automation

MDPI

Article

Few-Shot Surface Defect Detection in Sinusoidal Wobble Laser Welds Using StyleGAN2-AFMS Augmentation and YOLO11n-WAFE Detector

Guangkai Ma *,†, Jianwen Zhang † and Jiheng Jiang

College of Computer and Control Engineering, Northeast Forestry University, Harbin 150040, China; z1965024352@126.com (J.Z.); 13455053434@163.com (J.J.)

* Correspondence: maguangkai@nefu.edu.cn

† These authors contributed equally to this work.

Abstract

In the manufacturing of high-reliability components, sinusoidal wobble laser welding has gained preference due to its excellent performance. However, surface defect inspection for such welds is challenged by large variations in defect scales, the coexistence of multiple defects, and scarce samples, which collectively limit existing detection methods. To address these issues, this paper proposes a lightweight detection framework that integrates a generative adversarial network with an improved YOLO architecture. First, a frequency-domain-enhanced StyleGAN2-AFMS model is constructed to effectively augment high-quality defect samples. Second, a YOLO11n-WAFE detector is designed, which incorporates an ADownECA downsampling module to enhance the capability of capturing subtle defects and an Edge-Aware Semantic–Detail Fusion module to improve discriminative robustness under multi-defect conditions. To validate the approach, an industrial-level Sinusoidal Wobble Laser Weld Defect Dataset is built. Experiments reveal that the proposed framework boosts mAP@0.5 to 94.2% (an 8% improvement over the baseline) and mAP@0.5:0.95 to 77.4%, with an F1-score of 89.5%, while maintaining lightweight (2.15 M parameters) and fast (656 FPS) characteristics. This study provides a high-precision and efficient solution for few-shot industrial defect inspection.

Keywords: defect detection; sinusoidal wobble laser welding; few-shot learning; weld defect; YOLO; generative adversarial network

1. Introduction

Laser welding, as a critical process for achieving high-reliability connections in precision components within the electronics manufacturing industry, has been widely adopted in areas such as semiconductor packaging [1], precision instrument assembly, and the manufacturing of key structural components in aerospace [2]. Its welding quality directly impacts the reliability and service life of these structural parts [3]. Among various techniques, sinusoidal wobble laser welding, which produces welds with periodic surface patterns, is particularly applied in the manufacturing of high-critical components like laptop batteries, owing to its advantages in localized heat distribution and metallurgical bonding. However, in practical industrial production, due to fluctuations in process parameters, material properties, and environmental factors, surface defects such as fractures, excessive penetrations, and deposition variations may occur. These defects exhibit diverse morphologies and significant scale variations, often concealed within complex periodic

Check for updates

Academic Editor: Duc Truong Pham

Received: 17 January 2026
Revised: 14 February 2026
Accepted: 21 February 2026
Published: 26 February 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.

texture backgrounds [4]. Furthermore, multiple defect types may coexist within a single solder joint. Therefore, detection models must capture fine details of subtle defects while also recognizing large-scale defect patterns. This requirement places strict demands on the models' multi-scale feature extraction capabilities. Existing detection methods often exhibit limitations in meeting these complex demands, struggling to achieve rapid and accurate identification, thereby posing significant challenges to product quality control [5].

Traditional Automated Optical Inspection (AOI) methods, which rely on handcrafted image features and template-matching algorithms, can achieve satisfactory results in regular and relatively simple scenarios. However, their generalization capability and robustness are often limited when confronted with complex textures, multi-scale defects, and varying lighting conditions [6]. In recent years, deep learning techniques, particularly Convolutional Neural Networks (CNNs), have demonstrated significant progress in the field of visual inspection. Representative object detection models such as Faster R-CNN [7] and the YOLO series [8] have been widely applied to industrial quality inspection tasks owing to their remarkable performance.

Faster R-CNN-based detectors follow a two-stage pipeline: they first generate region proposals and then classify and refine bounding boxes. In industrial quality inspection, several studies have explored its application. For instance, Li et al. [9] introduced an improved Faster R-CNN incorporating ResNet50 and a feature pyramid structure, achieving a 2.2% mean Average Precision (mAP) improvement in aluminum surface defect detection. In the shipbuilding sector, Oh et al. [10] applied Faster R-CNN to automate the detection of welding defects in radiographic images, enhancing both recognition accuracy and efficiency through data augmentation. Similarly, Hu et al. [11] optimized the feature extraction module of Faster R-CNN for weld inspection in steel bridges, reporting a significant mAP increase of 28.3 percentage points. Nevertheless, the structural complexity of Faster R-CNN results in high computational costs, which remains a limitation in industrial inspection scenarios requiring high real-time performance.

Single-stage detectors, particularly the YOLO series, are favored in industrial defect detection due to their efficiency and balanced performance. To enhance their performance and practicality for industrial deployment, numerous enhancements to the YOLO architecture have been proposed. A common strategy centers on architectural and module-level innovations aimed at enhancing detection robustness and feature representation. For instance, Zhu et al. [12] embedded a Coordinate Attention mechanism into YOLOv5s, improving feature representation for weld detection and achieving an mAP@0.5 of 93.79%. Similarly, Yang et al. [13] incorporated larger convolutional kernels and an improved SPPSE module into YOLOv5, attaining 97% mAP in laser welding defect inspection. Liu et al. [14] further introduced MSFF-AConv and SPPELAN-SKNet modules into YOLOv9, strengthening multi-scale feature extraction and raising mAP to 56.8% for weld defects. Lightweight design is also a key consideration for industrial deployment. Wang et al. [15] developed YOLOv6-NW, a lightweight variant that reduced parameters to 16% of the original model while preserving accuracy in weld-crack detection. Xu et al. [16] integrated depthwise separable convolution and EIoU loss into YOLOv7, increasing mAP@0.5 by 11% while significantly lowering computational costs. Beyond architectural innovations, domain-specific adaptation has been achieved through customized detection heads and data strategies. Wang et al. [17] combined Mosaic and Copy–Pasting augmentations with optimized anchor boxes to raise mAP to 91.07% on transparent part defects. Çelik et al. [18] tailored YOLOv8s for defect detection in automotive plastic parts, using Pareto analysis to identify dominant defect types and performing hyperparameter optimization to achieve an mAP of 0.990, demonstrating the model's strong adaptability to industrial manufacturing environments. Cengil [19] customized YOLOv10 for three-class weld quality inspection, ef-

fectively distinguishing good, bad, and defective welds, and achieving 93.9% precision and 91.0% recall. Despite these advancements, YOLO-based models remain heavily data-dependent. Their performance can deteriorate significantly when training data is scarce, highlighting a critical limitation in real-world industrial applications.

Indeed, a common bottleneck across all deep learning-based detection methods is the severe shortage of high-quality, class-balanced training samples. In practical electronics manufacturing scenarios, defect samples are difficult to acquire in large quantities due to high production line yield and data sensitivity, leading to a critical insufficiency of high-quality samples available for research. This issue makes complex deep learning models prone to overfitting during training, hindering their ability to learn robust defect representations and consequently limiting their practical deployment on production lines [20]. To alleviate the data scarcity problem, traditional data augmentation techniques (e.g., rotation, flipping, color jittering) are widely employed [21]. However, these methods only provide limited geometric or color variation and struggle to generate defect morphologies with novel semantic information. In contrast, Generative Adversarial Networks (GANs) can achieve more substantial sample expansion at the data distribution level. Since the initial framework was proposed by Goodfellow et al. [22], GANs have seen rapid advancement in image generation and are increasingly being applied to defect sample augmentation. For example, Wang et al. [23] proposed a foreground-perception CycleGAN (FCGAN) that integrates attention mechanisms into both the generator and discriminator, enabling the synthesis of realistic pseudo-defect images and alleviating sample scarcity in surface defect detection. Huang et al. [24] developed an LT-CVAE-GAN model that combines a Conditional Variational Autoencoder and a Conditional GAN to generate samples for rare fault types, effectively rebalancing long-tailed datasets and improving defect diagnosis performance. Niu et al. [25] developed a weakly supervised surface defect segmentation framework based on an improved CycleGAN, which incorporates a defect attention module to generate defect-free templates using only image-level annotations. Xie et al. [26] proposed a High-quality Matching Transfer GAN (HMTGAN) that fuses CycleGAN and StyleGAN principles to produce realistic cross-material defect images, enhancing dataset diversity. Finally, Liang et al. [27] employed a Conditional GAN integrated with feature transformation and data optimization to address class imbalance in transformer fault diagnosis, demonstrating notable improvements in classification accuracy and data quality. It is important to note that although GAN technology can supplement the data distribution, it still faces challenges under limited sample conditions—including training instability, insufficient generation diversity, and the domain gap between synthetic and real samples.

The core scientific purpose of this work is to establish a rigorous, integrated framework that tackles the fundamental challenges of few-shot defect detection in sinusoidal wobble laser welding—a typical structured industrial manufacturing scenario. We aim to bridge the gap between data scarcity and model performance by co-designing a generative model that incorporates physical priors (periodic texture) and a lightweight detector explicitly optimized for capturing subtle, multi-scale defect features while satisfying real-time inference requirements.

Few-shot surface defect detection for sinusoidal wobble laser welds is challenged by data scarcity, complex multi-scale features, and stringent real-time requirements. To address these challenges, an end-to-end framework integrating generative data augmentation and lightweight detection is proposed. At the algorithmic level, StyleGAN2 and YOLO11n are collaboratively optimized to handle defect detection under periodic texture backgrounds. The overall architecture of the framework is illustrated in Figure 1, and the main contributions are summarized as follows:

