algorithms

MDPI

Article

GAD-YOLO: A Sight-Distance Adaptive Detection Algorithm for General Aviation Aircraft Skin Damage

Tao Wu [1], Jifei Zhong [2,*], Zhanhai Wang [1], Chen Chen [1] and Zhenghong Xia [2]

[1] China Academy of Civil Aviation Science and Technology, Beijing 100083, China; wutao@mail.castc.org.cn (T.W.); wangzhh@mail.castc.org.cn (Z.W.); chenchen@mail.castc.org.cn (C.C.)
[2] College of Air Traffic Management, Civil Aviation Flight University of China, Guanghan 618307, China; xcj3232@cafuc.edu.cn
* Correspondence: zhongjifei@cafuc.edu.cn

Abstract

To address the challenges in detecting surface damage on general aviation aircraft skin—such as feature degradation under varying imaging distances, significant target scale variations, and low recognition accuracy—this paper proposes GAD-YOLO, a sight-distance adaptive detection algorithm. First, a P2 small-target detection layer is integrated into the shallow network to enhance the capture of fine damage details. Second, an HMFHead detection head is introduced to mitigate scale variation effects through progressive semantic fusion and edge-aware constraints. Third, an LDown downsampling module is designed to construct a multi-scale feature fusion architecture. This module reduces redundancy via cross-level interaction and a lightweight kernel design, thereby decreasing the number of parameters and computational cost. Additionally, a DySample-based dynamic sampling operator is proposed to preserve local details through proximity-aware sampling while enriching the contextual semantics of distant damage features, effectively improving recognition performance. Experiments on a self-constructed general aviation aircraft skin damage dataset show that GAD-YOLO achieves 87.4% precision, 80.4% recall, 86.6% mAP@0.5, and 59.7% mAP@0.5:0.95. These results outperform the YOLOv11n baseline by 2.0%, 9.4%, 6.7%, and 7.6%, respectively. The proposed method significantly improves detection performance and provides a valuable reference for intelligent inspection and maintenance in general aviation.

Keywords: aviation safety; general aviation; aircraft skin damage; object detection; YOLO

Check for updates

Academic Editors: Frank Werner and Arslan Munir

Received: 5 November 2025
Revised: 22 December 2025
Accepted: 8 January 2026
Published: 10 January 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Aircraft skin, a core component for maintaining aerodynamic shape and structural integrity, directly affects operational safety when damaged. General Aviation, an essential segment of the aviation industry, plays an irreplaceable role in economic development and public service capacity. With the increasing demand for frequent low-altitude training and high-frequency, short-cycle takeoff and landing operations, general aviation aircraft skin is more prone to diverse damage types such as cracks, dents, detachment, scratches, and corrosion [1]. These damage characteristics differ significantly from the fatigue damage typically observed on large transport aircraft. Currently, the detection of general aviation aircraft skin damage still primarily relies on visual inspection by maintenance engineers [2]. This approach suffers from low efficiency, high labor intensity, and strong dependence on human factors, compromising detection accuracy and consistency. Although instrument-assisted

detection has partially improved capabilities, practical applications still face challenges such as feature resolution degradation under dynamically varying sight distances and interference from complex backgrounds, limiting stability and reliability. Consequently, there is an urgent need for an intelligent detection method that accounts for the unique damage characteristics of general aviation aircraft skin while effectively handling varying imaging distances, thereby improving the accuracy and robustness of damage recognition.

Existing aircraft skin damage detection technologies reveal several limitations. Non-destructive testing methods such as ultrasonic testing [3] and infrared testing [4], while capable of detecting damage, are often cumbersome, costly, and inefficient. Automated approaches, including wall-climbing robot inspection [5] and UAV-based inspection [6], exhibit inadequate adaptability and stability when dealing with complex aircraft surface environments. With the rapid development of deep learning, image-based object detection methods have become a research focus and can be broadly categorized into two-stage and single-stage detectors. Two-stage detectors, such as Faster R-CNN [7], Mask R-CNN [8], and Mask Scoring R-CNN [9], typically incur high computational complexity and resource consumption. In contrast, single-stage detectors offer advantages in efficiency, real-time performance, and multi-scale feature utilization, with representative models including SSD [10], RetinaNet [11], and the YOLO series [12].

Within the family of single-stage detectors, the YOLO series has achieved remarkable progress in detection speed and accuracy through continuous architectural innovation, multi-scale feature optimization, and lightweight network design [13]. However, existing studies predominantly target damage types common in large transport aircraft, such as fatigue cracks [14], and therefore cannot be directly transferred to general aviation scenarios. Compared with transport aircraft, general aviation aircraft operating under frequent low-altitude training and high-density takeoff and landing conditions are more susceptible to diverse damage types, including cracks, dents, scratches, and corrosion, which present distinct characteristics [15]. Furthermore, the high operational frequency of general aviation demands higher maintenance efficiency, which current models still struggle to satisfy in terms of rapid screening and real-time deployment. Additionally, numerous small structural components such as rivets [16] on general aviation aircraft skin increase detection difficulty. Under varying sight distances, target image size and resolution fluctuate substantially: close-range imaging suffers from intensified complex background interference, whereas long-range imaging leads to blurred target details, causing significant performance degradation. Therefore, achieving accurate recognition of general-aviation-specific damage patterns while overcoming the bottleneck of stable small-target detection under multi-distance conditions is a critical challenge for improving the practical utility of intelligent skin damage detection in general aviation.

To address the dual challenges of general aviation characteristics and sight-distance adaptability, this paper develops the GAD-YOLO integrated skin damage detection framework. The backbone network adopts lightweight convolutions to reduce computational overhead and meet the timeliness requirements of short maintenance cycles. The neck employs the DySample [17] distance-aware convolutional kernel mapping strategy, which achieves an adaptive trade-off between close-range detail preservation and long-range semantic enhancement via learnable upsampling. The detection head incorporates a progressive semantic fusion module and an edge constraint mechanism to improve localization accuracy for complex damage such as dents and corrosion. Furthermore, a high-resolution detection branch is embedded into the shallow network to enhance sensitivity to subtle damage.

The main contributions of this work are summarized as follows:

(1) Novel network architecture: A lightweight LDown downsampling module is designed to improve detail preservation while reducing parameters and computational cost through multi-path feature fusion and an optimized 3×3 convolutional kernel. Simultaneously, an HMFHead detection head is introduced to enhance recognition of complex damage features via progressive semantic fusion and edge-aware constraints. Furthermore, a P2 small-target detection layer is added by embedding a high-resolution branch into the shallow network, thereby strengthening detection capability for small targets and fine-grained damage.

(2) Distance-aware dynamic sampling: A DySample-based distance-aware convolutional kernel deformation mapping method is proposed. By introducing a dynamic upsampling strategy with learnable distance thresholds and modulation factors, an adaptive balance between close-range local detail preservation and long-range contextual semantic enhancement is achieved, enabling stable aircraft skin damage recognition across multiple sight distances.

(3) The GAD-YOLO model is built: Tailored for general aviation aircraft skin damage detection scenarios and addressing the unique damage characteristics caused by frequent low-altitude training and short-cycle takeoffs and landings, an adaptive and efficient detection model is designed. This model maintains high recognition performance across different sight distances, meets the rapid maintenance needs of general aviation aircraft, and strengthens localization capability for complex damage.

2. Relevant Studies

In recent years, alongside the rapid growth of the aviation industry, the safety and reliability of aircraft structures have become increasingly critical. As a core component of airframe structures, aircraft skin damage detection technology plays a pivotal role in ensuring flight safety. This section systematically reviews recent advancements of YOLO-series algorithms [18] in aircraft skin damage detection.

Within aircraft skin damage detection, YOLO-based algorithms have been widely adopted for industrial defect detection. Huang [19] proposed ASD-YOLO, which leverages deformable convolution and attention mechanisms to capture irregular defects and enhance small-defect features. However, this and most related studies mainly target scenarios such as fatigue cracks on large transport aircraft, where damage scales are relatively fixed and spatial patterns are regular. These models do not adequately consider the highly variable damage scales and strong spatial randomness characteristic of general aviation, which arise from frequent takeoffs and landings and intensive low-altitude training. Zhang [20] focused on coating defects and employed oriented bounding boxes to reduce background interference, but the improvements—primarily involving feature fusion and lightweight design—do not explicitly address severe scale fluctuations and resolution degradation under dynamic imaging distances. Li [21] enhanced fastener defect detection using additional detection layers and an adaptive feature pyramid but evaluated performance under ideal conditions without systematically analyzing the influence of small-target feature degradation on detection stability. Wang [22] incorporated multiple attention modules and refined loss functions into YOLOv8n; however, the network retains a standard cross-scale fusion architecture and lacks distance-aware adaptive sampling, limiting robustness against both close-range interference and long-range blurring. Liao [23] validated YOLOv9c in a mobile real-time visual inspection system for multi-distance scenarios, but the study emphasizes system deployment rather than dedicated structural mechanisms for sight-distance adaptation. Suvittawat [24] compared advanced models on UAV-acquired data and discussed deployment challenges. Nevertheless, the discussion still

centers on large transport aircraft, with limited emphasis on general aviation skin damage detection under multi-distance and complex operational conditions.

