YOLO26-SimAM: An Energy-Based Attention Augmented Detector for Aero-Engine Surface Defect Inspection

Xiao Wang*[1][0009−0006−6210−0711], Linhao Liu*[1][0009−0001−5795−1558], Yidi Song[1][0009−0002−6082−6229], Xiaotong He[1][0009−0000−6455−3278], Kai Chen[1][0009−0002−1981−3960], and Nianyin Zeng[1][0000−0002−6957−2942]†

[1]School of Aerospace Engineering, Xiamen University, Fujian 361005, China
{zny@xmu.edu.cn}
* Equal contribution. † Corresponding author.

Abstract. Surface defect detection on critical aero-engine components is pivotal for ensuring flight safety. Addressing challenges such as computational resource constraints, minute defect targets, and severe interference from metallic surface noise, this paper proposes a lightweight, high-precision real-time defect detection model. The approach adopts the latest YOLO26-n as the base network, fully leveraging its efficient, Non-Maximum Suppression (NMS)-free architecture optimized for edge devices. Innovatively, the Simple Attention Module (SimAM) parameter-free attention mechanism is integrated at a critical node within the feature fusion network. SimAM simultaneously derives three-dimensional channel and spatial attention weights through energy function theory, enabling adaptive enhancement of defect features and suppression of complex background interference without introducing any learnable parameters. Experiments on a self-built aerospace engine component defect dataset demonstrate that this model achieves a significant improvement in detection accuracy with minimal computational overhead, while maintaining YOLO26's original high inference speed. This provides an excellent solution for deploying reliable and efficient visual inspection systems in resource-constrained industrial environments.

Keywords: Aerospace engine · Defect detection · YOLO26 · Attention mechanism · Lightweight model.

1 Introduction

Aero-engine integrity directly determines flight safety and operational reliability [1]. As the core components of modern propulsion systems, aero-engine blades and related structures operate under extremely harsh thermo-mechanical conditions, including high rotational speeds, elevated temperatures, and complex aerodynamic loads. Under such environments, surface degradations such as fatigue cracks, oxidation-induced ablation, and foreign object damage are likely to occur and progressively accumulate [2]. If these defects are not detected and

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_41

addressed in a timely manner, they may propagate and ultimately lead to severe structural failure. Therefore, achieving accurate and efficient inspection of aero-engine components is essential for ensuring aviation safety and reducing maintenance costs. Conventional inspection approaches, including manual visual examination and traditional non-destructive testing techniques, have been widely applied in industrial practice. Nevertheless, these methods are inherently constrained by strong reliance on human expertise, limited efficiency, and insufficient consistency, which makes them inadequate for large-scale and high-frequency inspection requirements in modern intelligent manufacturing systems [3,4]. With the rapid development of deep learning, data-driven visual inspection methods have demonstrated substantial potential by automatically learning discriminative representations from complex data distributions. In particular, one-stage object detection frameworks represented by the YOLO family have been extensively adopted in industrial scenarios due to their favorable balance between detection accuracy and inference efficiency [5].

Despite these advances, aero-engine defect detection remains a highly challenging task due to several domain-specific factors. In practical inspection environments, imaging conditions are often non-ideal. Uneven illumination, strong metallic reflections, and contamination such as oil stains or dust can significantly degrade visual quality and obscure defect characteristics. In addition, many critical defects, especially early-stage cracks and micro-scale damages, exhibit extremely small spatial scales and very low contrast against complex backgrounds, which greatly increases the difficulty of reliable detection. These challenges are further intensified by the constraints of edge deployment, where models are required to deliver real-time performance under limited computational resources. Consequently, generic detection frameworks often struggle to simultaneously achieve robustness, accuracy, and efficiency in such scenarios.

Existing research has attempted to address these issues by enhancing feature representation capability through sophisticated attention mechanisms [25,15] or by increasing network depth and capacity [6,7]. Although these strategies can improve detection performance, they inevitably introduce additional parameters and computational cost, thereby limiting their practicality in real-time industrial applications. As a result, improving detection accuracy, particularly for small and low-contrast defects, while maintaining high efficiency remains an important and unresolved problem.