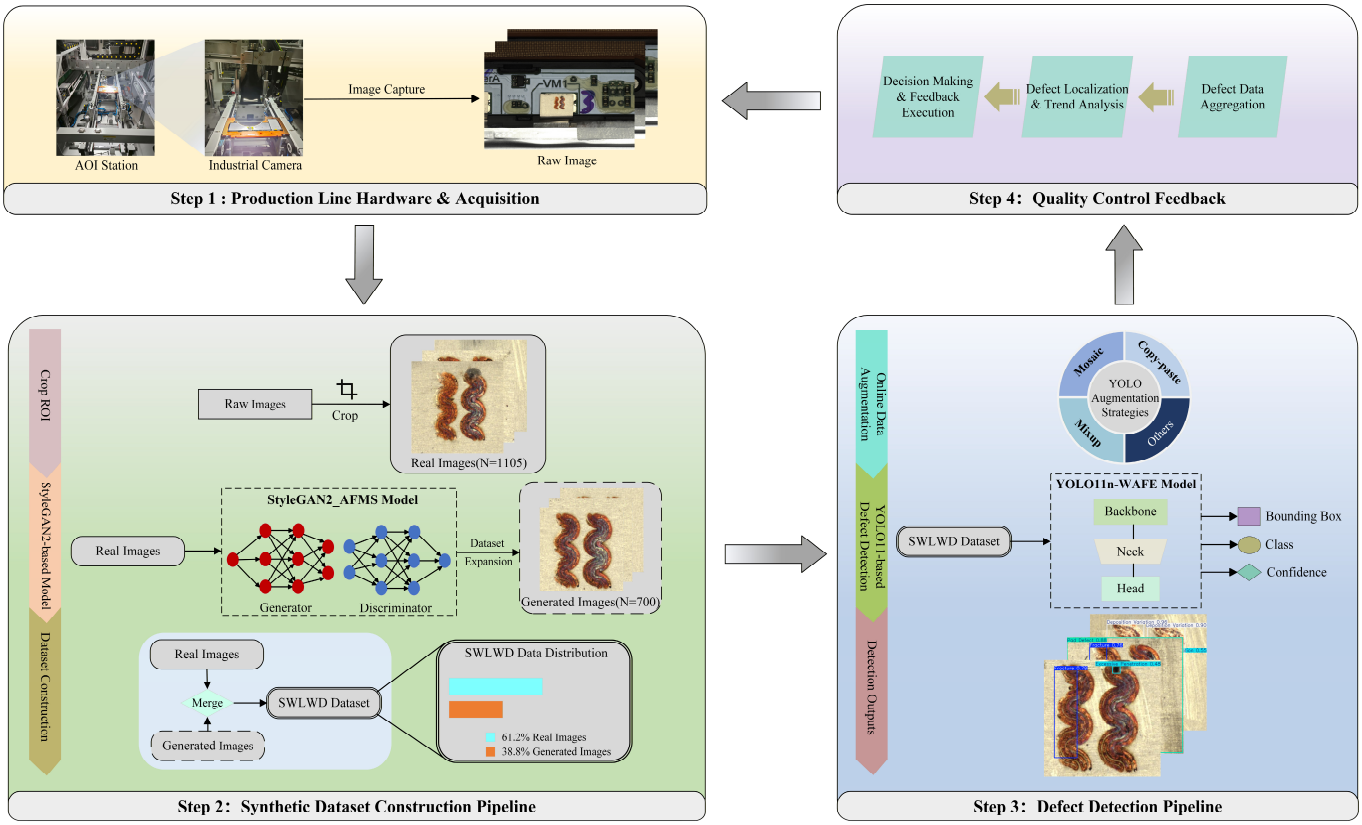


Figure 1. Overall framework for few-shot defect detection in sinusoidal wobble laser welds.

(i) To address the challenges of scarce defect samples and complex morphological variations, a frequency-domain-enhanced generative adversarial network named StyleGAN2-AFMS (Attentive Fourier-Modulation Synthesis) is proposed. This method employs a dual-stream mechanism to explicitly model the global periodic texture of weld images and local defect details: the frequency-domain stream reconstructs spatial structures via inverse Fourier transform, while the modulation stream captures non-periodic defect features using spatial convolutions. Adaptive fusion of the two types of features is then achieved through channel attention. To further enhance the model's performance under limited data, the Adaptive Discriminator Augmentation (ADA) strategy [28] is incorporated, which effectively mitigates overfitting during few-shot training and thereby boosts the realism and diversity of generated samples.

(ii) To balance detection accuracy and inference speed, a dedicated lightweight detector named YOLO11n-WAFE (Welding–ADownECA Fusion Enhancement) is designed. Its core innovations include: (1) an ADownECA downsampling module, which incorporates Efficient Channel Attention (ECA) within a multi-path structure to enhance discriminative capability for minute defect features, and (2) an Edge-Aware Semantic–Detail Fusion (EASDF) module, which employs Partial Differential Equations (PDEs) for explicit edge enhancement and combines an entropy-aware weighting strategy to optimize multi-scale feature fusion, significantly improving detection performance for multi-category defects.

(iii) An industrial-level Sinusoidal Wobble Laser Weld Defect Dataset (SWLWD Dataset) was constructed, comprising high-resolution samples of four typical defect categories. The dataset encompasses various morphologies and background combinations from real production scenarios, providing a crucial data foundation and performance benchmark for related research.

2. Materials and Methods

This section elaborates on the three core components of the proposed defect detection framework for sinusoidal wobble laser welds, namely: the construction and augmentation strategy of the dedicated dataset, the frequency-domain modeling-based image generation method, and the lightweight defect detection model. Their specific designs and implementations are as follows.

2.1. SWLWD Dataset Construction and Augmentation

(i) Original Data Acquisition and Annotation

To support the training and evaluation of defect detection models for sinusoidal wobble laser welds, the SWLWD Dataset was constructed. The original images were acquired from the production line of a battery manufacturer in Suzhou, China, using a Hikvision MV-CU120-10GC industrial camera (Hangzhou Hikvision Digital Technology Co., Ltd., Hangzhou, China) under OPT-LHDG3120-W (Guangdong OPT Co., Ltd., Dongguan, China) standard lighting conditions, yielding a total of 1105 high-resolution (4024×3036) color images. The welded material in all specimens is nickel-plated copper. To accommodate computational constraints and emphasize regions of interest, each original image was center-cropped to 512×512 pixels, ensuring that every sub-image contains complete weld features. The dataset split was performed at the original image level prior to center cropping, ensuring that all cropped patches derived from the same weld image were assigned exclusively to the same subset (training, validation, or test).

The annotation categories encompass four key defect types: Fracture, Excessive Penetration, Deposition Variation, and Pad Defect (including pad contamination and breakage). The dataset was randomly split into training, validation, and test sets. A detailed breakdown of the dataset partition is as follows: the training set contains 773 images (70%), the validation set contains 166 images (15%), and the test set contains 166 images (15%) [29]. In this work, the term "few-shot" is operationally defined at the defect-instance level in the training set, referring to the limited number of annotated defect instances per category rather than merely the number of images or weld samples. Following the dataset partition, all StyleGAN2-AFMS generated samples were added exclusively to the training set, while the validation and test sets remained unchanged for held-out evaluation. The selected 700 synthetic images were manually annotated by professional annotators using the LabelImg tool (version 1.8.6), following the same annotation guidelines as those applied to the real images, with secondary review for ambiguous cases.

To provide an intuitive representation of the defect characteristics within the SWLWD Dataset, Figure 2 shows sample examples of the four defect categories, while Table 1 summarizes the feature descriptions and detection challenges for each class. These samples not only reflect the complexity and diversity of weld defects in real production environments but also establish a reliable foundation for the subsequent training and evaluation of defect detection models.

(ii) Data Augmentation via StyleGAN2-AFMS

Because defect samples are extremely scarce and weld images often contain multiple coexisting defects, standard generative models struggle. Specifically, they have difficulty preserving global texture consistency while accurately reproducing fine defect details. To address this, the proposed StyleGAN2-AFMS generative model is employed for data augmentation. It collaboratively models the aforementioned two types of features through a frequency-space dual-stream mechanism, thereby generating higher-quality training samples.

The StyleGAN2-AFMS model was trained for 500k iterations on the original training set. Subsequently, the trained generator synthesized 2000 images, encompassing samples

with single defects as well as natural combinations of multiple defects. To ensure the quality and realism of the generated data, a rigorous manual and automatic screening pipeline was implemented. This pipeline leveraged Fréchet Inception Distance (FID) [30] and precision–recall metrics [31] to filter out low-quality and unrealistic samples.

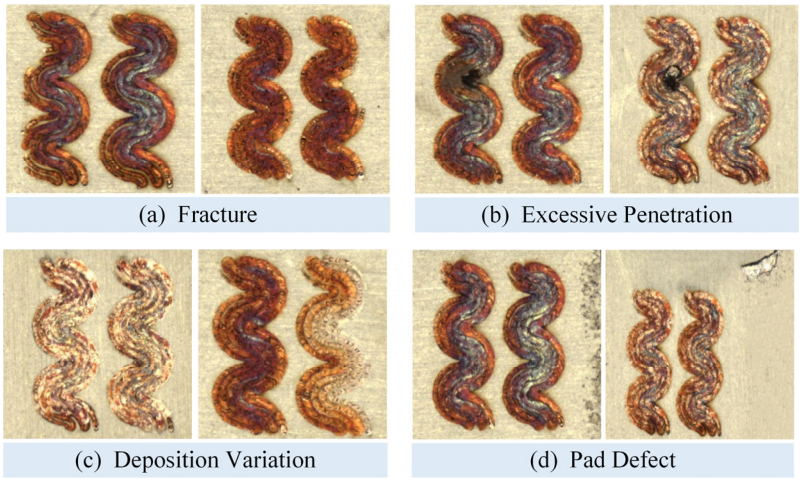


Figure 2. Original samples of sinusoidal wobble laser weld defects.

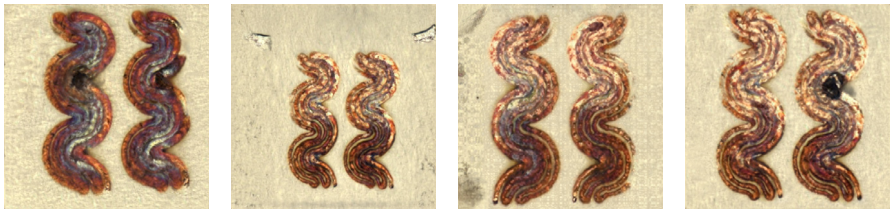
Table 1. Analysis and Evaluation of Sinusoidal Wobble Laser Weld Defects.

Defect Categories	Feature Description	Detection Challenges
Fracture	The weld exhibits obvious cracks or gaps, resulting in the failure of the weld.	The defects exhibit diverse morphologies, and some fractures are small in size, making them difficult to distinguish.
Excessive Penetration	The weld exhibits localized depressions or craters, which may be accompanied by oxide accumulation.	The shape is irregular, with blurred edges.
Deposition Variation	An anomaly in the deposited solder volume, encompassing both excessive accumulation and insufficient deposition, which causes the weld's morphology to deviate from the standard.	The surface features are subtle, showing little difference from normal welds; cases of insufficient deposition can be easily confused with fractures.
Pad Defect	Stains or geometric deformations appear on the pad, outside the weld seam area.	Pad contamination is readily identifiable, whereas geometric deformations require the distinction of subtle variations.

Ultimately, 700 generated images were selected and annotated to form the augmented dataset. The distribution of defect instances within these 700 synthetic images is detailed in Table 2 (under "Synthetic Label Count"). These synthetic instances are then added to the original training set, substantially increasing the total instance counts—particularly for scarce categories such as Excessive Penetration and Deposition Variation. To visually demonstrate the quality and diversity of the generated samples, Figure 3 displays representative images produced by StyleGAN2-AFMS. This hybrid dataset provides a more balanced and diverse data foundation for subsequent detector training.

Table 2. Statistics of instances in the augmented SWLWD Dataset training set.

Defect Categories	Original Label Count	Synthetic Label Count
Fracture	505	417
Excessive Penetration	190	286
Deposition Variation	254	415
Pad Defect	648	512



(**a**) F_ EP (**b**) DV + PD (**c**) F + DV + PD (**d**) F + EP + DV

Figure 3. Representative multi-defect samples generated by StyleGAN2-AFMS. Abbreviations: F—Fracture, EP—Excessive Penetration, DV—Deposition Variation, PD—Pad Defect. (**a**) F + EP; (**b**) DV + PD; (**c**) F + DV + PD; (**d**) F + EP + DV.