Despite significant advances in the accuracy and classification of aircraft skin damage detection in existing research, several bottlenecks persist. First, current datasets commonly suffer from insufficient sample size, incomplete coverage of damage types, and lack of scenario diversity, hindering comprehensive model training. Second, research on general aviation aircraft skin damage remains extremely scarce. The unique damage patterns in general aviation resulting from frequent low-altitude training and short-cycle takeoffs and landings differ substantially from those of large transport aircraft, yet existing technical solutions lack targeted adaptation. Third, no studies have systematically addressed the degradation in detection stability caused by dynamically varying imaging resolution and drastic target size fluctuations under multi-distance scenarios. Therefore, this paper proposes GAD-YOLO, a sight-distance adaptive detection algorithm for general aviation aircraft skin damage, aiming to resolve the inadequate adaptation to general-aviation-specific damage patterns and detection stability degradation under varying sight distances.

3. The Proposed GAD-YOLO Algorithm

3.1. GAD-YOLO Overall Architecture

Because general aviation operations involve frequent low-altitude training and high-frequency takeoffs and landings, their aircraft skin damage characteristics differ markedly from those of large transport aircraft. Damage on transport aircraft skin commonly exhibits concentrated regional distributions and relatively regular scale distributions. In contrast, general aviation skin damage is characterized by pronounced multi-scale feature variations, highly random spatial distribution, and significantly increased background interference in damaged areas due to the superposition of low-altitude complex environmental factors.

Therefore, GAD-YOLO is constructed based on YOLOv11n [25] and adopts a three-level modular architecture comprising a feature extraction backbone, a feature fusion neck, and a detection head. The backbone [26] is optimized to enhance multi-scale feature extraction and adapt to complex damage scale variations. The feature fusion neck [27] focuses on improving cross-scale feature association, suppressing background interference from low-altitude complex environments, and enhancing discrimination between damage and background. The detection head refines the decision process for target identification and localization, thereby increasing localization accuracy for irregular damage with random spatial distribution. Collectively, these modules address the unique challenges of general aviation aircraft skin surface damage detection. The overall architecture is shown in Figure 1.

3.2. Enhanced Small Target Detection with a P2 Feature Layer

In YOLOv11n, the multi-scale detection structure based on P3, P4, and P5 feature layers provides a balance in detecting small, medium, and large targets. However, in small-target detection tasks such as aircraft skin damage, performance remains sensitive to target size variations induced by sight distance changes and often exhibits significant instability. Specifically, at close range, targets appear larger and are relatively easy to detect. As distance increases, damaged regions shrink rapidly and become small targets. Under long-range conditions, severe target size reduction and inadequate small-target feature extraction and representation capabilities become critical bottlenecks that limit detection performance.

To strengthen small-target detection, this paper introduces a P2 small-target detection layer with a resolution of 160 × 160. This layer features a smaller receptive field and higher spatial resolution, enabling the model to capture detailed local damage information at

close range while preserving weak pixel-level features at long range. Consequently, it mitigates target size reduction and feature blurring caused by sight distance variation, thereby improving detection stability and robustness.

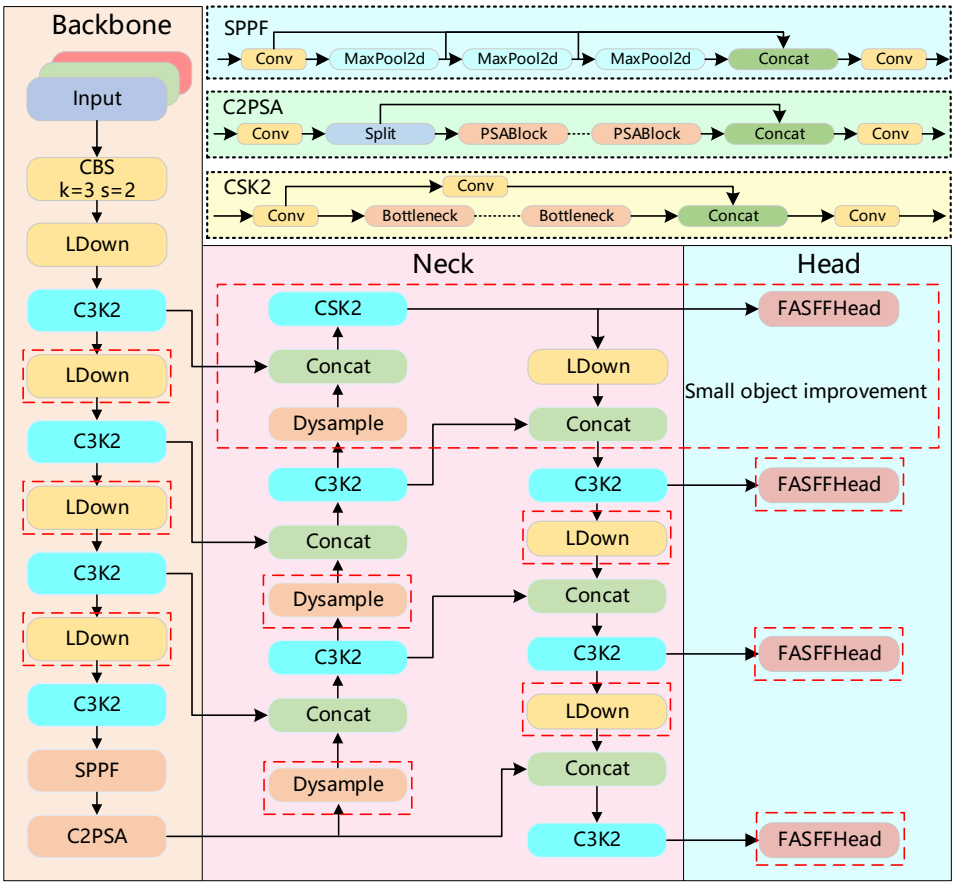


Figure 1. Network architecture of the GAD-YOLO algorithm. CBS denotes the basic convolutional block comprising convolution, batch normalization, and activation; C3K2 denotes an improved C3-based feature fusion block with dual convolutional paths; SPPF denotes the Spatial Pyramid Pooling-Fast module for multi-scale context aggregation; and C2PSA denotes a C2-based feature enhancement block integrated with an attention mechanism.

3.3. The HMFHead for Multi-Scale Feature Aggregation

To improve detection accuracy for small aircraft skin damage targets under varying sight distances, this paper designs HMFHead, a feature aggregation detection head. HMF-Head constructs hierarchical feature groups based on a decoupled structure and introduces an adaptive feature pyramid fusion mechanism to achieve efficient multi-scale information integration. During fusion, hierarchical fused feature groups HMF-1, HMF-2, and HMF-3 are generated using dynamic weight allocation and dimensional alignment strategies, enhancing the model's adaptability to targets at different scales.

Specifically, for the construction of HMF-1, the D-3 layer feature map first undergoes 3×3 max-pooling to extract key spatial information, followed by feature enhancement using a 3×3 convolution. The D-2 layer feature map is processed with a 3×3 convolution for channel alignment. For HMF-2, the D-3 layer feature map is processed by a 3×3 convolution for channel adjustment, whereas the D-1 layer feature map is first compressed via 1×1 convolution and then spatially aligned with D-3 through $2\times$ bilinear interpolation upsampling. For HMF-3, the D-2 layer feature map is compressed using a 1×1 convolution followed by $2\times$ nearest-neighbor upsampling to increase resolution, while the D-1 layer feature map undergoes $4\times$ upsampling after 1×1 convolution. This

results in consistent spatial and channel dimensions across fused features, as illustrated in Figure 2.

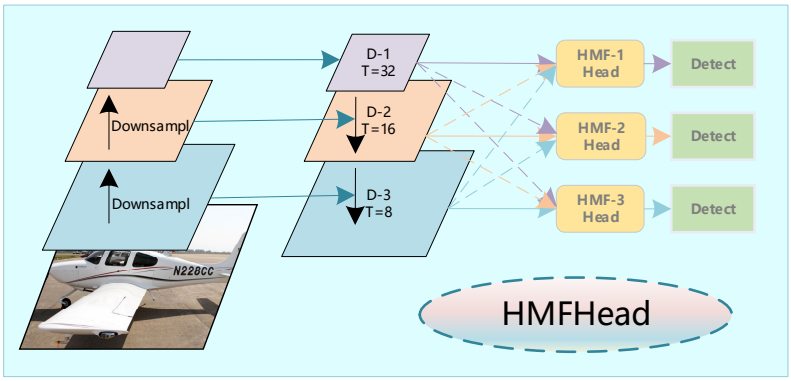


Figure 2. HMFHead architecture diagram.