Among recent real-time detection frameworks, YOLO26 provides an efficient end-to-end detection paradigm that avoids reliance on traditional post-processing operations such as Non-Maximum Suppression (NMS), thereby reducing latency and simplifying the inference process. The lightweight YOLO26-n variant is especially suitable for resource-constrained industrial inspection tasks due to its compact design and fast inference speed. Based on these considerations, YOLO26-n is selected as the baseline architecture in this study, and targeted improvements are introduced to better accommodate the characteristics of aero-engine defect inspection. Building upon this foundation, three primary contributions are summarized as follows:

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_41

1. We propose YOLO26-SimAM, a streamlined detection architecture that incorporates a parameter-free attention mechanism to enhance the discriminability of minute and low-contrast defects while preserving computational efficiency.
2. We analyze the energy-minimization principle of SimAM and investigate its integration strategy within multi-scale feature fusion, enabling effective feature recalibration without introducing additional learnable parameters.
3. Extensive experiments on a real aero-engine defect dataset demonstrate consistent improvements in mean Average Precision (mAP) with negligible inference overhead. The performance gains are particularly evident for small defects, which are further supported by qualitative heatmap visualizations.

2 Related Work

2.1 YOLO Series Object Detectors and Their Applications in Industrial Inspection

The YOLO series represents a fundamental paradigm for real-time object detection, emphasizing a unified end-to-end framework that directly predicts bounding boxes and class probabilities from input images. Early and recent versions, such as YOLOv5 and YOLOv8, adopt optimized backbone designs and effective multi-scale feature fusion strategies, enabling strong performance on large-scale benchmarks such as COCO while maintaining high inference speed. These characteristics make the YOLO family particularly suitable for industrial inspection tasks, where both accuracy and real-time responsiveness are required.

The newly proposed YOLO26 further streamlines the detection pipeline by removing the Distribution Focal Loss module and introducing a native end-to-end Non-Maximum Suppression (NMS)-free strategy, which reduces post-processing overhead and improves deployment efficiency on edge devices [27], as illustrated in Figure 1. This design simplifies the overall inference process and shortens latency, which is advantageous for time-sensitive inspection scenarios. However, despite these improvements, YOLO26 still faces limitations when dealing with weak signals and small targets embedded in complex industrial backgrounds. Such challenges are particularly prominent in aero-engine defect detection, where subtle defects are easily overwhelmed by noise and irrelevant textures. Therefore, further enhancing feature discrimination while preserving efficiency remains an important research direction.

2.2 Development of Attention Mechanisms in Visual Detection

Attention mechanisms have been widely introduced into visual detection networks to improve feature representation by selectively emphasizing informative regions. SENet [25] first introduced channel-wise attention by modeling inter-channel dependencies, significantly improving feature recalibration capability. Building upon this idea, CBAM [15] combines channel and spatial attention to further enhance representation power. Coordinate Attention [16] incorporates positional

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_41

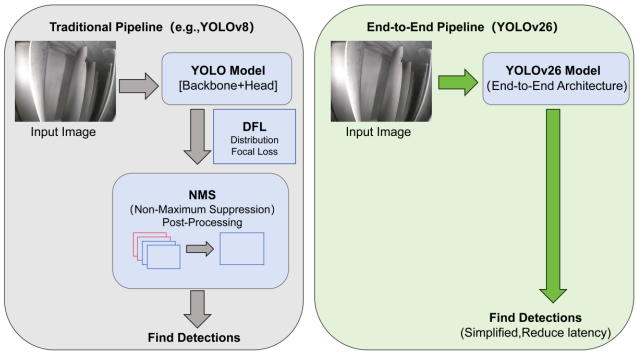

Fig. 1: Comparison of detection pipelines between traditional YOLO models and the end-to-end YOLO26 architecture.

information into channel attention with relatively low computational overhead, improving localization ability in lightweight models.

Despite their effectiveness, most of these methods rely on additional learnable parameters and auxiliary operations, which increase model complexity and may introduce overfitting risks, especially in small-sample industrial datasets. Moreover, the added computational burden can limit their applicability in resource-constrained environments. To address these issues, parameter-free attention mechanisms have been proposed. Among them, the Simple Attention Module (SimAM) [26] adopts an energy-based formulation to assign importance weights across spatial and channel dimensions without introducing extra parameters. This design maintains computational efficiency while still providing fine-grained feature modulation, making it well suited for practical visual inspection tasks where efficiency constraints are strict [13].