2.2. StyleGAN2-AFMS: A Frequency-Domain Modeling-Based Generative Network

High-quality data serves as the cornerstone of deep learning models. In industrial scenarios, the scarcity of defect samples and the complexity of their morphology—particularly the coexistence of global periodic textures and local non-periodic defects—pose stringent requirements on data generation models. Although the standard StyleGAN2 offers a degree of flexibility by using a learnable constant tensor as the starting point for generation, it lacks the capability for explicit modeling of intrinsic image structural priors. This results in inherent limitations when reconstructing the strong periodic patterns characteristic of solder joint surfaces, often yielding images with poor texture consistency and blurred defect details. To address this fundamental challenge, this paper proposes the StyleGAN2-AFMS model. Its core idea is to introduce frequency-domain modeling into the generation process, replacing the original fixed constant input. To realize this idea, an AFMS module based on a frequency-space dual-stream mechanism was designed. This module collaboratively encodes the global periodic structures and non-periodic local details of an image, thereby significantly enhancing the generator's capability to represent complex defect patterns. The overall architecture of the StyleGAN2-AFMS generator is illustrated in Figure 4.

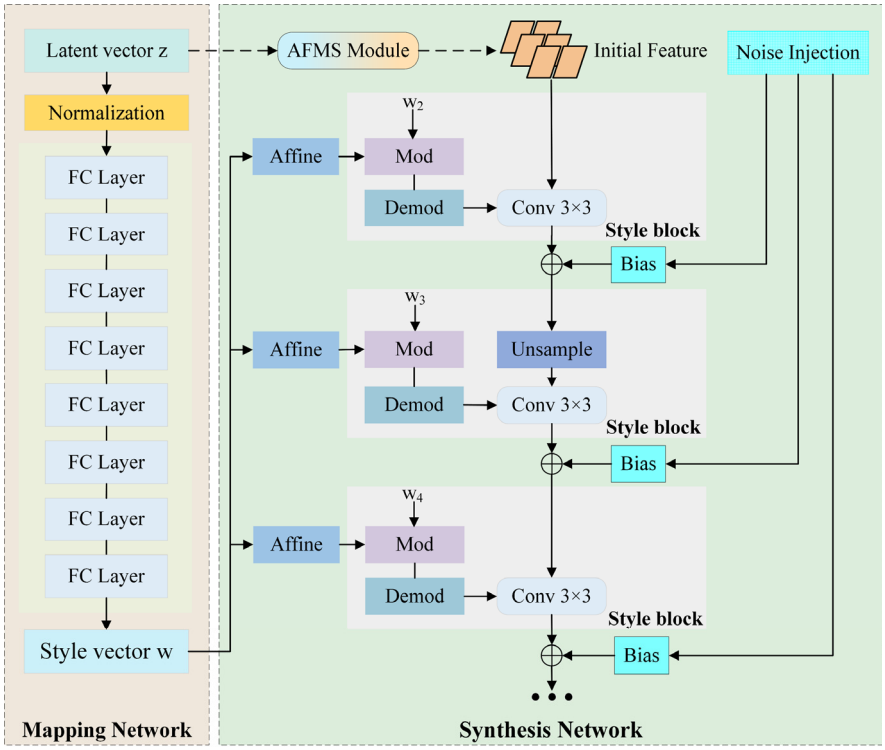


Figure 4. Overall architecture of the StyleGAN2-AFMS generator.

2.2.1. AFMS Module: Collaborative Frequency-Space Feature Generation

The design philosophy of the AFMS module is inspired by the complementary nature of frequency-domain and spatial-domain analysis in signal processing [32,33]. Fourier theory indicates that the frequency domain efficiently encodes global periodic structures of an image, while the spatial domain is more suitable for representing local detailed features. Given that the target weld images contain both strong periodic textures and non-periodic defects, the AFMS module employs a dual-stream architecture to process the input signal in parallel: one branch focuses on extracting spectral structural information from the image, while the other captures spatial detail features. The advantages of both domains are synergistically fused via an adaptive mechanism. Its structure is depicted in Figure 5.

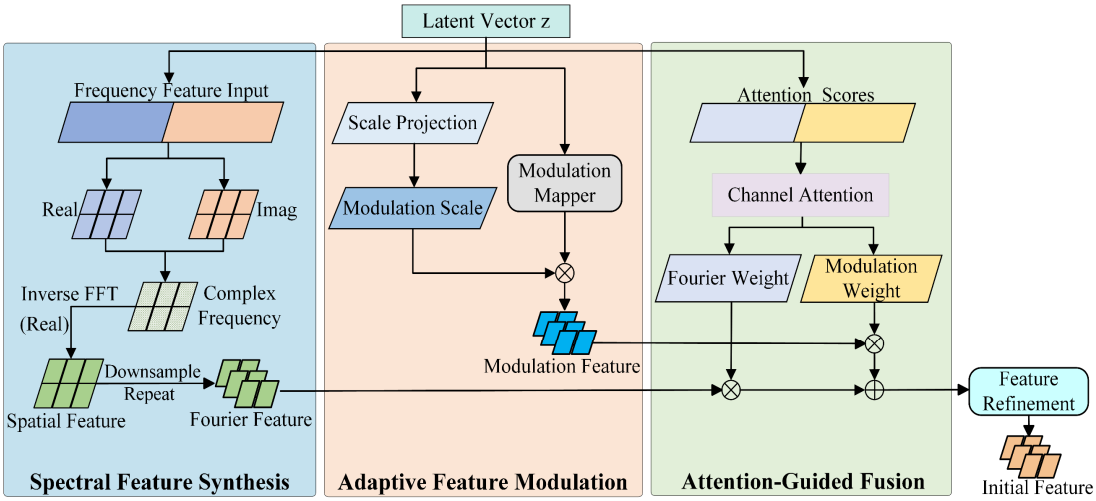


Figure 5. Architecture of the AFMS module.

The AFMS module adopts a hierarchical processing pipeline comprising three sub-modules: the Frequency-domain Feature Stream, the Modulation Feature Stream, and the Adaptive Channel Attention Fusion. These sub-modules are responsible for extracting spectral structures, generating local details, and dynamically integrating both types of information, respectively, thereby systematically constructing high-quality defect feature representations.

(i) Spectral Feature Stream

The latent vector z is first mapped by a fully connected layer into an 8×8 complex frequency-domain representation [34]. Compared with a 4×4 mapping, the 8×8 representation captures richer frequency components and finer periodic patterns. This representation is then projected back to the spatial domain using the Inverse Discrete Fourier Transform (IDFT) to reconstruct spatial structural features. For an M × N frequency input, the two-dimensional IDFT is defined by Equation (1).

$$f(x,y) = \frac{1}{MN} \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} F(u,v) e^{j2\pi(\frac{ux}{M} + \frac{vy}{N})} \quad (1)$$

where $f(x,y)$ denotes the resulting spatial domain feature map, represented as f_{freq}, and j is the imaginary unit. In practice, this transformation is implemented via its efficient algorithm, the Inverse Fast Fourier Transform (IFFT). This process enables the generator to directly perceive and reconstruct the inherent periodic patterns of the welds from the frequency dimension, thereby providing a strong global structural prior for subsequent synthesis.

(ii) Modulation Feature Stream

This branch focuses on capturing and generating local, non-periodic perturbations and structural variations—i.e., the fine details of various defects. In contrast to the global nature of the frequency-domain stream, the core objective of the modulation stream is to decode rich spatial detail features from the latent vector z. Specifically, the vector z is first processed by a feature mapping network (Modulation Mapper) composed of fully connected layers, non-linear activation functions, and transposed convolutions. This network generates a high-dimensional spatial feature map $F'_{\text{mod}} \in \mathrm{R}^{C\times 4\times 4}$ by transforming points in the latent space into feature representations with spatial dimensions through multi-layer non-linear mappings, thereby laying the foundation for generating the morphology of various defects. In parallel, a latent vector-based Channel-wise Modulation mechanism is introduced. This mechanism derives a set of adaptive, channel-specific scaling factors γ (Modulation Scale) from the latent vector z via a lightweight, fully connected layer.

$$\gamma = \sigma(\text{Linear}(z)) \tag{2}$$

where σ the Sigmoid activation function. This scaling factor γ is used for the channel-wise dynamic calibration of the modulated feature map. The final modulated features are given by Equation (3).

$$F_{\text{mod}} = \gamma \odot F'_{\text{mod}} \tag{3}$$

where $\odot$ denotes channel-wise multiplication. This design enables the generator to dynamically activate or suppress feature channels within the modulation stream according to different input vectors z, thereby allowing precise control over the intensity and combination of various defect details in the generated images. This significantly enhances the diversity and controllability of the model's output.

(iii) Adaptive Channel Attention Fusion

To achieve adaptive fusion of the global structural features from the frequency domain and the local detail features from the spatial domain, a lightweight channel attention mechanism based on the latent vector z is introduced. This mechanism maps the latent vector z to a weight vector via a fully connected layer. After being normalized via the Softmax function, this vector is decomposed into a weight $w_{\text{freq}}(z)$ dedicated to the frequency-domain stream and a weight $w_{\text{mod}}(z)$ for the modulation stream, satisfying the constraint $w_{\text{freq}}(z) + w_{\text{mod}}(z) = 1$. The fusion process is given by the following equation.

$$w_{\text{freq}}(z), w_{\text{mod}}(z) = \text{Softmax}(\text{FC}(z)) \tag{4}$$

$$F_{\text{fused}} = w_{\text{freq}}(z) \odot F_{\text{freq}} + w_{\text{mod}}(z) \odot F_{\text{mod}} \tag{5}$$

This mechanism enables the model to dynamically adjust the contribution of the two types of features based on the input vector z, thereby flexibly accommodating the varying emphasis on global structure versus local details required by different defect patterns. The fused feature F_{fused} is then further refined by a lightweight convolutional layer to enhance spatial consistency and suppress redundant features. Finally, the processed feature map serves as the input to the generator.

2.2.2. ADA Strategy

During the training of StyleGAN2-AFMS, the discriminator typically relies on a large volume of training data to avoid overfitting. However, on small datasets with limited defect samples, the discriminator is highly prone to overfitting the training data, leading to an unstable training process. To mitigate this issue, this study adopts the ADA strategy.

This strategy randomly applies a series of augmentation operations to the images input to the discriminator during training. The overall workflow is illustrated in Figure 6.

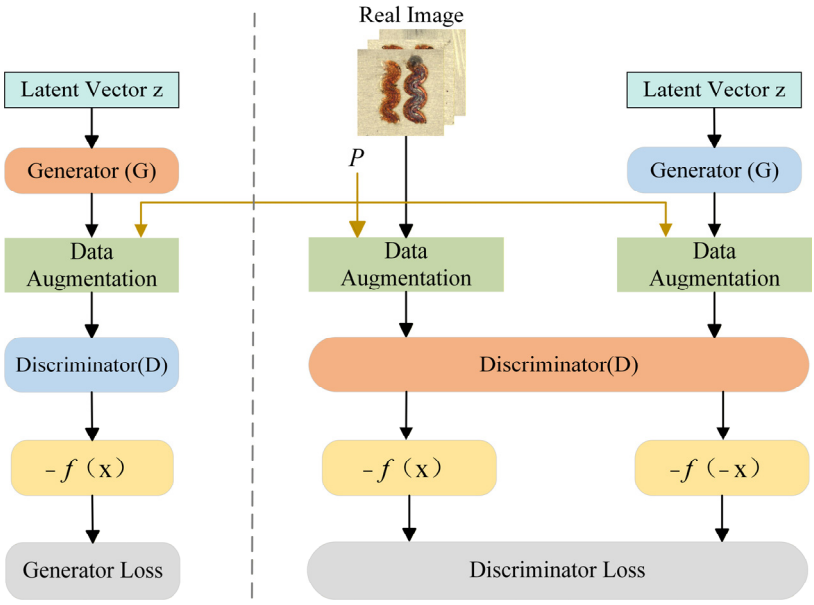


Figure 6. Architecture of the Adaptive Discriminator Enhancement.