In the adaptive feature pyramid fusion module, taking HMF-3 as an example, features from the D-1, D-2, and D-3 layers are denoted as f_1, f_2, and f_3, respectively. To generate the final fused feature, these features are first multiplied by learnable weight coefficients ω_3, φ_3, and ψ_3, respectively, followed by summation as shown in Equation (1).

$$f_i^l = \omega_3 \cdot f_1^{1\to l} + \phi_3 \cdot f_2^{2\to l} + \psi_3 \cdot f_3^{3\to l} \tag{1}$$

Here, ${f_1}^{1\to l}$ is the feature obtained by adjusting the spatial resolution of f_1 from the D-1 layer to that of the target layer through bilinear interpolation upsampling; ${f_2}^{2\to l}$ is the feature obtained by adjusting f_2 from the D-2 layer to the target layer through nearest-neighbor upsampling; and ${f_3}^{3\to l}$ is the feature obtained by adjusting f_3 from the D-3 layer to the target layer through 3×3 pooling followed by convolution.

The weight coefficients are generated by first bringing the feature maps from each hierarchy to the same spatial size via bicubic interpolation and then passing them through a 1×1 convolution layer to produce corresponding weight maps. After concatenation, normalization is performed using the Softmax function to constrain the weights to [0, 1] with a sum of one, as shown in Equation (2). This enables rational allocation of contributions from different feature maps and achieves distance-adaptive feature fusion.

$$\omega_i^l = \frac{e^{\eta_{\omega i}^l}}{e^{\eta_{\omega 1}^l} + e^{\eta_{\omega 2}^l} + e^{\eta_{\omega 3}^l}}, i \in \{1, 2, 3\} \tag{2}$$

In Equation (2), ω_1^l, ω_2^l, and ω_3^l correspond to the normalized weights assigned to the D-1, D-2, and D-3 layer features in the target layer l. The terms $\eta_{\omega 1}^l$, $\eta_{\omega 2}^l$, and $\eta_{\omega 3}^l$ denote the unnormalized weight indices of the original features from the D-1, D-2, and D-3 layers, respectively, representing the contribution weights of each feature layer prior to normalization.

3.4. The LDown Module for Lightweight and Efficient Downsampling

Traditional downsampling methods often cause substantial loss of detailed information when reducing feature map size, which is particularly detrimental for small-target detection. To address this issue, this paper introduces LDown, a lightweight downsampling module that replaces conventional structures to enhance detail retention and detection accuracy while reducing computational load and parameter count. LDown combines average pooling and max-pooling through multi-path feature fusion, effectively capturing

multi-scale features and strengthening the model's representation capability for complex damage patterns.

As shown in Figure 3, LDown uses a 3 × 3 convolution as its core structure and omits the 1 × 1 convolution typically found in traditional architectures. Because 3 × 3 convolutions provide a larger receptive field and capture spatially correlated features more comprehensively, whereas 1 × 1 convolutions may introduce redundant information, this design helps suppress irrelevant features, reduce parameters, and accelerate forward computation. As a result, GAD-YOLO can focus more effectively on key feature extraction.

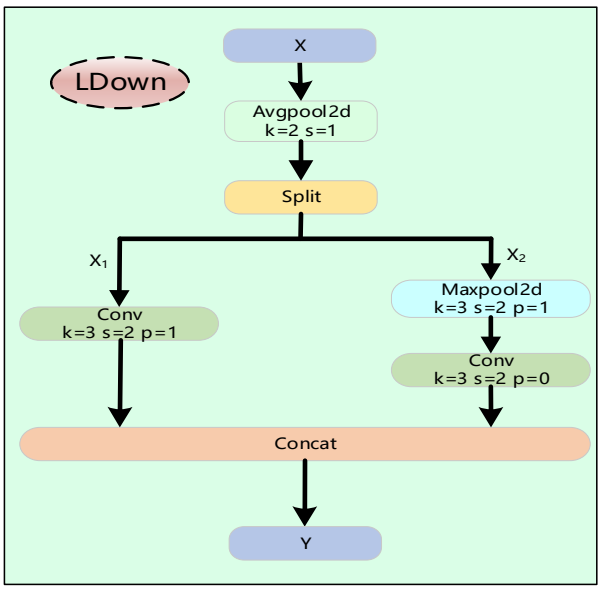


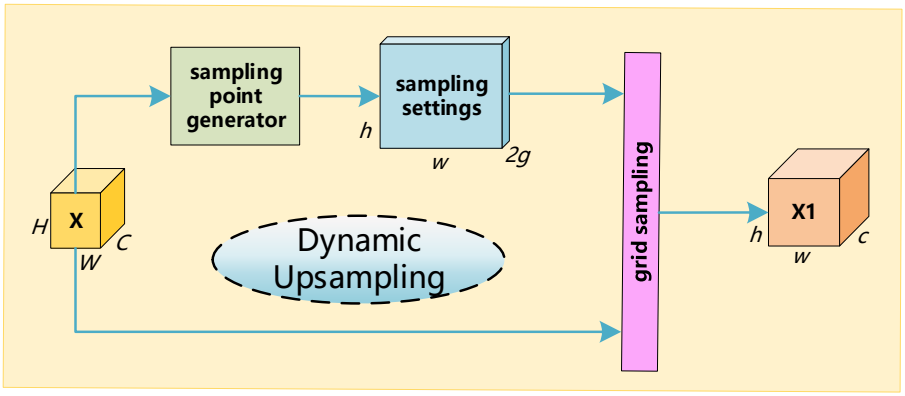
Figure 3. LDown architecture diagram.

Furthermore, LDown exhibits strong learning and adaptation capabilities. Its parameters are not fixed but are dynamically optimized during training via backpropagation according to the data distribution. When processing images with complex textures or significant variance, the convolutional kernel weights and biases adaptively adjust to capture diverse feature patterns. Thanks to this mechanism, LDown maintains stable performance across multiple tasks such as image classification and object detection, substantially improving model flexibility and robustness.

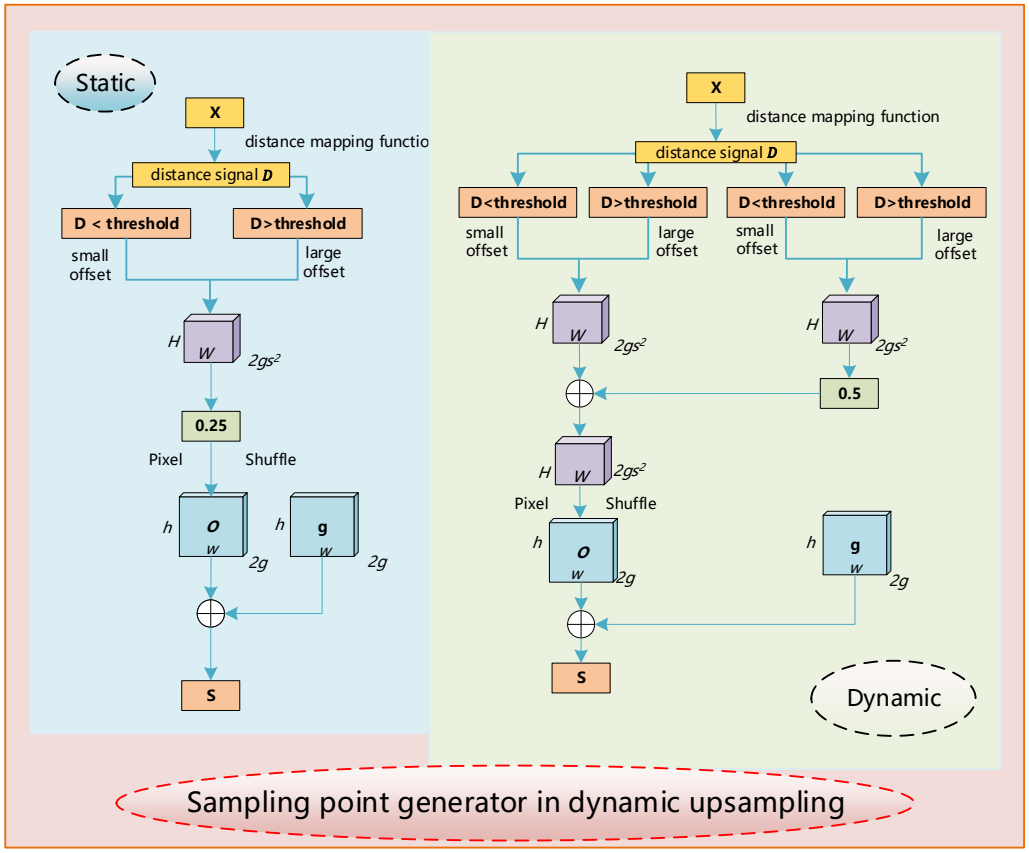
3.5. Distance-Aware Convolutional Kernel Deformation for Dynamic Sampling

Building upon the DySample dynamic upsampling strategy, this paper proposes a distance-aware convolutional kernel deformation mapping function. Based on the sight distance signal D, the function dynamically adjusts parameters such as the close-range modulation factor, the long-range modulation factor, and the response speed. The distance signal D is estimated through the ratio relationship between the camera calibration parameters and the pixel size of a reference object in the image, and the initialization value of the learnable distance threshold T is set to 1.0 m. When D is smaller than the preset threshold, the function generates micro-scale convolutional kernel offsets, causing the sampling range to highly converge within the target's local neighborhood. Through localized high-density sampling, it maximizes the preservation of high-frequency detail information in the input feature map while maintaining spatial topology relationships, thereby effectively preventing texture blurring. When D exceeds the preset threshold, the function produces large-scale spatial offsets through the convolutional kernel, breaking the local receptive field constraints of traditional convolutions and expanding the sampling range to include semantically relevant contextual regions. This enhances long-range se-

mantic information transmission, jointly constructing a multi-scale feature fusion network, as illustrated in Figure 4.