2.3 Challenges and Current Status of Aero-Engine Defect Detection

Applying object detection techniques to aero-engine surface inspection remains challenging due to several inherent factors, including the presence of small-scale defects, complex and noisy backgrounds, and significant intra-class variation. Existing approaches attempt to mitigate these issues through strategies such as image preprocessing to enhance defect visibility [11] and the incorporation of semantic prior-aware modules to guide feature learning [3]. While these methods can improve detection performance to some extent, they often introduce additional computational complexity or require carefully designed components, which may limit real-time performance and lightweight deployment in practical aviation maintenance scenarios [12,14].

In view of the above challenges, achieving an effective balance between detection accuracy and computational efficiency remains a key problem. To this

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_41

end, this work integrates the efficient YOLO26 framework with the parameter-free SimAM attention mechanism, aiming to enhance feature discrimination for weak and small defects while maintaining the lightweight characteristics of the model. This combination provides a practical solution for improving the precision-efficiency trade-off in real-world aero-engine inspection tasks.

3 Methodology

3.1 Overall Framework of YOLO26-SimAM

The proposed YOLO26-SimAM model is developed upon the YOLO26-n baseline, inheriting its advantages in efficiency, accuracy, and suitability for edge deployment [28]. As illustrated in Figure 2, the overall framework follows a standard three-stage design, consisting of a Backbone, a Neck, and a Head.

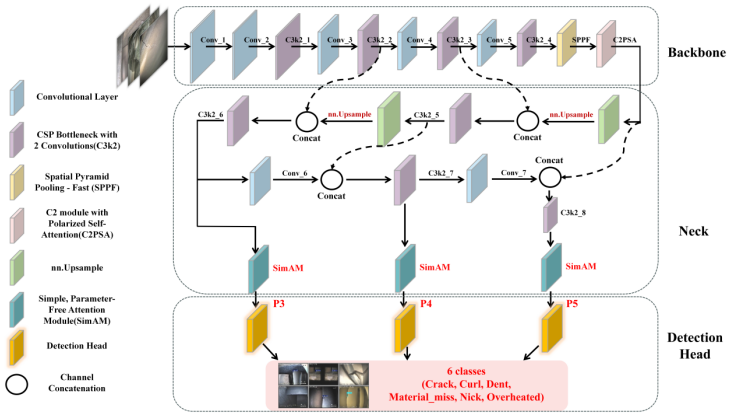

Fig. 2: The overall architecture of the YOLO26-SimAM model.

The Backbone resizes the input to 640×640 and extracts hierarchical features through convolutional layers, C3k2, Conv_3, and CSP Bottleneck modules, which help balance feature reuse and computational cost. An SPPF module is further introduced to aggregate multi-scale contextual information and enlarge the receptive field [10]. This enables the model to capture both local details and higher-level semantics that are important for defect recognition.

The Neck constructs a bidirectional feature pyramid to fuse multi-level features. Through upsampling and skip connections, features from different stages are effectively integrated to produce three feature maps, P3, P4, and P5, corresponding to different resolutions. To enhance feature quality, a SimAM module is inserted before each output. It adaptively reweights feature responses based on energy distribution, strengthening informative regions while suppressing background interference, without introducing additional parameters.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_41

The Head retains the decoupled design of YOLO26 for efficient multi-scale prediction [28]. The P3, P4, and P5 feature maps are processed by independent branches for classification and regression, enabling the model to better handle defects at different scales. This design maintains the efficiency of the baseline while improving feature discrimination.

3.2 SimAM Parameter-Free Attention Mechanism

In aero-engine blade inspection, micro-defects often exhibit weak signals and are easily affected by complex backgrounds such as texture, oil contamination, and illumination variation. Attention mechanisms can improve feature discriminability, but commonly used modules such as SE and CBAM introduce additional parameters and computational cost, which is not ideal for edge deployment.

To address this issue, SimAM is adopted as a parameter-free attention mechanism. It evaluates the importance of each neuron by measuring the energy difference between the neuron and its surrounding context, and assigns corresponding weights. In this way, more informative features are emphasized while less relevant responses are suppressed. Compared with conventional attention modules, SimAM introduces no learnable parameters and can be easily integrated into existing networks.