In the diagram, the orange-colored Generator (G) and Discriminator (D) represent the networks undergoing training. Image augmentation operations—including rotation, color transformation, noise addition, etc.—are applied, each with a preset probability value p. During each training iteration, these augmentations are randomly applied to the images received by the discriminator, thereby increasing the diversity of the training data. Simultaneously, the application probability of these augmentations is dynamically adjusted based on the discriminator's degree of overfitting, enhancing training stability and improving the quality of the generated images.

2.3. YOLO11n-WAFE: A Lightweight Detector for Weld Defects

The defect detection task for sinusoidal wobble laser welds is challenged by multiple factors simultaneously: minute defects that result in missed detections, diverse defect morphologies, and stringent real-time requirements. Single-stage detectors like the YOLO series have become the preferred architecture for industrial quality inspection due to their favorable balance between speed and accuracy. Among them, YOLO11n [35], an excellent and stable lightweight representative of the series, introduces a range of novel modules such as C3k2 and C2PSA. This achieves a superior balance between computational efficiency, feature representation capability, and model lightweighting, providing a powerful modern baseline for real-time detection tasks.

However, when applied to the specific scenario of sinusoidal wobble laser welds, the model still faces two severe challenges. Firstly, the downsampling strategy adopted to meet real-time requirements may still lead to the loss of critical features when processing minute defects, resulting in missed detections. Secondly, the discriminative capability of its standard FPN for multi-scale coexisting defects against the strong periodic texture background requires further enhancement. To address these issues, an improved detection architecture named YOLO11n-WAFE (Welding–ADownECA Fusion Enhancement) is proposed, specifically designed for weld defects based on the efficient foundation of YOLO11n, as shown in Figure 7.

(i) ADownECA Downsampling Module

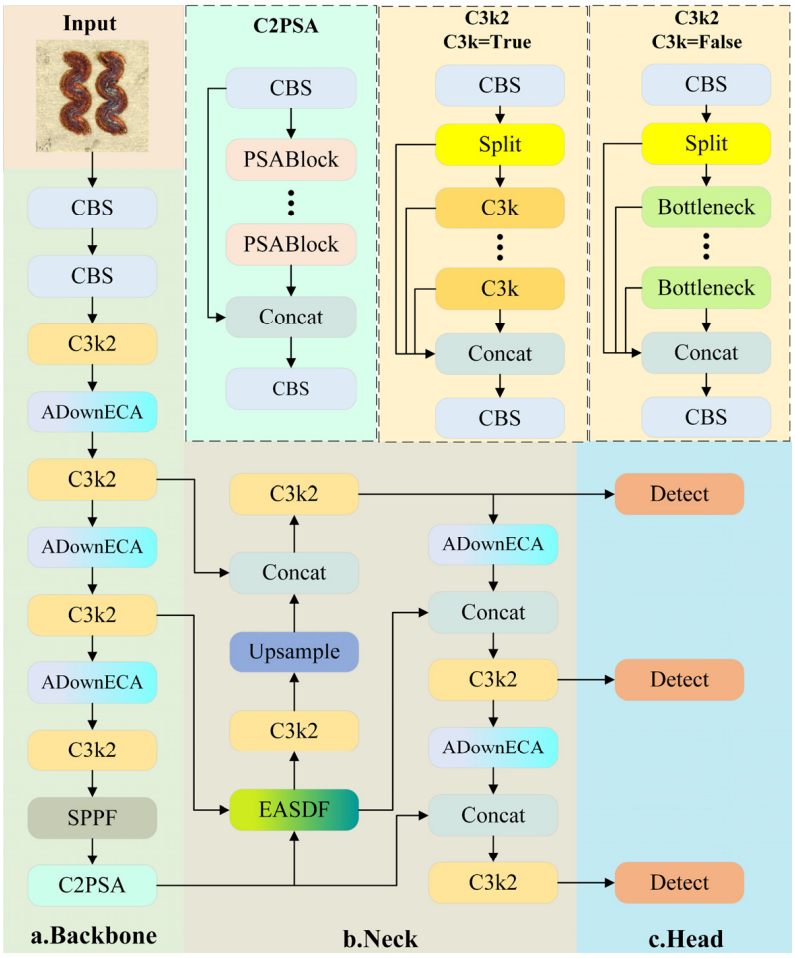


Figure 7. Overall architecture of the proposed YOLO11n-WAFE network.

This module is designed to replace the original downsampling operation. By incorporating a multi-path structure and a lightweight attention mechanism, it enhances the retention and discrimination of minute defect features while maintaining inference speed, with the explicit aim of reducing the missed detection rate.

(ii) EASDF Module

This module optimizes the multi-scale feature fusion process through explicit edge enhancement and an entropy-aware weighting strategy. It enables the model to more accurately distinguish defects of different categories and sizes, thereby adapting to scenarios where multiple defects coexist within a single weld.

2.3.1. ADownECA Downsampling Module

In the task of detecting defects in sinusoidal wobble laser welds, the traditional convolutional downsampling method employed by the YOLO11n model, while prioritizing computational efficiency, is highly prone to losing critical features of minute defects (such as micro-scale Excessive Penetration). This often leads to missed detections and poses a significant quality control risk. To address this issue, the ADown module [36], originating from YOLOv9, is introduced as a foundation. Its multi-path parallel structure is more effective than standard convolution at preserving spatial detail information.

However, while the dual-path structure of the ADown module processes features separately along the channel dimension, enabling parallel extraction of diverse features, it may weaken the synergy and information interaction between the paths. This is particularly

critical for accurately identifying subtle defects that require the fusion of multi-channel contextual information. To reinforce this cross-channel interaction and focus on key features, an Efficient Channel Attention (ECA) module [37] is embedded after feature concatenation. The ECA mechanism acquires channel-wise statistics via lightweight global average pooling and implements local cross-channel interaction using a one-dimensional convolution with an adaptive kernel size k. This effectively enhances the responses of channels sensitive to minute defects, achieving this with nearly no increase in computational complexity. Ultimately, the combination of the refined downsampling capability of the ADown module and the intelligent channel calibration of the ECA mechanism forms the proposed ADownECA module. This module significantly enhances the retention and discrimination of minute defect features while preserving the model's lightweight nature, thereby laying a solid foundation for subsequent accurate detection. The structure of the ADownECA module is illustrated in Figure 8.

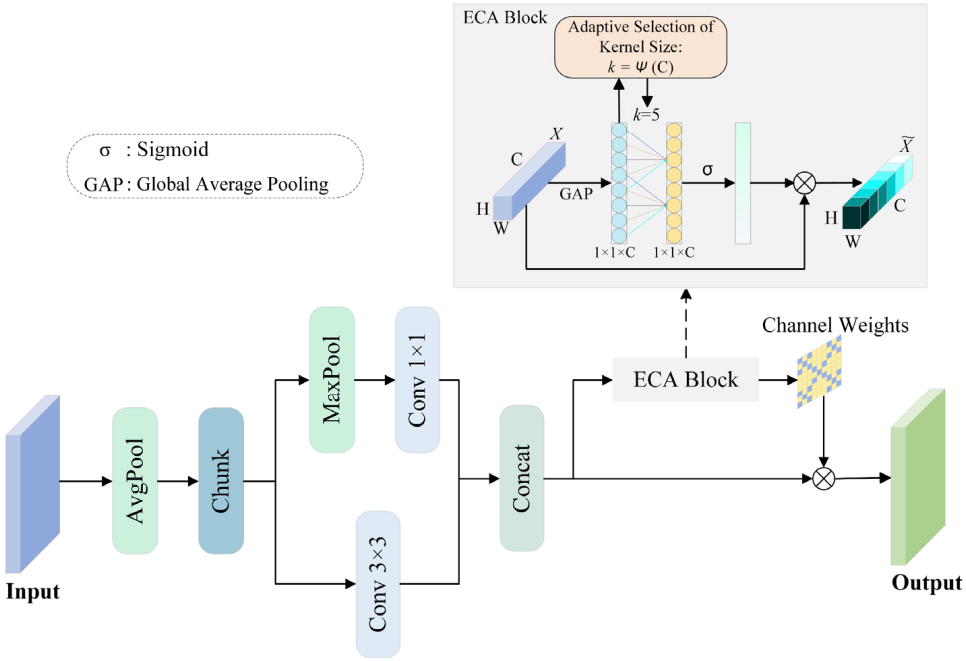


Figure 8. Structure of the proposed ADownECA module.

2.3.2. EASDF Module

During the laser welding process, the quality of welds is critical to product performance. However, the multi-scale nature of weld defects and the frequent coexistence of multiple defects within a single weld pose severe challenges for feature fusion: minute-scale defects (e.g., micro-scale Excessive Penetration) rely on fine local gradient information from low-level feature maps, while larger-scale defects (e.g., Deposition Variation, large-scale Fractures) require the global semantic context provided by high-level feature maps. Although FPN and their variants have become the standard solution for multi-scale detection, their direct fusion operations—such as concatenation or element-wise addition—exhibit inherent limitations in sinusoidal wobble laser weld defect inspection. The underlying issue is twofold. Firstly, subtle pad defects such as contamination in low-level features can be easily confused with the background, introducing significant noise. Secondly, the fine spatial details lost in high-level features due to repeated downsampling are difficult to recover effectively, leading to blurred defect localization. Moreover, simple linear fusion strategies lack adaptability, failing to distinguish between information-rich critical regions and redundant background textures.

To address the aforementioned challenges, inspired by anisotropic diffusion theory [38] and variational principles, the EASDF module is proposed, whose overall structure is illustrated in Figure 9. The module consists of three synergistic sub-modules: Edge Gradient Enhancement, Semantic–Detail Separation, and Entropy-Aware Dynamic Fusion. Unlike traditional FPN, EASDF incorporates a lightweight edge enhancement operator based on PDEs to preserve and amplify edge information critical for defect discrimination, while simultaneously adopting an entropy-aware dynamic weighting strategy to adaptively suppress background redundancy. This approach significantly enhances the robustness and precision of cross-scale feature fusion by maximizing the retention of discriminative details while inhibiting noise interference. The module takes the low-level feature map L and the high-level feature map H from the backbone network as inputs, where L contains rich spatial details and H encodes high-level semantics. These inputs first undergo edge enhancement to extract and modulate gradient responses, are then processed via semantic–detail separation, and are finally fused adaptively using entropy-aware dynamic weights to produce the final feature representation for detection.