(**a**)



(**b**)

Figure 4. Structure of DySample. (**a**) Dynamic upsampling: A process involving a sampling point generator, sampling settings, and grid sampling to upsample the input X to X1. (**b**) Sampling point generator for dynamic upsampling: Comprises static and dynamic modules, utilizing a distance mapping function, threshold judgments, and Pixel Shuffle to generate sampling points for dynamic upsampling.

The original regular sampling grid ($G \in R^{2\times h\times w}$) defines the normalized coordinates of each sampling location on the input feature map to accommodate the computation of dynamic offsets in subsequent deformable convolutions, as shown in Equation (3).

$$G_{i,j} = \left(\frac{i+0.5}{h} \times 2 - 1, \frac{j+0.5}{w} \times 2 - 1 \right) \tag{3}$$

where $G_{i,j}$ denotes the coordinates of the original regular sampling grid; the input feature map is indexed by (i,j); and h and w represent the height and width of the scaled image, respectively.

The base offset prediction is defined by Equation (4).

$$\Delta p_{i,j} = \begin{pmatrix} \Delta p_{i,j}^{x} \\ \Delta p_{i,j}^{y} \end{pmatrix} \tag{4}$$

where $\Delta {p_{i,j}}^{x}$, and $\Delta {p_{i,j}}^{y}$ denote the offset components in the x and y directions, respectively.

The distance-aware convolutional kernel deformation mapping function is given by Equation (5).

$$L(D;T) = \frac{\alpha}{1+e^{-k(D-T)}} + \frac{\beta}{1+e^{k(D-T)}} \tag{5}$$

where D denotes the distance signal; T is the learnable distance threshold, initialized to 1 m and updated during training via backpropagation; α is the close-range modulation factor, initialized to 0.1, which controls the intensity of local sampling; β is the long-range modulation factor, initialized to 0.9, which controls the strength of contextual semantic enhancement; and k regulates the response rate near the threshold, initialized to 2, thereby controlling the smoothness of the transition around the threshold.

Static and dynamic factors jointly regulate the range of offset O to ensure sampling smoothness and minimize artifacts. The controlled offset O is superimposed onto the original grid G to generate the dynamic sampling grid S, as defined in Equations (6) and (7).

$$O_{i,j} = \begin{pmatrix} \Delta x_{i,j} \\ \Delta y_{i,j} \end{pmatrix} = L(D;T) \cdot \Delta p_{i,j} \tag{6}$$

$$S_{i,j} = \begin{pmatrix} S_{i,j}^{x} \\ S_{i,j}^{y} \end{pmatrix} = \begin{pmatrix} G_{i,j}^{x} + \Delta x_{i,j} \\ G_{i,j}^{y} + \Delta y_{i,j} \end{pmatrix} \tag{7}$$

where O denotes the displacement; $S_{i,j}$ represents the coordinates in the dynamic sampling grid; $G_{i,j}^{x}$ indicates the horizontal coordinate of the point at row i and column j in the original regular sampling grid; $\Delta x_{i,j}$ signifies the offset in the x-axis direction; $G_{i,j}^{y}$ denotes the vertical coordinate of the point at row i and column j in the original regular sampling grid; and $\Delta y_{i,j}$ represents the offset in the y-axis direction.

4. Experiments

4.1. Dataset Construction

The quality and scale of a dataset are crucial for training and evaluating deep neural networks. High-quality, large-scale datasets enable models to learn complex features, thereby improving recognition accuracy and generalization. Because publicly available aircraft skin surface damage datasets are scarce, a dedicated data collection campaign was conducted at multiple locations—including the Civil Aviation Flight University of China, Guanghan Airport, and the Guanghan Branch Maintenance Facility—to build a high-quality general aviation aircraft skin surface damage dataset. The dataset covers multiple aircraft types such as the Cessna 172R, Diamond DA42NG, TB-20, Citation CJ1, and Cheyenne IIIA. These aircraft represent key categories, including single-engine primary trainers, twin-engine intermediate trainers, high-performance trainers, and light business jets, and their fuselage skin materials and damage types are broadly representative of the general aviation fleet.

Data were collected using a Hikvision MV-CU013-A0GM industrial camera with a resolution of 1280 × 1024 pixels. To specifically address recognition instability caused by

varying sight distances, three sight distances (0.5 m, 1.0 m, and 1.5 m) between the camera and the aircraft skin were established, systematically capturing dimensional variations in the skin under different distances.

Data collection environments covered diverse conditions, including sunny weather, rainy weather, indoor lighting, and non-uniform illumination, as well as multiple inspection viewpoints such as frontal, oblique, and top-down views. A total of 4851 damage images were collected and manually annotated using LabelImg. Six categories of damage—cracks, dents, rivet detachment, skin detachment, scratches, and corrosion—were labeled, as shown in Figure 5a–f. In total, 16,061 annotated instances were generated, with category ratios of 28%, 15%, 17%, 18%, 12%, and 10%, respectively.

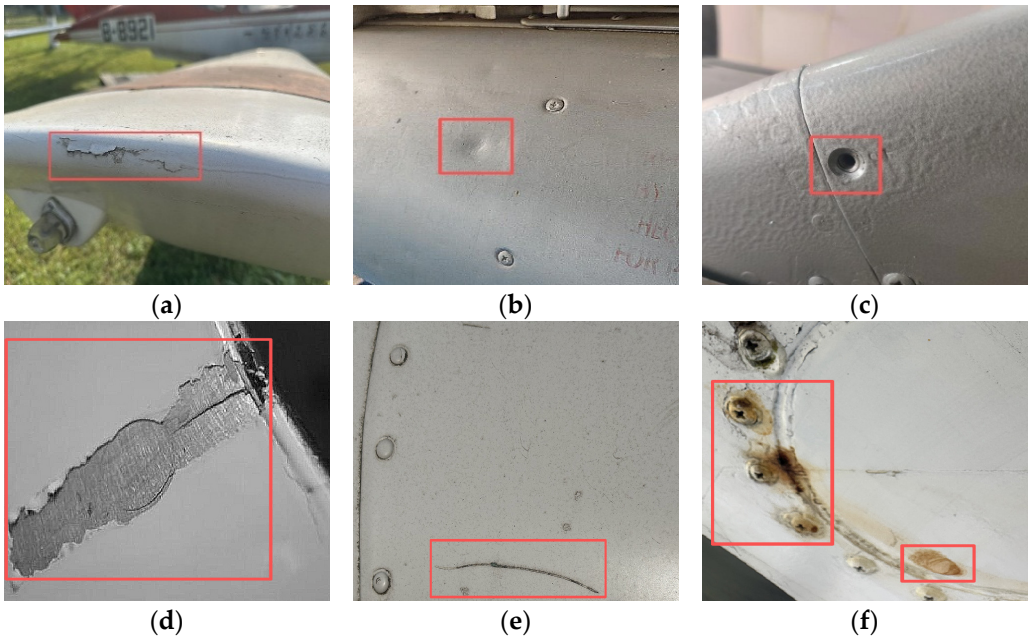


Figure 5. General aviation aircraft skin surface damage categories. (**a**) Crack; (**b**) Dent; (**c**) Rivet detachment; (**d**) Skin detachment; (**e**) Scratch; (**f**) Corrosion.