By providing lightweight feature recalibration, SimAM enhances the representation of subtle defects while preserving the overall efficiency of the model. Given a feature tensor:

$$\mathbf{X} \in \mathbb{R}^{C \times H \times W} \tag{1}$$

the energy e_t^* of one target neuron t on channel c can be efficiently computed via the following closed-form solution:

$$e_t^* = \frac{4(\hat{\sigma}^2 + \lambda)}{(t - \hat{\mu})^2 + 2\hat{\sigma}^2 + 2\lambda} \tag{2}$$

In this equation, $\hat{\mu}$ and $\hat{\sigma}^2$ are the mean and variance of all neurons in that channel respectively, and λ is a hyperparameter for numerical stability. The term $(\hat{\sigma}^2 + \lambda)$ reflects the overall variability of features in that channel, while $(t - \hat{\mu})^2$ measures the deviation of the target neuron relative to the channel's global context. Therefore, a smaller e_t^* indicates a greater difference between the target neuron t and the global context of its channel, warranting a higher attention weight. Finally, the module obtains a three-dimensional attention weight map by calculating the reciprocal of the energy values and normalizing via a Sigmoid function, then enhances the original features through scaling:

$$\tilde{\mathbf{X}} = \text{sigmoid}\left(\frac{1}{\mathbf{E}}\right) \odot \mathbf{X} \tag{3}$$

Based on this design, the assessment of neuron importance relies entirely on the statistical properties of the features themselves, requiring no additional parameters. Consequently, its impact on the model's computational overhead and deployment complexity is minimal.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_41

3.3 Deployment Location of SimAM: Feature Fusion Optimization in the Neck Layer

Effective multi-scale fusion is critical for detecting both large ablation areas and subtle cracks. Analysis shows that after upsampling, downsampling, and concatenation in the Neck, feature maps combine semantic and spatial information from different levels. This stage is optimal for introducing attention mechanisms for feature selection and enhancement.

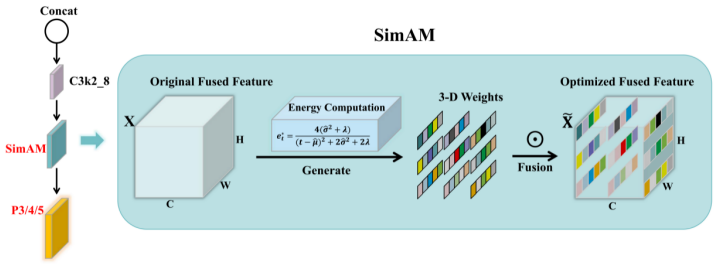

Fig. 3: Detailed workflow of the SimAM module within a fusion node.

We embed SimAM after each feature fusion node in the Neck. During feature pyramid construction, features from shallow or deep layers are aligned, concatenated, and fused. SimAM then performs feature recalibration, as detailed in Figure 3. It evaluates the importance of each element, generating implicit attention weights that are multiplied element-wise with fused features. This adaptively enhances discriminative responses related to defect edges and texture mutations while suppressing noise from metal glare, uneven illumination, and slow-varying backgrounds. This continuous purification ensures P3, P4, and P5 maps passed to the detection head are optimized and highly discriminative, forming a solid foundation for high-precision multi-scale defect detection in complex industrial scenes.

4 Experiments and Results Analysis

4.1 Experimental Setup

We use a self-collected aero-engine defect dataset consisting of high-resolution inspection images of turbine blades, stator blades, and disks. The dataset covers a variety of typical surface defects, including cracks, ablation, dents, coating spallation, and corrosion, which exhibit diverse scales and visual characteristics. To ensure a reliable evaluation, the dataset is randomly divided into training, validation, and testing sets with a ratio of 7:2:1. Performance is evaluated using standard detection metrics, including mean Average Precision (mAP), Precision,

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_41

and Recall, together with efficiency-related indicators such as parameter count, GFLOPs, and Frames Per Second (FPS), providing a comprehensive assessment of both accuracy and computational cost.

All experiments are implemented using PyTorch 1.12 and the Ultralytics framework. The input resolution is set to 640×640, and the model is initialized with pretrained weights from YOLO26-n to accelerate convergence. The optimizer is stochastic gradient descent (SGD) with a momentum of 0.937 and a weight decay of 0.0005. The initial learning rate is set to 0.01 and scheduled using cosine annealing to ensure stable training. The model is trained for 300 epochs with a batch size of 32. Data augmentation strategies include Mosaic augmentation, brightness and contrast adjustment, and simulated oil contamination noise, which help improve robustness under realistic industrial conditions. All experiments are conducted on an NVIDIA RTX 4090 GPU.