(i) Edge Gradient Enhancement

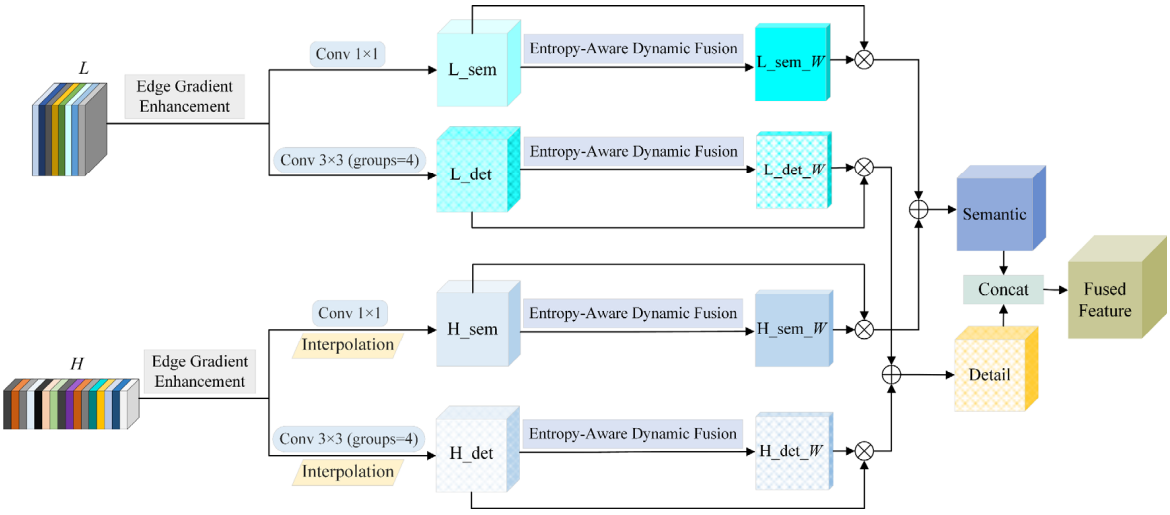


Figure 9. Structure of the proposed EASDF module.

This step is designed to explicitly enhance the edge and detail information within the feature maps that is critical for defining defect morphology. Its core involves constructing a controlled diffusion process, the mathematical essence of which is to find the minimum of an energy functional. This functional is dedicated to preserving edge information while smoothing the interior regions of the image.

First, the Sobel operator is employed to extract the gradient magnitude from both L and H to capture edge intensity, as calculated by Equation (6).

$$\nabla mag = \sqrt{(\nabla_x \mathrm{F})^2 + (\nabla_y \mathrm{F})^2 + \epsilon} \tag{6}$$

where $\nabla_x \mathrm{F}$ and $\nabla_y \mathrm{F}$ represent the operations of the Sobel convolution kernels in the x and y directions, respectively, and $\epsilon = 10^{-6}$ is a constant for numerical stability. This gradient magnitude map defines the conduction coefficient for the anisotropic diffusion. Subsequently, an adaptive weight map is generated via a 1×1 convolution followed by a Sigmoid activation, serving as the modulation coefficient g for edge diffusion. Its calculation is given by Equation (7).

$$g = \sigma(\mathrm{conv}_{1\times1}(\nabla \mathrm{mag})) \tag{7}$$

where σ denotes the Sigmoid activation function.

Based on the above definitions, the following discrete iterative update rule is derived, as shown in Equation (8), which approximates the solution to the corresponding partial differential equation.

$$\mathbf{F}^{(t+1)} = \mathbf{F}^{(t)} + \tau \cdot g \cdot \nabla^2 \mathbf{F}^{(t)} \tag{8}$$

where $\mathbf{F}^{(t)}$ denotes the feature map at the t-th iteration, τ is a learnable evolution step size parameter (initialized to 0.1), ∇^2 is the Laplacian operator used to compute the diffusion flow, and the number of iterations is set to 2. During this process, the feature map evolves iteratively based on the gradient information, enhancing edges while smoothing homogeneous regions, thereby suppressing noise and preserving critical information. Finally, a gated residual mechanism is introduced to adaptively fuse the PDE-enhanced features with the original features, yielding the edge-aware enhanced feature representation as formulated in Equation (9).

$$\mathbf{F}_{enhanced} = w \odot \mathbf{F} + (1 - w) \odot \mathbf{F}_{orig} \tag{9}$$

where $\odot$ denotes element-wise multiplication, and w represents the gating weights learned through a 1×1 convolution and a Sigmoid function, which regulate the flow of enhanced information. This mechanism ensures training stability and allows the network to adaptively select the degree of enhancement. The internal structure of the Edge Gradient Enhancement sub-module is illustrated in Figure 10.

(ii) Semantic–Detail Separation

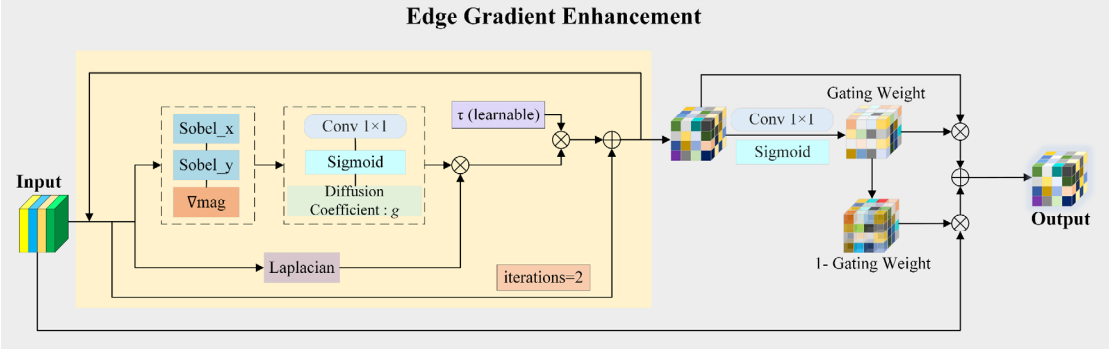


Figure 10. Structure of the Edge Gradient Enhancement block.

To prevent interference between different types of information, the enhanced high-level and low-level features are fed into two parallel branches for preprocessing. The semantic branch employs a 1×1 convolutional layer to distill global contextual information from $H_{enhanced}$, focusing on capturing the overall structure of large-scale defects. The detail branch, conversely, extracts local texture and edge features from $L_{enhanced}$ using grouped convolution, aiming to preserve the subtle variations characteristic of minute defects.

(iii) Entropy-Aware Dynamic Fusion

A key innovation of this module is its adoption of a weighting strategy based on Shannon Entropy from information theory to achieve adaptive fusion of the branch features. This strategy effectively quantifies the richness of semantic information and the uncertainty in different regions by calculating the spatial entropy E of each branch's feature map, as formulated in Equations (10) and (11).

$$p_c = \frac{|\mathbf{X}_c|}{\sum_{c=1}^{C} |\mathbf{X}_c| + \epsilon} \tag{10}$$

$$\mathrm{H} = -\sum_{c=1}^{C} p_c \cdot \log(p_c + \epsilon) \tag{11}$$

where p_c represents the absolute probability distribution along the channel dimension. A region with a higher entropy value H indicates that the corresponding spatial location carries greater informational complexity and uncertainty, which are predominantly located near defect edges or texturally rich areas. To balance the overall distribution characteristics of the features and mitigate potential local sensitivity arising from reliance on entropy alone, the feature mean is also introduced as an auxiliary regulating factor. This mean term provides a global perspective on feature intensity, forming an effective complement to the structural complexity captured by entropy. The final fusion weight map W is determined jointly by the spatial entropy map and the feature mean, as calculated by Equation (12).

$$W = \sigma\left(\alpha \cdot \mathrm{H} + \beta \cdot \frac{1}{C}\sum_{c=1}^{C} \mathbf{x}_c\right) \tag{12}$$

where α and β are learnable scaling parameters that balance the contributions of the entropy and mean values in the weight calculation, and σ denotes the Sigmoid activation function. This weight map is ultimately used to guide the weighted fusion of semantic and detail features, enabling the model to dynamically focus on informationally richer regions, thereby achieving more precise feature integration. Figure 11 illustrates the structure of the Entropy-Aware Dynamic Fusion sub-module.

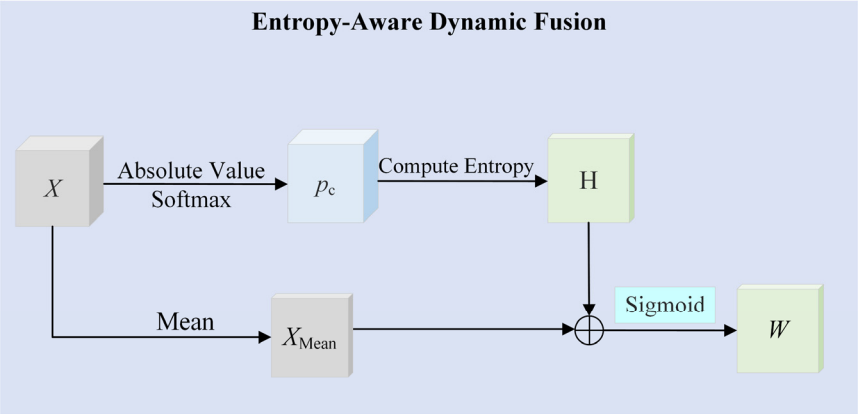


Figure 11. Structure of the Entropy Weight Calculation block.

The final output is produced by first applying the adaptive weights to their respective features through weighting and summation, which yields the enhanced semantic and detail representations. These two streams of information are then concatenated to form the fused feature map for subsequent detection stages.

Through the above process, the EASDF module demonstrates superior performance in multi-scale defect detection, providing a more robust and precise feature representation for laser welding quality assessment.

2.4. Model Evaluation Metrics

2.4.1. Generative Model Quality Assessment

In industrial defect detection tasks, the quality of generated samples must simultaneously satisfy two core requirements: high fidelity (avoiding introducing noise that could interfere with the detection model) and high coverage (covering diverse defect morphologies), to effectively support the training and evaluation of downstream detection models. This paper employs three metrics for quantitative assessment: FID, Precision, and Recall.

These metrics provide a comprehensive analysis of model performance from three distinct dimensions: distribution similarity, authenticity of generated samples, and diversity.

(i) FID

FID is a widely used metric for evaluating generative models, designed to measure the overall distribution similarity between generated and real images within a feature space. Specifically, FID leverages a pre-trained Inception V3 model to extract image features and assesses generation quality by computing the Fréchet distance between the mean and covariance of the features of the generated and real image sets. The calculation of FID is given by Equation (13).

$$\mathrm{FID} = \| \mu_r - \mu_g \|^2 + Tr\left(\Sigma_r + \Sigma_g - 2\left(\Sigma_r\Sigma_g\right)^{\frac{1}{2}}\right) \tag{13}$$

where r and g denote the real and generated images, μ_r and Σ_r are the mean and covariance of the real image feature distribution, μ_g and Σ_g represent the mean and covariance of the generated image feature distribution, and Tr denotes the trace of a matrix. A lower FID value indicates that the distribution of generated images is closer to the real distribution, reflecting higher generation quality.

(ii) Precision and Recall

Precision measures the authenticity of generated samples by calculating the proportion that falls within the neighborhood of the real data manifold, while Recall reflects the diversity by quantifying the proportion of real samples covered by the generated data manifold.