To further improve the model's generalization, mitigate overfitting, and ensure robust performance in complex real-world conditions, the initial dataset was expanded via data augmentation. Multiple augmentation operations were applied, including color jittering, random scaling, sharpening, cropping, flipping, translation, rotation, and noise injection. The original 4851 images were expanded to 34,721 images. After augmentation, the category proportions were as follows: cracks 19%, dents 18%, rivet detachment 16%, skin detachment 16%, scratches 17%, and corrosion 14%, thereby substantially increasing dataset diversity and scale. To mitigate potential overfitting introduced by augmentation, a controlled partition strategy was adopted: the augmented dataset was randomly split into training, validation, and test sets in a ratio of 7:2:1, ensuring fairness and reliability in model evaluation. This dataset provides a solid foundation for training and analyzing the proposed GAD-YOLO model for general aviation aircraft skin damage recognition.

4.2. Experimental Setup

Experiments were conducted on a 64-bit Windows 11 operating system equipped with a 32-vCPU Intel® Xeon® Platinum 8481C CPU. The GPU was an NVIDIA GeForce RTX 4090D with 24 GB GDDR6 VRAM and 80 GB system RAM. The software environment consisted of Python 3.10.12 and PyTorch 2.1.0 with CUDA 11.8. No pretrained weights were used during training. The SGD optimizer was adopted, and considering the stochastic

nature of SGD, each model configuration was trained independently with three different random seeds. The training parameters are summarized in Table 1.

Table 1. Training parameter settings.

Parameter	Setting
Input image size	640 × 640
Training epochs	50
Batch size	32
Initial learning rate	0.001
Learning rate momentum	0.937
Weight decay coefficient	0.0005
Warm-up strategy	Linear warm-up
Optimizer	SGD

4.3. Evaluation Metrics

To evaluate model performance in the object detection task, this paper employs the following metrics: precision (P), recall (R), parameter count (Params), computational load (GFLOPs), average precision (AP), and mean average precision (mAP), with calculation formulas defined by Equations (8)–(12).

$$P = \frac{TP}{TP + FP} \tag{8}$$

$$R = \frac{TP}{TP + FN} \tag{9}$$

$$AP = \int_0^1 P(R)dR \tag{10}$$

$$mAP = \frac{\sum_{m=1}^{M} AP}{M} \tag{11}$$

$$FPS = \frac{N(p)}{T(p)} \tag{12}$$

where TP denotes the number of correctly classified positive samples, specifically instances that are actually positive and predicted as positive by the model; FP represents false positives, specifically instances that are actually negative but incorrectly predicted as positive; FN signifies false negatives, specifically instances that are actually positive but erroneously classified as negative by the model; M is the total number of categories in the dataset; $N(p)$ indicates the total number of images processed by the algorithm; $T(p)$ denotes the complete processing time per image from preprocessing to result output.

4.4. Ablation Study

To quantify the contribution of each improved module to the performance of YOLOv11n in general aviation aircraft skin damage detection, ablation experiments were conducted using identical experimental configurations, progressively introducing and combining different modules. Using YOLOv11n as the baseline (Model 1), Models 2–5 individually incorporate the P2 small-target detection layer, replace the original detection head with HMFHead, introduce the LDown downsampling module, and apply the DySample-based distance-aware kernel deformation mapping for upsampling, respectively, to analyze the main effect of each module. Models 6–10 then combine multiple modules to investigate overall performance under multi-module integration. The results are reported in Table 2.

Table 2. Results of ablation experiment.

Model	P2	HMF-Head	LDown	Dysample	P/%	R/%	mAP @0.5/%	mAP @0.5:0.95/%	FPS /fps	Params /10^6	GFLOPs/M
1	×	×	×	×	85.4	71.0	79.9	52.1	167.2	2.6	6.4
2	√	×	×	×	84.3	74.8	82.3	55.6	105.9	2.9	12.5
3	×	√	×	×	83.7	73.8	80.6	53.9	117.2	4.0	8.6
4	×	×	√	×	86.5	74.6	81.9	54.7	176.1	2.1	5.3
5	×	×	×	√	87.3	72.2	81.2	54.3	157.3	2.6	6.5
6	√	√	×	×	84.7	75.9	83.1	56.1	98.3	4.1	13.4
7	√	×	√	×	85.2	74.8	83.4	57.2	110.7	3.5	9.7
8	×	√	√	×	86.8	76.2	82.8	56.4	123.4	2.8	7.1
9	√	√	√	×	87.1	78.6	85.8	58.5	119.0	3.7	12.3
10	√	√	√	√	87.4	80.4	86.6	59.7	125.1	3.6	12.3

To further evaluate the impact of data augmentation, additional comparisons were conducted for both the baseline and GAD-YOLO. Under identical architectures and hyperparameters, whether or not data augmentation and dataset expansion were used was treated as the experimental variable, and the results are summarized in Table 3. The P, R, mAP@0.5, and mAP@0.5:0.95 values are averaged over three independent training runs to reflect overall stability and variability. For intuitive comparison, detection results for general aviation aircraft skin damage using YOLOv11n and GAD-YOLO are visualized in Figure 6.

Analysis of Table 2 shows that introducing the P2 small-target detection layer increases mAP@0.5 and mAP@0.5:0.95 to 82.3% and 55.6%, representing gains of 2.4% and 3.5% over the baseline. With the addition of HMFHead, mAP@0.5 and mAP@0.5:0.95 reach 80.6% and 53.9%, i.e., 0.7% and 1.8% higher than the baseline. Incorporating LDown yields an mAP@0.5 of 81.9% and mAP@0.5:0.95 of 54.7%, corresponding to improvements of 2.0% and 2.6%. Introducing DySample leads to an mAP@0.5 of 81.2% and mAP@0.5:0.95 of 54.3%, improving by 1.3% and 2.2% relative to the baseline.

When all four modules—P2, HMFHead, LDown, and DySample—are combined, the improved model achieves P, R, mAP@0.5, and mAP@0.5:0.95 of 87.4%, 80.4%, 86.6%, and 59.7%, respectively. These results represent improvements of 6.7% and 7.6% in mAP@0.5 and mAP@0.5:0.95 compared with the baseline YOLOv11n, while maintaining a speed of 125.1 FPS, 3.6 $\times$ 10^6 parameters, and 12.3 $\times$ 10^9 GFLOPs. Thus, significant performance gains are achieved without a substantial increase in model complexity, providing an efficient and accurate solution for multi-scale detection tasks.

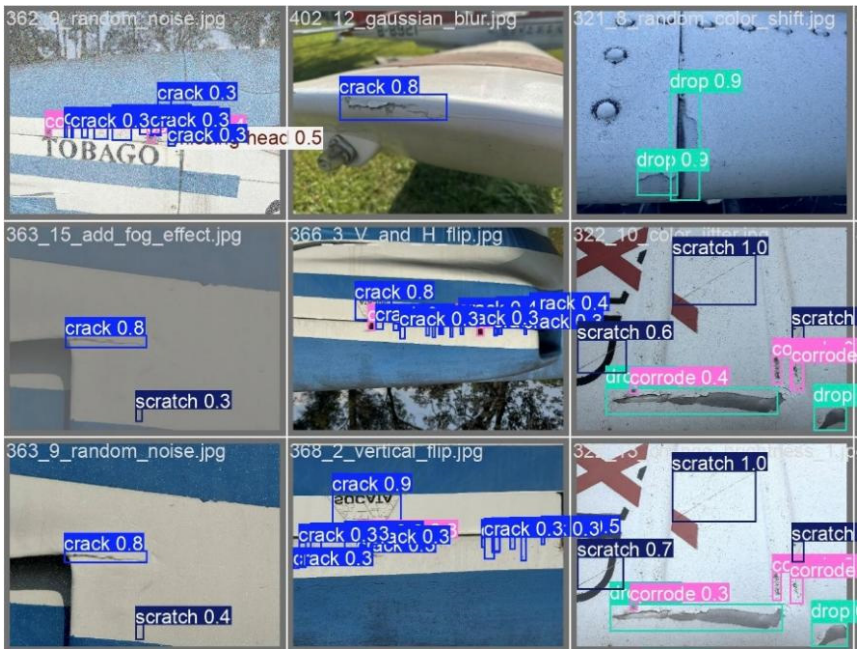
According to Table 3, when trained on the original set of 4851 images, the baseline achieves P, R, mAP@0.5, and mAP@0.5:0.95 of 72.7%, 59.2%, 63.1%, and 38.6%, whereas GAD-YOLO achieves 75.2%, 66.4%, 70.6%, and 45.1%, respectively. After data augmentation and dataset expansion to 34,721 images, the baseline's metrics increase to 85.4%, 71.0%, 79.9%, and 52.1%, while GAD-YOLO further improves to 87.4%, 80.4%, 86.6%, and 59.7%. Compared with training on the original dataset alone, augmentation yields approximate mAP@0.5:0.95 gains of 13.5% and 14.6% for the baseline and GAD-YOLO, respectively. These results indicate that data augmentation effectively increases sample diversity and feature coverage, mitigates overfitting caused by limited samples, and enables stronger generalization and robustness without increasing computational cost.

Table 3. Ablation study results on the original dataset and the augmented dataset.