4.2 Comparative Analysis with the Baseline Model

Table 1 presents a quantitative comparison between YOLO26-SimAM and the baseline YOLO26-n. The proposed model achieves a clear improvement in detection performance, with mAP50 increasing from 0.701 to 0.792 and mAP[50:95] improving from 0.439 to 0.484. Meanwhile, Recall shows a slight increase from 0.624 to 0.630, indicating a stable ability to capture positive samples. Precision improves more significantly, rising from 0.706 to 0.81, which suggests that the proposed method effectively reduces false positives and enhances prediction reliability.

In terms of efficiency, the number of parameters and GFLOPs remain unchanged at 2.506 MB and 3.0392 GFLOPs, respectively, demonstrating that the integration of SimAM does not increase model complexity. The inference speed shows a slight decrease, with FPS dropping from 170.79 to 158.07. This reduction is mainly attributed to the additional element-wise operations introduced by the SimAM module during feature recalibration in the Neck stage. Although SimAM does not involve learnable parameters, the computation of energy-based weights and subsequent feature scaling still introduce minor overhead. Overall, the results indicate that the proposed method achieves a favorable trade-off, delivering noticeable accuracy gains with only a marginal impact on inference speed.

Table 1: Performance comparison between YOLO26-SimAM and the baseline model YOLO26-n

Model	**mAP$_{50}$**	**mAP$_{50:95}$**	**Recall**	**Precision**	**Params(MB)**	**GFLOPs**	**FPS**
Baseline	0.701	0.439	0.624	0.706	2.5061	3.0392	170.79
Ours	**0.792**	**0.484**	**0.63**	**0.81**	2.5061	3.0392	158.07

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_41

4.3 Comparative Analysis with Mainstream Lightweight Detection Models

A comparative analysis was conducted between YOLO26-SimAM and current mainstream lightweight detectors, including YOLOv8-n, YOLOv10-n and RT-DETR-L. Results in Table 2 demonstrate that YOLO26-SimAM achieves superior accuracy with the highest mAP50 of 0.792 and a competitive mAP50:95 of 0.484, outperforming alternative models such as Gold-YOLO and TOOD while surpassing DETR by a significant margin.

Table 2: Comparison of different detectors on the aero-engine defect test set

Methods	mAP_{50}	$mAP_{50:95}$	Recall	Precision	Params(MB)	GFLOPs	FPS
TOOD[18]	0.729	0.485	0.701	0.695	32.03	172.1	31.2
Retinanet[19]	0.577	0.256	0.491	0.411	21.41	163.84	13.8
YOLOv6[29]	0.643	0.365	0.567	0.724	4.23	11.80	340
Faster-RCNN[22]	0.685	0.301	0.689	0.429	41.75	182.3	38.6
DynamicRCNN[20]	0.698	0.349	0.615	0.406	41.75	182.3	36.5
DETR[21]	0.737	0.464	0.734	0.672	41.56	81.63	49.0
YOLOv5[30]	0.761	0.386	0.671	0.847	2.5	7.1	444
YOLOF[23]	0.765	0.431	0.650	0.656	42.46	83.36	59.9
Gold-YOLO[24]	0.789	0.448	0.699	0.847	5.98	10.2	444.6
Ours	**0.792**	**0.484**	**0.63**	**0.81**	**2.5061**	**3.0392**	**158.07**

The model maintains high efficiency with minimal parameter usage and computational cost, requiring only 2.506 MB parameters and 3.039 GFLOPs. This represents a substantial reduction compared to models such as TOOD and Faster-RCNN. While inference speed is lower than some highly optimized variants, the achieved frame rate remains fully adequate for industrial real-time inspection. YOLO26-SimAM thus offers an effective balance of accuracy, compactness and efficiency suitable for deployment in resource-limited aero-engine inspection environments.