Let the set of feature vectors of real samples and the set of feature vectors of generated samples be denoted as above. For each sample, the distance to its k-th nearest neighbor (k = 3) within the same set is computed. The calculations for Precision and Recall are given by Equations (14)–(16).

$$f(\phi, \Phi) = \begin{cases} 1, \text{if } \| \phi - \phi' \|_2 \leq \| \phi' - \mathrm{NN}_k(\phi', \Phi) \|_2 \text{ for at least one } \phi' \in \Phi \\ 0, \text{otherwise,} \end{cases} \tag{14}$$

$$\mathrm{precision}\left(\Phi_r, \Phi_g\right) = \frac{1}{\left|\Phi_g\right|} \sum_{\phi_g \in \Phi_g} f\left(\phi_g, \Phi_r\right) \tag{15}$$

$$\mathrm{recall}\left(\Phi_r, \Phi_g\right) = \frac{1}{\left|\Phi_r\right|} \sum_{\phi_r \in \Phi_r} f\left(\phi_r, \Phi_g\right) \tag{16}$$

where ϕ_r and ϕ_g represent feature vectors of real and generated images respectively, $\| \phi - \phi' \|_2$ denotes the Euclidean distance between vectors, and $\mathrm{NN}_k(\phi', \Phi)$ returns the k-th nearest neighbor of ϕ' in set Φ. The function $f(\phi, \Phi_r)$ evaluates the authenticity of a given image, while $f\left(\phi, \Phi_g\right)$ assesses its likelihood of being generated.

2.4.2. Detection Model Performance Evaluation

This experiment employs six key metrics to comprehensively evaluate model performance: F1-score, mAP@0.5, mAP@0.5:0.95, number of parameters (Parameters), inference speed (FPS), and model size (Size). The F1-score, which balances both precision and recall, measures the overall detection capability of the model in the weld defect detection task. Its calculation is given by Equations (17)–(19).

$$\mathrm{Precision} = \frac{TP}{TP + FP} \tag{17}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{18}$$

$$\text{F1} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{19}$$

Here, *TP* (True Positives) denotes the number of actual positive samples correctly identified as positive by the model, *FP* (False Positives) represents the number of negative samples incorrectly classified as positive, and *FN* (False Negatives) indicates the number of positive samples mistakenly predicted as negative by the model.

mAP is used to measure the model's comprehensive accuracy in detecting various types of weld defects, reflecting its detection capability across different defect categories. The calculation is given by Equations (20) and (21).

$$\text{AP} = \int_0^1 P(R)dr \tag{20}$$

$$\text{mAP} = \frac{\sum \text{AP}}{N_{cls}} \tag{21}$$

where N_{cls} denotes the total number of defect categories.

The model's parameter count (Parameters) reflects its structural complexity and storage requirements. The inference speed (FPS, Frames Per Second) measures the processing speed during inference, indicating the number of images processed per second. The model size represents the disk storage footprint of the final weight file.

3. Results and Analysis

3.1. Experimental Platform and Parameter Settings

All experiments in this study were conducted on the SWLWD Dataset. The experimental setup encompasses both the hardware environment and model training configurations, with detailed specifications provided in Tables 3–5. Specifically, Table 3 outlines the hardware environment, while Tables 4 and 5 present the training configurations for StyleGAN2-AFMS and YOLO11n-WAFE, respectively.

Table 3. Experimental Environment Configuration.

Parameter	Parameter Value
Operating system	Ubuntu 20.04 LTS
CPU	Intel Xeon Silver 4310
GPU	RTX 3090
Programming language	Python 3.8.19
Deep learning framework	Pytorch 1.13
CUDA Version	11.7

Table 4. StyleGAN2-AFMS Training Configuration.

Parameter	Parameter Value
Batch Size	4
Total Training	500 kimg
Optimizer	Adam
Path Length Regularization	8.2
Data Augmentation	Mirror

Table 5. YOLO11n-WAFE Training Configuration.

Parameter	Parameter Value
Batch Size	16
Epochs	100
Optimizer	SGD
Initial learning rate	0.003
Learning rate schedule	Cosine annealing

3.2. Optimization of Data Augmentation Strategy for YOLO11n-WAFE Model

To address the characteristics of the SWLWD Dataset in industrial scenarios, this paper optimizes the online data augmentation strategy of YOLO to enhance the robustness and generalization capability of the YOLO11n-WAFE model. Key modifications include: reducing the probability of Mosaic augmentation to avoid introducing overly complex backgrounds in few-sample scenarios, which could impede model convergence; introducing Copy–Paste augmentation to increase the diversity of defect samples; expanding the range of rotation angles to simulate angular variations in workpieces in production environments; and adopting Mixup augmentation to improve model generalization. The specific parameter settings are provided in Table 6.

Table 6. Optimized Configuration of YOLO11n-WAFE Data Augmentation Parameters.

Parameter	Parameter Value
Operating system	Ubuntu 20.04 LTS
CPU	Intel Xeon Silver 4310
GPU	RTX 3090
Programming language	Python 3.8.19
Deep learning framework	Pytorch 1.13
CUDA Version	11.7

3.3. Quality Evaluation of Generated Images

For evaluation, the proposed StyleGAN2-AFMS model is compared against several generative models, including StyleGAN2, StyleGAN2-ADA, and StyleGAN3-t. The evaluation employs FID, Precision, and Recall to quantitatively analyze distribution similarity, fidelity, and diversity, respectively. The results are summarized in Table 7.

Table 7. Performance comparison of different generative models and ablation variants. Significant values are in bold.

Network Architecture	FID	Precision	Recall
StyleGAN2	22.36	0.495	0.234
StyleGAN2-ADA	16.79	0.526	0.380
StyleGANA3-t	16.44	0.538	0.410
AFMS (Frequency-only)	17.22	0.540	0.445
AFMS (Modulation-only)	16.90	0.537	0.423
AFMS (Fixed-weight)	16.33	0.542	0.458
StyleGAN2-AFMS(Ours)	**15.60**	**0.551**	**0.470**

As shown in Table 7, among the compared StyleGAN series models, the proposed StyleGAN2-AFMS achieves the most competitive results across key metrics. Specifically, it attains the lowest FID score (15.60), indicating superior distributional alignment between the generated and real images. The model also obtains the highest Recall (0.470), demonstrating its ability to cover a broader manifold of real data and generate more diverse defect

patterns, which is crucial for mitigating data scarcity in few-sample scenarios. Furthermore, its Precision (0.551) surpasses other baseline models, reflecting that the generated samples maintain high authenticity and low noise levels while preserving diversity. These results validate that the introduced frequency-space synergistic mechanism effectively enhances generation quality by jointly optimizing the representation of global periodic structures and local non-periodic defects.

To further validate the design of StyleGAN2-AFMS, we include its ablation variants in Table 7. The AFMS (Frequency-only) model retains only the frequency-domain stream for periodic texture modeling, achieving an FID of 17.22. The AFMS (Modulation-only) model retains only the spatial modulation stream for defect detail generation, yielding an FID of 16.90. The significant performance gap between these single-stream variants and the complete model (FID = 15.60) demonstrates the complementary and synergistic effect of the proposed dual-stream architecture. The AFMS (Fixed-weight) variant, which replaces the adaptive fusion mechanism (Equations (4) and (5)) with a fixed, equal weighting scheme (i.e., $w_{\text{freq}} = w_{\text{mod}} = 0.5$), achieves an FID of 16.33. While this fixed-weight fusion outperforms both single-stream variants—confirming the benefit of combining frequency and modulation features—it still lags considerably behind the fully adaptive model (FID = 15.60). This result underscores the importance of the dynamic, input-dependent fusion strategy for optimally balancing global texture and local defect features. We also investigated the sensitivity to the frequency mapping resolution. Among configurations of 4×4, 8×8, and 16×16, the 8×8 resolution yielded the optimal FID score. The 4×4 resolution provided insufficient frequency components for faithful periodic texture reconstruction. The 16×16 resolution captures finer details, yet it introduces high-frequency noise that degrades generation quality and increases parameter count and computational overhead, reducing efficiency. Thus, the 8×8 resolution offers the best trade-off between capturing characteristic patterns and maintaining model efficiency for our dataset.

3.4. Impact of Generated Sample Quantity on Detection Performance

To determine the optimal volume of data generated by StyleGAN2-AFMS while mitigating the risks of both insufficient augmentation and potential overfitting [39], a systematic sensitivity analysis was conducted. Specifically, the original training set was augmented by synthesizing images in a range from 100 to 1000 with a step size of 100 using the trained StyleGAN2-AFMS model, thereby constructing multiple progressively expanded training sets. The YOLO11n-WAFE detector was then independently trained and evaluated on each set under the data augmentation strategy detailed in Section 3.2, with the corresponding detection performance trends shown in Figure 12.

Figure 12 demonstrates a clear non-monotonic relationship between detection performance and the number of generated samples, characterized by an initial improvement followed by a gradual decline. During the enhancement phase spanning from 0 to 700 samples, all evaluation metrics showed consistent improvement: mAP@0.5 increased from the baseline of 88.3% to a peak of 91%, while mAP@0.5:0.95 rose significantly from 63.5% to 74.2%. This progression indicates that the additional synthetic samples effectively enriched the training data diversity, providing more comprehensive defect feature representations and enabling the learning of more robust features. The model achieved an optimal performance balance at 700 samples, establishing this as the preferred augmentation volume. In the subsequent phase from 700 to 1000 samples, metrics displayed a slight deterioration—for instance, mAP@0.5 decreased from 91.0% to 90.5%—likely attributable to the introduction of low-diversity samples or subtle noise from excessive synthetic images, which may interfere with training and compromise generalization capability [40].

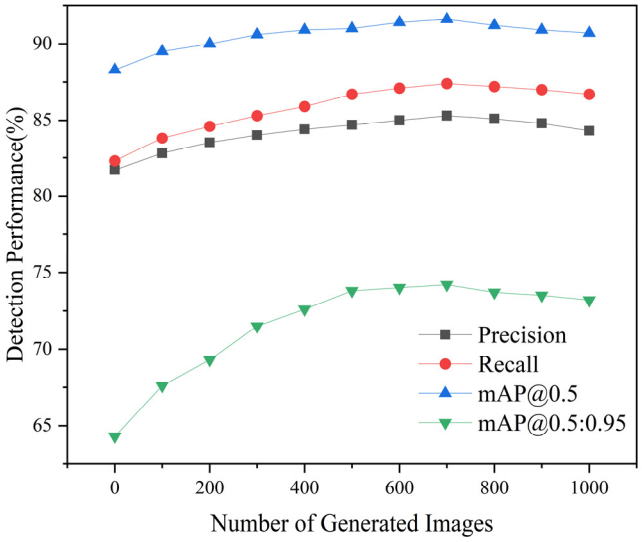


Figure 12. Impact of the Number of Generated Samples on Detection Performance.

The experimental results underscore that in few-shot scenarios, employing more generated samples does not invariably lead to better performance; instead, an optimal augmentation range exists. For the SWLWD dataset in this study, 700 generated samples represent the ideal augmentation scale, maximizing performance gains while circumventing the negative effects of excessive augmentation. Consequently, all ablation and comparison experiments in subsequent Sections 3.5 and 3.6 adopt this optimal data configuration.