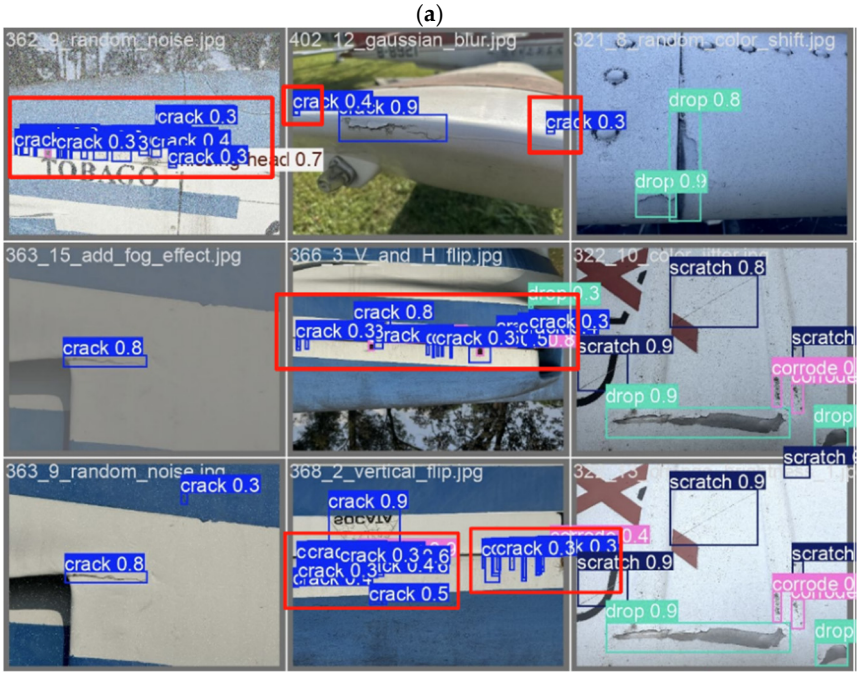
Training Data Setup	P/%	R/%	mAP @0.5/%	mAP @0.5:0.95/%	FPS /fps	Params /10^6	GFLOPs /M
Original baseline	72.7	59.2	63.1	38.6	171.7	2.6	6.4
Original improved	75.2	66.4	70.6	45.1	124.3	3.6	12.3

Table 3. *Cont.*

Training Data Setup	P/%	R/%	mAP @0.5/%	mAP @0.5:0.95/%	FPS /fps	Params /10^6	GFLOPs /M
Augmented baseline	85.4	71.0	79.9	52.1	167.2	2.6	6.4
Augmented improved	87.4	80.4	86.6	59.7	125.1	3.6	12.3



(**a**)



(**b**)

Figure 6. Visual comparison of GAD-YOLO and YOLOv11n detection results. (**a**) YOLOv11n(Baseline): Detection results of the baseline model YOLOv11n on aircraft defect images; (**b**) GAD-YOLO: Detection results of the GAD-YOLO model on aircraft defect images, showing improved detection performance.

The visualization results further demonstrate that GAD-YOLO offers remarkable advantages in detecting general aviation aircraft skin damage, especially for tiny defects that differ from those on large transport aircraft. Compared with YOLOv11n, GAD-YOLO significantly improves small-target detection and more stably captures tiny scratches,

dents, and cracks that are easily missed by traditional algorithms. Moreover, the predicted bounding boxes exhibit more accurate localization and better alignment with damage areas, along with higher confidence scores. This comprehensive improvement in detection accuracy, localization precision, and confidence validates the effectiveness and superiority of GAD-YOLO for high-precision aircraft skin damage detection, substantially increasing the detection rate and localization accuracy of small defects while reducing the risk of missed detections.

4.5. Sight-Distance Experimental Validation

To validate the effectiveness of GAD-YOLO in enhancing aircraft skin damage detection stability across varying distances, this study employs aircraft skin rivets as key visual references. Rivets, being small in size, exhibit imaging characteristics highly sensitive to sight distance, are prone to blurring, exhibit significant scale variations, and may even lose discernible features.

To quantify performance differences between YOLOv11n and GAD-YOLO, average confidence scores at 0.5 m, 1.0 m, and 1.5 m were computed, as summarized in Table 4. At 0.5 m, YOLOv11n achieves an average confidence of 0.79, whereas GAD-YOLO reaches 0.84, an improvement of approximately 5%. At 1.0 m, YOLOv11n's confidence drops to 0.67, while GAD-YOLO remains higher at 0.73, corresponding to a 6% improvement. At 1.5 m, YOLOv11n further declines to 0.51, whereas GAD-YOLO achieves 0.62, yielding an improvement of around 11%. These results show that YOLOv11n suffers substantial performance degradation as distance increases, while GAD-YOLO maintains consistently higher accuracy, with the performance gap widening at longer distances.

Table 4. Comparison of average confidence scores at three sight distances.

Detection Distance/m	Model	Average Confidence Score
0.5	YOLOv11n	0.79
	GAD-YOLO	0.84
1.0	YOLOv11n	0.67
	GAD-YOLO	0.73
1.5	YOLOv11n	0.51
	GAD-YOLO	0.62

For an intuitive assessment of sight-distance effects on stability, visualization results of both algorithms at 0.5 m, 1.0 m, and 1.5 m are presented in Figure 7, including bounding boxes, category labels, and confidence scores. At 0.5 m, both algorithms can clearly identify rivets with well-aligned bounding boxes. At 1.0 m, as targets shrink and image resolution decreases, YOLOv11n exhibits noticeable bounding-box deviations and confidence fluctuations, whereas GAD-YOLO maintains relatively high confidence and fewer localization errors. At 1.5 m, rivets become even smaller, leading to an increased missed-detection rate for YOLOv11n and substantial confidence drops, with several rivets being undetected. In contrast, GAD-YOLO continues to detect most rivets robustly with more accurate bounding boxes. Under the optimal distance of 0.5 m, we statistically evaluated detectable damage targets in the test set and determined, by segmenting minimally detectable targets and converting to image scale, that the mean size of the smallest reliably detectable damages is approximately 32×32 pixels.

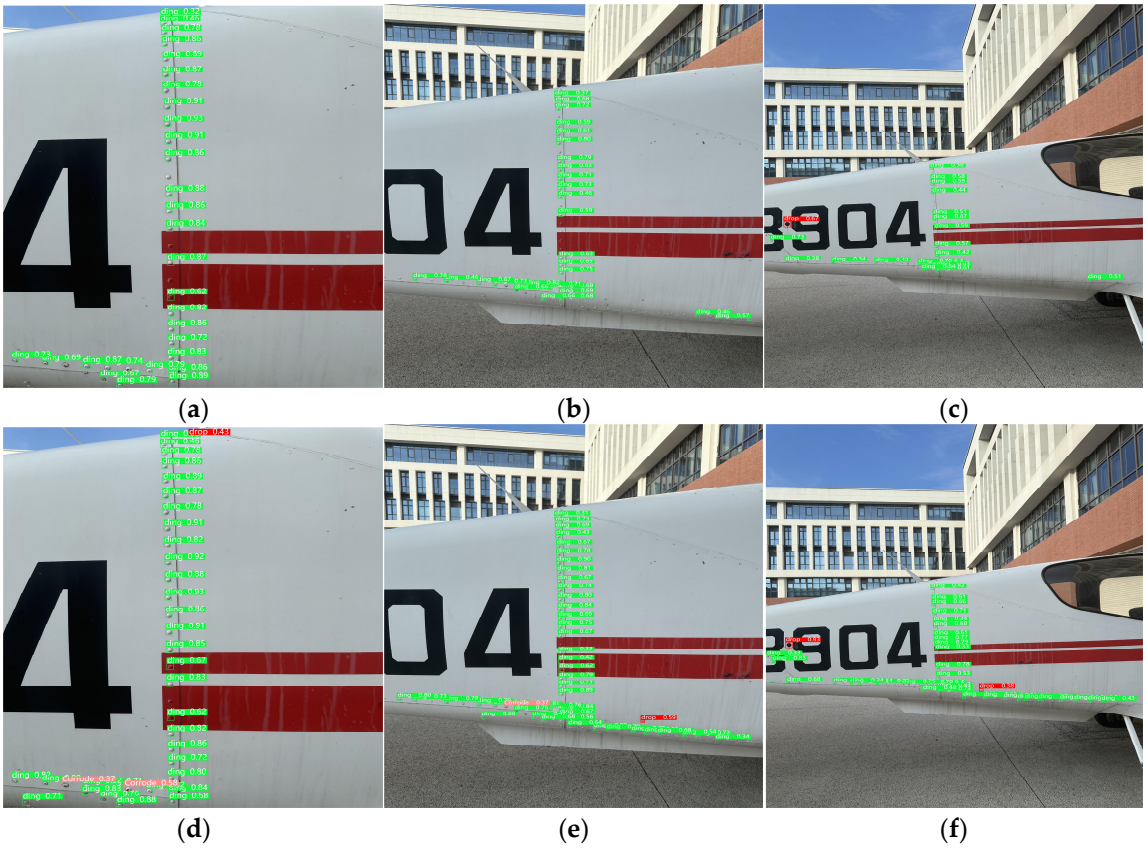


Figure 7. Comparison of detection results of aircraft rivets between original and improved algorithms with different sight distances. (**a**) Original algorithm at a distance of 0.5 m: Detection result of aircraft rivets by the original algorithm at 0.5 m distance; (**b**) 1 m original algorithm: Detection result of aircraft rivets by the original algorithm at 1 m distance; (**c**) 1.5 m original algorithm: Detection result of aircraft rivets by the original algorithm at 1.5 m distance; (**d**) 0.5 m GAD-YOLO algorithm: Detection result of aircraft rivets by the GAD-YOLO algorithm at 0.5 m distance; (**e**) 1 m GAD-YOLO algorithm: Detection result of aircraft rivets by the GAD-YOLO algorithm at 1 m distance; (**f**) 1.5 m GAD-YOLO algorithm: Detection result of aircraft rivets by the GAD-YOLO algorithm at 1.5 m distance.