4.4 Visualization Results and Analysis

Figure 4 shows qualitative comparisons between YOLO26-n and YOLO26-SimAM. The proposed model demonstrates stronger detection confidence and improved localization, particularly for small defects. In challenging scenarios with low contrast or background interference, YOLO26-SimAM produces more concentrated and accurate bounding boxes, while reducing missed detections and false positives. In contrast, the baseline YOLO26-n occasionally exhibits weaker responses to subtle defect regions. These observations are consistent with the quantitative results and further illustrate the effectiveness of the proposed attention mechanism in enhancing feature discrimination.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_41

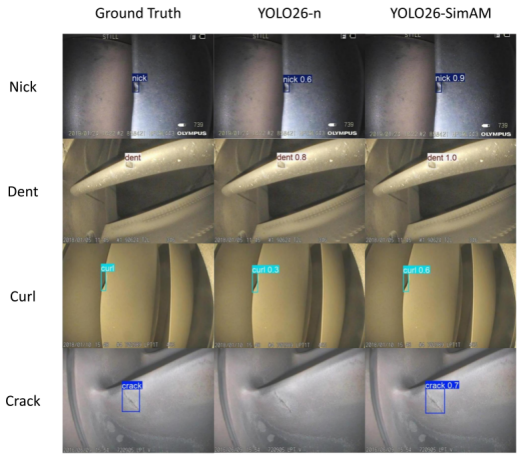

Fig. 4: Qualitative comparison of detection results for YOLO26-n and YOLO26-SimAM.

5 Conclusion

We present YOLO26-SimAM for aero-engine surface defect inspection. By integrating parameter-free SimAM attention into YOLO26-n's feature fusion network, we balance detection accuracy and computational efficiency. Experiments on the aero-engine defect dataset show significant mAP improvements over the baseline with no added parameters or computation. The model outperforms other lightweight detectors, offering high accuracy, compact size, and low complexity. Visual results confirm enhanced detection of small, low-contrast defects. The framework provides a practical solution for reliable real-time visual inspection in resource-limited industrial settings.

Limitations include a slight inference speed reduction versus the baseline, though throughput remains sufficient for real-time use. In addition, future research will evaluate the generalization capability of the proposed model on other public industrial defect datasets such as NEU-DET or DAGM to further validate its robustness.

The proposed method demonstrates strong potential for practical deployment in industrial inspection systems where both efficiency and robustness are critical. Future work will further explore adaptive attention placement strategies and cross-dataset generalization to enhance model robustness.

Acknowledgments

This work was supported in part by the XMU Training Program of Innovation for Undergraduates under Grant APSR-202517, the Natural Science Foundation

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_41

of China under Grant T2541027, the Natural Science Foundation for Distinguished Young Scholars of the Fujian Province under Grant 2023J06010, and the Independent Innovation Foundation of AECC under Grant ZZCX-2023-005.

Disclosure of Interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Xiao, Y., Shao, H., Feng, M., Han, T., Wan, J., Liu, B.: Towards trustworthy rotating machinery fault diagnosis via attention uncertainty in transformer. Journal of Manufacturing Systems **70**, 186–201 (2023)
2. Yang, P., Yue, W., Li, J., Bin, G., Li, C.: Review of damage mechanism and protection of aero-engine blades based on impact properties. Engineering Failure Analysis **140**, 106570 (2022)
3. Wu, P., Li, H., Luo, X., Hu, L., Yang, R., Zeng, N.: From data analysis to intelligent maintenance: a survey on visual defect detection in aero-engines. Measurement Science and Technology **36**, 062001 (2025)
4. Abdulrahman, Y., Eltoum, M.A.M., Ayyad, A., Moyo, B., Zweiri, Y.: Aero-engine blade defect detection: A systematic review of deep learning models. IEEE Access **11**, 53048–53061 (2023)
5. Hui, Y., Wang, J., Li, B.: WSA-YOLO: Weak-supervised and adaptive object detection in the low-light environment for YOLOV7. IEEE Transactions on Instrumentation and Measurement **73**, 1–12 (2024)
6. Wang, Y., Wang, H., Xin, Z.: Efficient detection model of steel strip surface defects based on YOLO-V7. IEEE Access **10**, 133936–133944 (2022)
7. Shang, H., Wu, J., Sun, C., Liu, J., Chen, X., Yan, R.: Global prior transformer network in intelligent borescope inspection for surface damage detection of aeroengine blade. IEEE Transactions on Industrial Informatics **19**, 8865–8877 (2023)
8. Li, X., Liu, M., Ling, Q.: Pixel-wise gamma correction mapping for low-light image enhancement. IEEE Transactions on Circuits and Systems for Video Technology **34**, 681–694 (2024)
9. Zhang, Y., Liu, X., Wang, D.: Semantic prior-aware network for pixel-level defect detection in complex industrial surfaces. IEEE Transactions on Industrial Informatics **18**, 6123–6132 (2022)
10. Jin, H., Ouyang, A., Wang, Q., Yang, D., Gan, X., Yue, X.: Helipad target detection method for low-altitude rotor UAVs based on improved YOLOv11 network model. Journal of Wireless Communications and Networking **2025**, 72 (2025)
11. Chen, H., Wu, P., Wen, W., Zeng, N.: DLA-Net: A dynamically learnable attention network for intelligent surface visual inspection of aero-engine blades. IEEE Transactions on Instrumentation and Measurement **74**, 1–14 (2025)
12. Chen, T., Zhang, C., Jing, W., Foo, E.Y.S., Lai, X., Zeng, N.: State of health estimation for lithium-ion batteries using separable LogSparse self-attention transformer. IEEE Transactions on Instrumentation and Measurement **74**, 1–13 (2025)