3.5. Ablation Studies

Building on the established optimal data augmentation strategy (Section 3.2), validated generative data quality (Section 3.3), and determined augmentation scale (Section 3.4), this subsection presents a systematic ablation study to evaluate the individual and collective contributions of the proposed components in the YOLO11n-WAFE detector. The experimental design compares the original ADown module against the improved ADownECA module and introduces the EASDF module, assessing the separate and combined impacts of these components on detection performance, computational complexity, and inference efficiency. Evaluation metrics include the F1-score, mAP@0.5 and mAP@0.5:0.95, number of parameters, inference speed (FPS), and model file size. All experiments were conducted under identical environmental configurations and parameter settings. The results of the ablation study are summarized in Table 8.

Table 8. Ablation experiment. Significant values are in bold.

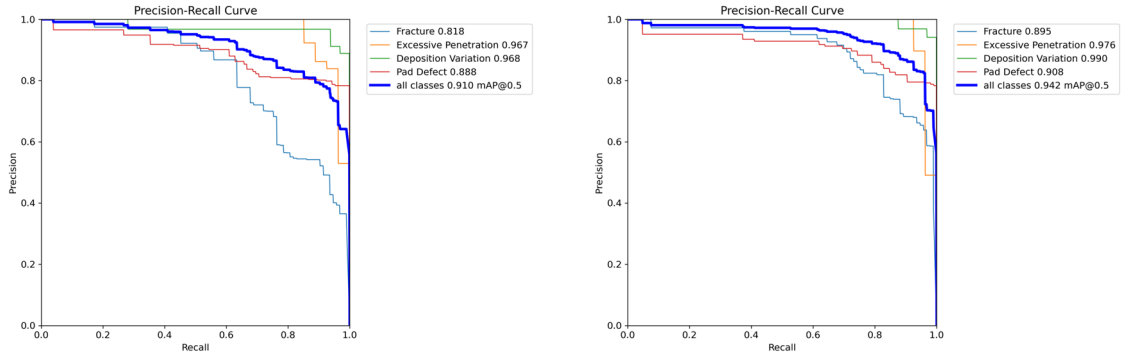
YOLO11	ADown	ADownECA	EASDF	F1	map@0.5	map@0.5:0.95	Parameters/M	FPS	Size/MB
√				86.2	91	74.2	2.58	**688**	5.20
√	√			87.4	91.8	74.9	**2.10**	684	**4.3**
√		√		88.1	92.1	75.8	**2.10**	682	4.31
√			√	88.3	93.8	77.2	2.59	661	5.24
√		√	√	**89.5**	**94.2**	**77.4**	2.15	656	4.36

√ indicates that the corresponding module is used in the experiment.

This study builds upon YOLO11 as the baseline, progressively integrating proposed modules to enhance detection performance for sinusoidal wobble laser weld defects. The ablation study reveals the following: (1) Introducing the ADown module increases the F1-score from 86.2% to 87.4% and mAP@0.5 from 91% to 91.8%, while the number of parameters and model size decrease from 2.58 M and 5.20 MB to 2.10 M and 4.30 MB,

respectively, demonstrating that multi-path downsampling enhances defect feature capture capability. (2) Replacing ADown with ADownECA further improves the F1-score to 88.1% and mAP@0.5 to 92.1% with negligible increase in parameters and model size, confirming the advantage of the attention mechanism in strengthening feature representation. (3) Incorporating EASDF alone elevates mAP@0.5 to 93.8% (F1 = 88.3%), indicating its effectiveness in suppressing interference across multi-scale defects through edge enhancement and semantic–detail fusion. (4) With the synergistic combination of ADownECA and EASDF, the complete YOLO11n-WAFE model achieves an F1-score of 89.5% and mAP@0.5 of 94.2%, with 2.15 M parameters, 4.36 MB model size, and an inference speed of 656 FPS (≈1.52 ms/frame). Although inference speed shows a slight decrease compared to the baseline, it still substantially exceeds the 30 FPS capture rate of typical industrial cameras. This demonstrates an excellent balance among detection accuracy, model lightweight design, and real-time performance.

Furthermore, Figure 13 presents a comparison of Precision–Recall (PR) curves between the YOLO11 baseline model and the improved YOLO11n-WAFE model. The PR curves for each defect category demonstrate that the enhanced model achieves higher precision across most recall ranges, with particularly superior performance in high-recall regions. These results validate the effectiveness of the ADownECA and EASDF modules in enhancing multi-scale defect detection capability and further confirm the overall efficacy of the proposed approach.



(**a**) (**b**)

Figure 13. P-R curves of YOLO11n and YOLO11n-WAFE on the SWLWD Dataset:(**a**) YOLO11n; (**b**) YOLO11n-WAFE.

The above ablation results (Table 8) demonstrate that integrating both ADownECA and EASDF yields the best detection performance. To further investigate the internal design of the EASDF module—particularly the choice of PDE iteration number—and to verify its advantage over simpler edge enhancement alternatives under the same backbone, we conducted an additional ablation study based on the fixed YOLO11n + ADownECA foundation. The results are summarized in Table 9.

Compared with simply replacing the fusion module with static Sobel edge concatenation (+0.6% mAP@0.5 over the ADownECA baseline) or SE attention (+0.9% mAP@0.5), the proposed EASDF module achieves a substantially higher gain (+2.1% mAP@0.5) when integrated with the same ADownECA backbone. This confirms that PDE-based iterative edge enhancement is more effective than simple feature concatenation or channel-wise recalibration for distinguishing subtle defects from periodic backgrounds.

Regarding the PDE iteration number, one iteration yields 93.5% mAP@0.5. Increasing to two iterations achieves the peak performance of 94.2% mAP@0.5, while three iterations cause a slight degradation to 93.8% mAP@0.5 and reduce FPS from 656 to 640 due to over-

smoothing. These results indicate that two iterations provide the optimal balance between detection accuracy and inference speed; thus, we adopt this setting in our framework.

Table 9. Ablation on EASDF iteration number and comparison with lightweight edge enhancement methods.

Variant	F1	map@0.5	map@0.5:0.95	Parameters/M	FPS
YOLO11n + ADown(baseline)	88.1	92.1	75.8	2.10	682
+Replace fusion with Sobel edge concat	88.5	92.7	76.2	2.11	675
+Replace fusion with SE attention	88.8	93.0	76.5	2.12	668
+EASDF (iter = 1)	89.0	93.5	76.9	2.15	662
+EASDF (iter = 2, Ours)	89.5	94.2	77.4	2.15	656
+EASDF (iter = 3)	89.2	93.8	77.1	2.15	640

3.6. Comparison of Different Detection Models

To evaluate the overall performance of YOLO11n-WAFE, comparative experiments were conducted with other mainstream models, with quantitative results presented in Table 10.

Table 10. Comparison of experimental results of mainstream models. Significant values are in bold.

Model	F1	map@0.5	map@0.5:0.95	Parameters/M	FPS	Size/MB
YOLOv5n	85.3	88.8	66.3	2.5	**702**	5.02
YOLOv6n	84.6	90.1	68.4	4.23	694	8.28
YOLOv8n	86.8	90.8	75	3.0	696	6.2
YOLOv9-tiny	87.2	91.3	68.5	**1.97**	535	4.38
YOLOv10n	82.5	89	70.1	2.69	680	5.47
YOLO11n	86.2	91	74.2	2.58	688	5.20
YOLO11s	88.9	93.6	77.8	9.41	345	18.2
RT-DETR	89.3	**94.3**	**78.5**	42.7	98	81.9
YOLO11n-WAFE	**89.5**	94.2	77.6	2.15	656	**4.36**

Table 10 presents a comprehensive performance comparison between YOLO11n-WAFE and mainstream detection models on the SWLWD dataset. The results demonstrate the superior detection accuracy of YOLO11n-WAFE, achieving 94.2% mAP@0.5 and 89.5% F1-score, which significantly outperforms other lightweight models and approaches the performance of larger-scale models such as YOLO11s and RT-DETR. This validates the effectiveness of the proposed ADownECA and EASDF modules in feature preservation and multi-scale fusion. In terms of model efficiency, YOLO11n-WAFE maintains only 2.15 M parameters with a model size of 4.36 MB, while achieving an inference speed of 656 FPS—far exceeding the real-time detection requirement of 30 FPS—thus balancing efficiency and accuracy among lightweight models. Although RT-DETR achieves the highest mAP@0.5:0.95 of 78.5%, its substantial parameter count of 42.7 M and lower inference speed of 98 FPS make it unsuitable for resource-constrained scenarios. In summary, YOLO11n-WAFE achieves an optimal balance among accuracy, speed, and model complexity, making it particularly suitable for high-real-time industrial defect inspection under few-sample conditions.

3.7. Visualization of Detection Results

To further investigate the advantages of YOLO11n-WAFE in defect localization tasks, this study randomly selected multiple defect samples and employed a normalized bounding box heatmap method to compare the full-image visualizations between the baseline YOLO11n model and the improved YOLO11n-WAFE model. The comparison highlights

the models' attention distribution within defect regions and their responses to background information, as shown in Figure 14.

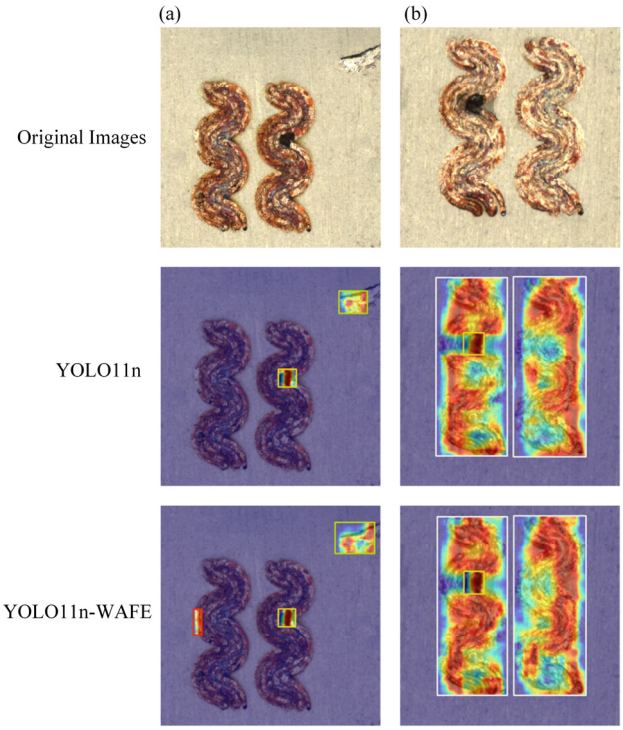


Figure 14. Qualitative comparison of defect detection using heatmaps for YOLO11n and YOLO11n-WAFE on weld images: (**a**) Sample image containing three defects: Fracture, Excessive Penetration, and Pad Defect. (**b**) Sample image containing two defects: Excessive Penetration and Deposition Variation.

The heatmaps demonstrate that YOLO11n-WAFE successfully detects minute defects missed by the baseline model. Moreover, its detection bounding boxes cover larger areas of defective regions, with high-intensity responses distinctly concentrated on the actual defects, effectively suppressing background interference. Furthermore, to more intuitively demonstrate the detection performance and robustness of the YOLO11n-WAFE model in multi-defect scenarios, four additional sinusoidal wobble laser weld images containing mixed defects were selected, with comparative detection results between the baseline YOLO11n model and the improved YOLO11n-WAFE model presented in Figure 15.