As sight distance increases, rivet sizes gradually shrink and image resolution degrades, leading to a higher missed-detection rate and markedly reduced detection stability, which is consistent with theoretical expectations. Under identical conditions, however, GAD-YOLO demonstrates superior stability and accuracy, effectively alleviating the impact of inadequate resolution, maintaining higher rivet detection rates, providing more precise bounding-box localization and size estimation, and yielding higher and more stable confidence scores.

4.6. Comparative Experiment

To further validate the practical effectiveness of GAD-YOLO, multi-baseline comparative experiments were conducted against YOLOv11n, YOLOv10n, YOLOv9t, and YOLOv8n. These models represent different trade-offs in terms of parameter count and computational complexity within the YOLO family, enabling a comprehensive assessment of the proposed method's advantages in P, R, mAP@0.5, and mAP@0.5:0.95.

Table 5 reports the detailed results, while Figure 8 illustrates improvements in key metrics. GAD-YOLO achieves the best overall performance, with P and R reaching 87.4% and 80.4%, respectively, significantly exceeding other YOLO variants. The mAP@0.5 reaches 86.6%, and mAP@0.5:0.95 reaches 59.7%, both substantially higher than those of the com-

parison models. Although GAD-YOLO has a larger parameter count and computational cost than some lighter models, its performance advantages remain prominent, making it particularly suitable for scenarios requiring high precision and recall.

Table 5. Comparative experiment results.

Model	P/%	R/%	mAP@0.5/%	mAP@0.5:0.95/%	FPS/fps	Params/10^6	GFLOPs/M
YOLOv11n	85.4	71.0	79.9	52.1	167.2	2.6	6.4
YOLOv10n	82.7	69.6	77.9	50.6	138.1	2.7	8.4
YOLOv9t	82.0	65.1	73.8	46.7	161.7	1.8	6.7
YOLOv8n	85.5	70.2	78.6	51.5	143.8	2.7	6.9
GAD-YOLO	87.4	80.4	86.6	59.7	125.1	3.6	12.3

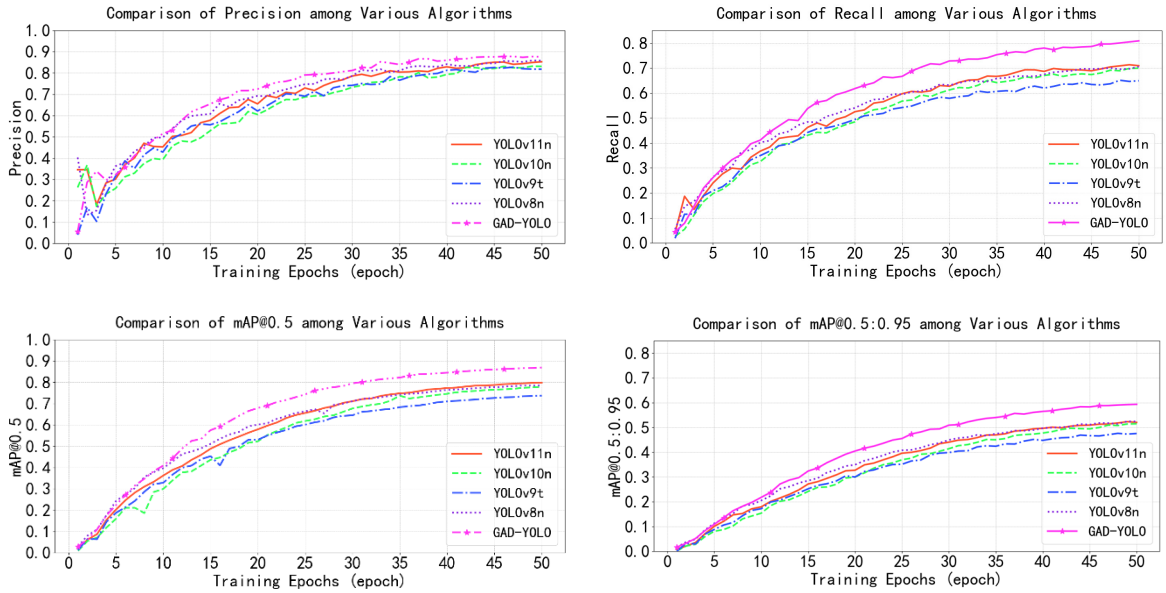


Figure 8. Comparison between GAD-YOLO and different algorithmic key evaluation metrics.

4.7. Model Generalization Evaluation

To evaluate the generalization capability of the proposed method, both YOLOv11n and GAD-YOLO were tested on the NEU Surface Defect Database [28] and the VisDrone2019-DET dataset [29]. The NEU dataset is a representative metal surface defect dataset containing 1800 grayscale images of hot-rolled steel strips, categorized into six defect types: rolled-in scale, patches, crazing, pitted surface, inclusion, and scratches. The defects are typically small with subtle texture differences and complex backgrounds, making NEU a widely used benchmark for surface defect detection. VisDrone2019-DET is a large-scale UAV-view benchmark with 6471 training images, 548 validation images, and 1610 test images, covering 10 object categories including pedestrians, bicycles, and tricycles. It contains numerous small and occluded objects, posing significant challenges for object detection.

As shown in Table 6, GAD-YOLO exhibits clear generalization ability. On the NEU dataset, YOLOv11n achieves P, R, mAP@0.5, and mAP@0.5:0.95 of 73.5%, 64.3%, 71.9%, and 42.8%, whereas GAD-YOLO reaches 75.6%, 67.8%, 74.6%, and 44.7%, respectively, indicating consistent improvements across all metrics and stronger feature representation for small-scale defects. On VisDrone2019-DET, YOLOv11n and GAD-YOLO achieve mAP@0.5:0.95 values of 23.1% and 24.9%, respectively, with GAD-YOLO improving by 1.8%. GAD-YOLO also obtains higher P, R, and mAP@0.5 values, demonstrating enhanced stability and accuracy in complex multi-object scenes.

Table 6. Generalization experiments results.

Networks	Model	P/%	R/%	mAP@0.5/%	mAP@0.5:0.95/%
NEU Surface Defect Dataset	YOLOv11n	73.5	64.3	71.9	42.8
	GAD-YOLO	75.6	67.8	74.6	44.7
VisDrone2019-DET Dataset	YOLOv11n	47.1	37.2	39.4	23.1
	GAD-YOLO	51.3	42.5	42.8	24.9

5. Conclusions

To tackle feature attenuation, severe scale fluctuations, and complex background interference caused by dynamically varying sight distances in general aviation aircraft skin damage detection, this paper proposes GAD-YOLO, a sight-distance-adaptive detection algorithm built upon YOLOv11n. The algorithm integrates a P2 small-target detection layer, an HMFHead feature aggregation head, the lightweight LDown downsampling module, and a DySample-based distance-aware convolutional kernel deformation mapping function. Experiments on a self-constructed multi-sight-distance general aviation aircraft skin damage dataset demonstrate that GAD-YOLO achieves 87.4% precision, 80.4% recall, 86.6% mAP@0.5, and 59.7% mAP@0.5:0.95, which represent improvements of 2.0%, 9.4%, 6.7%, and 7.6%, respectively, over the YOLOv11n baseline. Visualization experiments at different sight distances further confirm that GAD-YOLO provides superior detection stability and performance in multi-sight-distance scenarios.