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_41

13. Tan, W., Zhang, H., Wang, Y., Wen, W., Chen, L., Li, H., Gao, X., Zeng, N.: SEDA-EEG: A semi-supervised emotion recognition network with domain adaptation for cross-subject EEG analysis. Neurocomputing **622**, 129315 (2025)
14. Yu, K., Tan, W., Ge, J., Li, X., Wang, Y., Huang, J., Chen, X., Li, S., Zeng, N.: M-GENE: Multiview genes expression network ensemble for bone metabolism-related gene classification. Neurocomputing **622**, 129318 (2025)
15. Woo, S., Park, J., Lee, J., Kweon, I.S.: CBAM: Convolutional block attention module. In Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) ECCV 2018. LNCS, vol. 11211, pp. 3–19. Springer, Cham (2018).
16. Hou, Q., Zhou, D., Feng, J.: Coordinate attention for efficient mobile network design. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 13713–13722 (2021)
17. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Guo, B.: Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 10012–10022 (2021)
18. Feng, C., Zhong, Y., Huang, W., Li, Y., Chen, Z., Li, X.: TOOD: Task-aligned one-stage object detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 3490–3499 (2021)
19. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 2980–2988 (2017)
20. Zhang, H., Wang, Y., Dayoub, F., Sünderhauf, N.: Dynamic R-CNN: Towards high quality object detection via dynamic training. In Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (eds.) ECCV 2020. LNCS, vol. 12346, pp. 260–275. Springer, Cham (2020).
21. Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., Zagoruyko, S.: End-to-end object detection with transformers. In Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (eds.) ECCV 2020. LNCS, vol. 12346, pp. 213–229. Springer, Cham (2020).
22. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis and Machine Intelligence **39**, 1137–1149 (2017)
23. Chen, Q., Wang, Y., Yang, T., Zhang, X.: You only look one-level feature. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 13039–13048 (2021)
24. Wang, L., Xu, Y., Wang, Y., Yang, S., Zhang, Z., Xie, W.: Gold-YOLO: Efficient object detector via gather-and-distribute mechanism. In Advances in Neural Information Processing Systems (NeurIPS), pp. 51094–51112 (2023)
25. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 7132–7141 (2018)
26. Yang, L., Zhang, R.Y., Li, L., Xie, X.: SimAM: A simple, parameter-free attention module for convolutional neural networks. In Proceedings of the 38th International Conference on Machine Learning (ICML), pp. 11863–11874 (2021)
27. Wang, C.Y., et al.: End-to-End Object Detection without NMS. arXiv:2601.12882 (2025)
28. Sapkota, R., Cheppally, R.H., Sharda, A., Karkee, M.: YOLO26: Key architectural enhancements and performance benchmarking for real-time object detection. arXiv:2509.25164 (2025)
29. Li, C., Li, L., Jiang, H., et al.: YOLOv6: A single-stage object detection framework for industrial applications. arXiv:2209.02976 (2022)

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_41

30. Ultralytics: YOLOv5. https://github.com/ultralytics/yolov5, last accessed 2026/03/30

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_41