth, K. Manes, D. Kalantar, P. Miller, J. Heebner, E. Bliss, D. Spec, T. Parham, P. Whitman, and P. Wegner, Fusion Sci. Technol. **69**, 25 (2016).
5. A. B. Zylstra, O. A. Hurricane, D. A. Callahan, *et al.*, Nature **601**, 542 (2022).
6. M. Luttmann, V. Denis, P. Gendeau, C. Lanternier, P. Baudon, Y. Schiano, L. Espert, F. Seguineau, O. Henry, and D. Raffestin, J. Phys. Conf. Ser. **717**, 012106 (2016).
7. J. Miquel, C. Lion, and P. Vivini, J. Phys. Conf. Ser. **688**, 012067 (2016).
8. J. Zhu, P. Yang, Y. Zhang, C. Liu, S. Zhou, Z. Liu, S. Tang, A. Guo, Q. Fan, and D. Liu, Acta Opt. Sin. **43**, 0822005 (2023).
9. Y. Ke, C. Zeng, P. Xie, Q. Jiang, K. Liang, Z. Yang, and M. Zhao, Appl. Opt. **54**, 4876 (2015).
10. N. Otsu, Automatica **11**, 23 (1975).
11. T. W. Ridler and S. Calvard, IEEE Trans. Syst. Man Cybern. **8**, 630 (1978).
12. J. N. Kapur, P. K. Sahoo, and A. K. Wong, Comput. Vis. Graph. Image Process. **29**, 273 (1985).
13. J. Kittler and J. Illingworth, Pattern Recogn. **19**, 41 (1986).
14. I. Sobel and G. Feldman, "A 3×3 isotropic gradient operator for image processing," a talk at the Stanford Artificial Project (1968), pp. 271–272.
15. J. M. Prewitt, *Picture Processing and Psychopictorics* (Academic Press, 1970), p. 75149.
16. D. Marr and E. Hildreth, Proc. R Soc. Lond. Ser. B. Biol. Sci. **207**, 187 (1980).
17. M.-K. Hu, IRE Trans. Inf. Theory **8**, 179 (1962).
18. M. H. Singer, Pattern Recogn. **26**, 1019 (1993).
19. A. Krizhevsky, I. Sutskever, and G. E. Hinton, Commun. ACM **60**, 84 (2017).
20. J. Long, E. Shelhamer, and T. Darrell, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2015), p. 3431.
21. R. K. Gupta, G. D. Bruce, S. J. Powis, and K. Dholakia, Laser Photonics Rev. **14**, 2000120 (2020).
22. J. Zhu, J. Zhu, X. Li, B. Zhu, W. Ma, X. Lu, W. Fan, Z. Liu, S. Zhou, and G. Xu, High Power Laser Sci. Eng. **6**, e21 (2018).
23. N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, in *European Conference on Computer Vision* (2020), p. 213.
24. Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, in *Proceedings of the IEEE/CVF International Conference on Computer Vision* (2021), p. 10012.
25. J. Hu, L. Shen, and G. Sun, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2018), p. 7132.
26. Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2020), p. 11534.
27. M.-H. Guo, C.-Z. Lu, Z.-N. Liu, M.-M. Cheng, and S.-M. Hu, Comput. Vis. Media **9**, 733 (2023).
28. S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, in *Proceedings of the European Conference on Computer Vision (ECCV)* (2018), p. 3.
29. D. H. Douglas and T. K. Peucker, Cartographica **10**, 112 (1973).
30. L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, in *Computer Vision – ECCV 2018* (2018), p. 833.
31. E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, and P. Luo, in *Proceedings of the 35th International Conference on Neural Information Processing Systems* (2021), p. 12077.

https://doi.org/10.1017/hpl.2026.10167 Published online by Cambridge University Press