Despite its excellent detection performance, GAD-YOLO still has several limitations. First, it provides only pixel-level localization of damage and cannot directly yield accurate physical dimensions. Second, detection accuracy may degrade under extreme weather (e.g., strong backlighting, rain, fog) or ultra-low-resolution imaging. Future work will focus on the following: (1) integrating multi-view digital image correlation and reference-object calibration to support precise quantification of physical damage dimensions; (2) incorporating multimodal data fusion to enhance robustness under extreme environmental conditions; and (3) further optimizing the lightweight model architecture and leveraging hardware acceleration to facilitate deployment on portable inspection devices and other edge terminals, thereby improving the practical applicability of the model in real-world general aviation maintenance scenarios.

Author Contributions: Conceptualization, T.W.; Methodology, J.Z.; Software, Z.X.; Validation, Z.W.; Investigation, C.C.; Resources, T.W.; Writing—original draft, J.Z.; Writing—review and editing, Z.X. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported in part by the National Key R&D Program of China (No.2024YFC3014400) and Civil Aviation Administration of China (CAAC) Safety Capability Project (ASSA2024/95).

Data Availability Statement: All data included in this study are available upon request by contacting the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Yandouzi, M.; Gaydos, S.; Guo, D.; Ghelichi, R.; Jodoin, B. Aircraft skin restoration and evaluation. *J. Therm. Spray Technol.* **2014**, *23*, 1281–1290. [CrossRef]
2. Karaoğlu, U.; Mbah, O.; Zeeshan, Q. Applications of machine learning in aircraft maintenance. *J. Eng. Manag. Syst. Eng.* **2023**, *2*, 76–95. [CrossRef]

3. Liu, Y.; Shan, Y.; Liu, P.; Li, Y. Research on Wing Skin Damage Detection and Positioning Method Based on Ultrasonic Lamb Wave. In Proceedings of the 2024 18th Symposium on Piezoelectricity, Acoustic Waves, and Device Applications (SPAWDA), Dongguan, China, 8–11 November 2024; IEEE: Piscataway, NJ, USA, 2024; pp. 400–404. [CrossRef]
4. Zhang, X.; Bai, L.; Liang, Y.; Yang, J.; Ai, J.; Ren, C. Surface Defect 3-D Profile Reconstruction Using Eddy Current Pulsed Thermography. *IEEE Trans. Ind. Inform.* **2024**, *20*, 10092–10102. [CrossRef]
5. Viraj, M.A.; Muthugala, J.; Vega-Heredia, M.; Lin, N.H.; Bhagya, S.M.; Samarakoon, P.; Elara, M.R. Adhesion risk assessment of an aircraft inspection robot for improving operator awareness. In Proceedings of the 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Kyoto, Japan, 23–27 October 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 5131–5138. [CrossRef]
6. Shao, L.; He, J.; Lu, X.; Hei, B.; Qu, J.; Liu, W. Aircraft skin damage detection and assessment from UAV images using GLCM and cloud model. *IEEE Trans. Intell. Transp. Syst.* **2023**, *25*, 3191–3200. [CrossRef]
7. Nguyen, H. Improving Faster R-CNN Framework for Fast Vehicle Detection. *Math. Probl. Eng.* **2019**, *2019*, 3808064. [CrossRef]
8. He, K.; Gkioxari, G.; Dollar, P.; Girshick, R. Mask R-CNN. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *42*, 386–397. [CrossRef] [PubMed]
9. Huang, Z.; Huang, L.; Gong, Y.; Huang, C.; Wang, X. Mask scoring r-cnn. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 6409–6418. Available online: https://ieeexplore.ieee.org/document/8953609 (accessed on 21 October 2025).
10. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. SSD: Single shot multibox detector. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; Springer International Publishing: Cham, Switzerland, 2016; pp. 21–37. [CrossRef]
11. Zhang, H.; Chang, H.; Ma, B.; Shan, S.; Chen, X. Cascade RetinaNet: Maintaining consistency for single-stage object detection. *arXiv* **2019**, arXiv:1907.06881. [CrossRef]
12. Jiang, P.; Ergu, D.; Liu, F.; Cai, Y.; Ma, B. A Review of Yolo algorithm developments. *Procedia Comput. Sci.* **2022**, *199*, 1066–1073. [CrossRef]
13. Zhou, R.; Xiang, B.; Wu, L.; Hu, Y.; Dou, L.; Huang, K. Research on Intelligent Identification Method for Pantograph Positioning and Skateboard Structural Anomalies Based on Improved YOLO v8 Algorithm. *Algorithms* **2024**, *17*, 574. [CrossRef]
14. Tavares, S.M.O.; De Castro, P.M.S.T. An overview of fatigue in aircraft structures. *Fatigue Fract. Eng. Mater. Struct.* **2017**, *40*, 1510–1529. [CrossRef]
15. Yang, X. Research of General Aviation flight data analysis. In Proceedings of the 2011 International Conference on Electronic & Mechanical Engineering and Information Technology, Harbin, China, 12–14 August 2011; IEEE: Piscataway, NJ, USA, 2011; Volume 9, pp. 4497–4499. Available online: https://ieeexplore.ieee.org/document/6023064 (accessed on 7 January 2026).
16. Sun, H.; Feng, Q.; Huang, S.; Li, J.; Peng, L.; Li, S.; Tu, J. Quantification Method for Rivet Hole Cracks in an Aircraft Fuselage Using Guided Waves: A BTBD-Theory-Hybrid Space-Time Cross Fusion SpatNet. *IEEE Trans. Ind. Inform.* **2025**, *21*, 3996–4007. [CrossRef]
17. Yu, A.; Zhang, J.; Dang, W.; Zhang, H. Learning to Upsample Smoke Images via a Deep Convolutional Network. *IEEE Access* **2019**, *7*, 138932–138940. [CrossRef]
18. Zhu, X.B. Aviation rivet classification and anomaly detection based on deep learning. *Int. J. Aerosp. Eng.* **2023**, *2023*, 3546838. [CrossRef]
19. Huang, B.; Ding, Y.; Liu, G.; Tian, G.; Wang, S. ASD-YOLO: An aircraft surface defects detection method using deformable convolution and attention mechanism. *Measurement* **2024**, *238*, 115300. [CrossRef]
20. Zhang, Y.; Wang, W.; Guo, Z.; Ji, Y. Research on oriented surface defect detection in the aircraft skin-coating process based on an attention detector. *IET Image Process.* **2024**, *18*, 1213–1228. [CrossRef]
21. Li, H.; Wang, C.; Liu, Y. YOLO-FDD: Efficient defect detection network of aircraft skin fastener. *Signal Image Video Process.* **2024**, *18*, 3197–3211. [CrossRef]
22. Wang, H.; Fu, L.; Wang, L. Detection algorithm of aircraft skin defects based on improved YOLOv8n. *Signal Image Video Process.* **2024**, *18*, 3877–3891. [CrossRef]
23. Liao, K.C.; Lau, J.; Hidayat, M. Aircraft skin damage visual testing system using lightweight devices with yolo: An automated real-time material evaluation system. *AI* **2024**, *5*, 1793–1815. [CrossRef]
24. Suvittawat, N.; Kurniawan, C.; Datephanyawat, J.; Tay, J.; Liu, Z.; Soh, D.W.; Ribeiro, N.A. Advances in Aircraft Skin Defect Detection Using Computer Vision: A Survey and Comparison of YOLOv9 and RT-DETR Performance. *Aerospace* **2025**, *12*, 356. [CrossRef]
25. Khanam, R.; Hussain, M. YOLOv11: An overview of the key architectural enhancements. *arXiv* **2024**, arXiv:2410.17725. [CrossRef]
26. Xue, G.; Li, S.; Hou, P.; Gao, S.; Tan, R. Research on lightweight Yolo coal gangue detection algorithm based on resnet18 backbone feature network. *Internet Things* **2023**, *22*, 100762. [CrossRef]

27. Chen, F.; Deng, M.; Gao, H.; Yang, X.; Zhang, D. NHD-YOLO: Improved YOLOv8 using optimized neck and head for product surface defect detection with data augmentation. *IET Image Process.* **2024**, *18*, 1915–1926. [CrossRef]
28. Aşar, E.; Özgür, A. *Systematic Review of Steel Surface Defect Detection Methods on the Open Access Datasets of Severstal and the Northeastern University (NEU). Steel 4.0: Digitalization in Steel Industry*; Springer: Cham, Switzerland, 2024; pp. 37–72. Available online: https://link.springer.com/chapter/10.1007/978-3-031-57468-9_3 (accessed on 7 January 2026).
29. Du, D.; Zhu, P.; Wen, L.; Bian, X.; Lin, H.; Hu, Q.; Zhang, L. VisDrone-DET2019: The vision meets drone object detection in image challenge results. In Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, Seoul, Republic of Korea, 27 October–2 November 2019. Available online: https://openaccess.thecvf.com/content_ICCVW_2019/html/VISDrone/Du_VisDrone-DET2019_The_Vision_Meets_Drone_Object_Detection_in_Image_Challenge_ICCVW_2019_paper.html (accessed on 7 January 2026).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.