As observed in the four comparative results in Figure 15a–d, YOLO11n exhibits certain limitations in complex defect scenarios. For instance, in Figure 15a, the baseline model only detects the main fracture region while missing the accompanying Excessive Penetration, whereas YOLO11n-WAFE successfully identifies both "Fracture" and "Excessive Penetration", demonstrating stronger fine-grained perception capability. In Figure 15b, YOLO11n performs well in detecting single Deposition Variation—correctly identifying both excessive deposition (left weld) and insufficient deposition (Figure 15c). However, when Excessive Penetration coexists with Deposition Variation, its confidence in detecting Deposition Variation significantly decreases, accompanied by slight bounding box localization deviations, and moreover, it erroneously fragments the Excessive Penetration into two separate bounding boxes rather than recognizing it as a unified defect, revealing limited adaptability in multi-defect scenarios. In contrast, YOLO11n-WAFE maintains stable high confidence and precise localization under the same conditions, accurately delineating the Excessive Penetration as a single coherent defect alongside the Deposition Variation, thereby exhibiting stronger robustness and feature discrimination capability in

complex environments. For the scenario involving both Fracture and Pad Defect shown in Figure 15d, both models complete the detection task, but YOLO11n-WAFE achieves more prominent confidence levels, further verifying its more stable recognition capability in multi-category defect tasks.

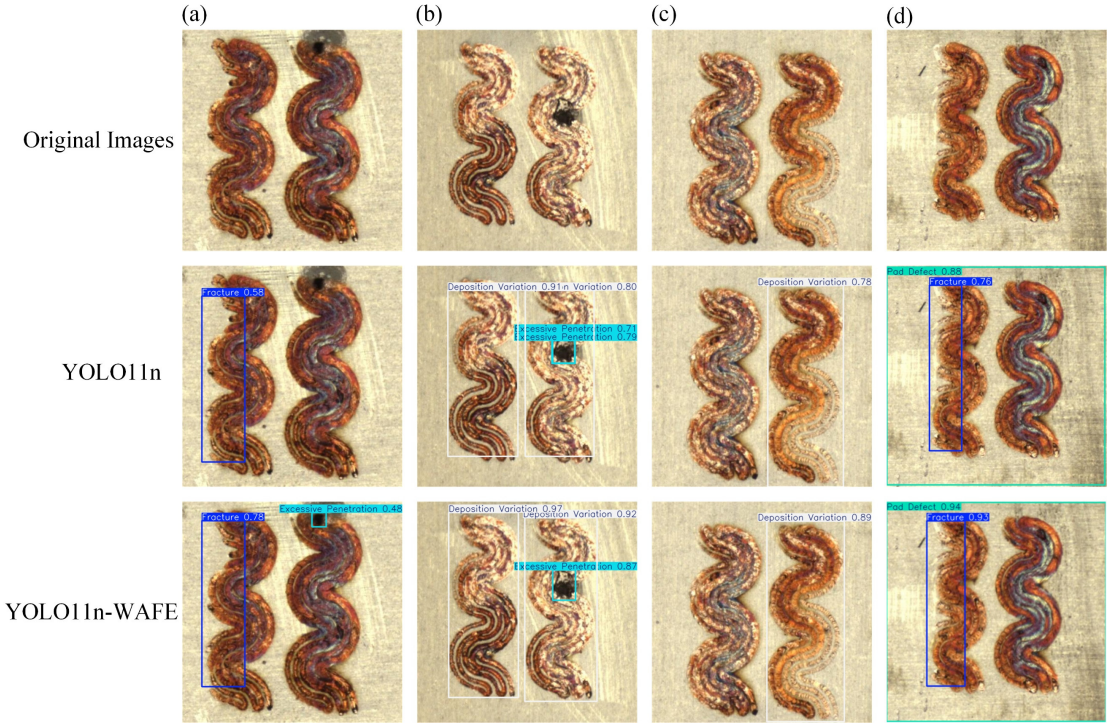


Figure 15. Comparison of detection results between YOLO11n and YOLO11n-WAFE on weld images with mixed defects: (**a**) Fracture and Excessive Penetration; (**b**) Excessive Penetration and Deposition Variation; (**c**) Deposition Variation; (**d**) Fracture and Pad Defect.

The results demonstrate that the improved model comprehensively captures multiple defect types in complex industrial environments, significantly reducing missed detections while substantially increasing detection confidence—reflecting strong adaptability to diverse defect morphologies and environmental interference. The visualization results further indicate that clear bounding box annotations enable quality inspectors to rapidly locate problematic areas, supporting automated screening and manual verification, thereby shortening defect response time and enhancing production line efficiency. Particularly under few-sample conditions with high-precision requirements, the model's consistent performance provides efficient and scalable technical support for solder quality control in electronic manufacturing.

4. Discussion

The core contribution of this study lies in its synergistic integration of StyleGAN2-AFMS and YOLO11n-WAFE, establishing a new paradigm for addressing the challenge of few-shot defect detection in sinusoidal wobble laser welds. While the demonstrated performance stems from the effective response of two core mechanisms to specific industrial challenges, certain limitations remain worthy of consideration.

First, the success of the StyleGAN2-AFMS generative model validates the superiority of the "physics-aware" generation paradigm over general data augmentation methods in highly structured industrial vision tasks. Its frequency-space dual-stream architecture does not merely learn pixel distributions but explicitly decouples and collaboratively models the global periodic texture and local non-periodic defects in weld images, thereby

incorporating domain knowledge as a strong inductive bias into the generation process. The significantly reduced FID score and improved Recall indicate that the synthetic samples not only achieve high fidelity but, more critically, cover a broader manifold of real defect data. This demonstrates that for industrial objects with strong regularity, the generative model's understanding of underlying physical semantics is key to effective data augmentation. Second, the improvements in the YOLO11n-WAFE detector precisely address the inherent weaknesses of lightweight models against complex textured backgrounds. The ADownECA module mitigates the loss of subtle features caused by aggressive downsampling through its multi-path structure and channel attention mechanism, while the EASDF module effectively distinguishes periodic backgrounds from defect features via PDE-based edge enhancement and dynamic information fusion, overcoming the performance limitations of standard FPN in such scenarios. Working synergistically, these components deliver substantial improvements in mAP@0.5 and F1-score without significantly increasing computational overhead, demonstrating the effectiveness of optimizing the feature pyramid network for specific application scenarios.

Nevertheless, this study has certain limitations. Beyond the aforementioned advantages, the StyleGAN2-AFMS model also introduces specific trade-offs. The frequency-space dual-stream architecture, while crucial for capturing periodic textures, increases the model's training complexity and computational overhead compared to standard generators or augmentation-focused variants like StyleGAN2-ADA. Furthermore, its design is particularly optimized for structured scenes with strong periodic patterns; the performance gains may be less pronounced on datasets dominated by non-periodic or highly irregular textures. In addition to these methodological considerations, the primary constraint lies in the SWLWD dataset's origin from a single production line under controlled conditions, where its limited scale and diversity are insufficient for comprehensively validating the model's cross-domain generalization capability when facing different manufacturers, lighting conditions, and viewing angles. Additionally, as this study relies exclusively on visual surface images, the lack of metallographic cross-sectional data limits our analysis to surface manifestations, precluding investigation into the subsurface morphology and material genesis of the defects. This represents a valuable avenue for future interdisciplinary work.

Future work will focus on three directions: enhancing generalization through cross-factory data collection and unsupervised domain adaptation techniques; exploring conditional generation frameworks to achieve fine-grained controllable synthesis of defect attributes; and optimizing deployment efficiency via model pruning and quantization for comprehensive evaluation on embedded platforms.

5. Conclusions

This study effectively addresses the challenges of few-shot surface defect detection in sinusoidal wobble laser welds by proposing and validating an integrated framework that combines a physically informed data augmentation model with a lightweight, high-precision detector.

The main findings are as follows: The StyleGAN2-AFMS generative model achieved a Fréchet Inception Distance (FID) of 15.60 and a Recall of 0.470 on the SWLWD Dataset, demonstrating its ability to generate high-fidelity and diverse defect samples that expand the limited training data manifold. The YOLO11n-WAFE detector attained a mean Average Precision (mAP@0.5) of 94.2% and an F1-score of 89.5%, while maintaining a lightweight profile of 2.15 million parameters and a real-time inference speed of 656 frames per second (FPS).

In conclusion, this work provides a quantitatively validated, practical solution for automated weld defect inspection under few-shot conditions. By co-designing data generation

and detection, the framework achieves high-precision, real-time, and deployable inspection, offering a directly applicable pathway for quality control in precision manufacturing.

Author Contributions: Conceptualization, G.M. and J.Z.; methodology, G.M.; validation, J.Z. and J.J.; investigation, J.Z.; resources, J.J.; data curation, G.M. and J.J.; writing—original draft preparation, G.M. and J.Z.; writing—review and editing, G.M., J.Z. and J.J.; visualization, J.J.; supervision, J.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data utilized in this study were obtained from Celxpert (Kunshan) Electronics Co., Ltd. Data are available from the corresponding author upon reasonable request and with permission from Celxpert (Kunshan) Electronics Co., Ltd.

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

Abbreviations

The following abbreviations are used in this manuscript:

AFMS	Attentive Fourier-Modulation Synthesis
SWLWD Dataset	Sinusoidal Wobble Laser Weld Defect Dataset
FID	Fréchet Inception Distance
mAP	Mean Average Precision
FPS	Frames Per Second
ECA	Efficient Channel Attention
PDE	Partial Differential Equation
EASDF	Edge-Aware Semantic–Detail Fusion Module
ADownECA	A downsampling module integrating ADown and ECA
ADA	Adaptive Discriminator Augmentation
F, EP, DV, PD	Fracture, Excessive Penetration, Deposition Variation, Pad Defect
F_{mod},F_{freq}	Modulated feature map and frequency-domain feature map in StyleGAN2-AFMS
γ	Channel-wise modulation scaling factor
w_{mod},w_{freq}	Adaptive fusion weights for modulation streams and frequency
τ	Learnable evolution step size in the PDE-based edge enhancement
∇	Gradient operator

References

1. Rahim, K.; Mian, A. A review on laser processing in electronic and MEMS packaging. *J. Electron. Packag.* **2017**, *139*, 030801. [CrossRef]
2. Kashaev, N.; Ventzke, V.; Çam, G. Prospects of laser beam welding and friction stir welding processes for aluminum airframe structural applications. *J. Manuf. Process.* **2018**, *36*, 571–600. [CrossRef]
3. Stenberg, T.; Barsoum, Z.; Åstrand, E.; Öberg, A.E.; Schneider, C.; Hedegård, J. Quality control and assurance in fabrication of welded structures subjected to fatigue loading. *Weld. World* **2017**, *61*, 1003–1015. [CrossRef]
4. Sun, Y.; Xie, Y.; Peng, R.; Zhang, Y.; Chen, W.; Guo, Y. DEIM-SFA: A Multi-Module Enhanced Model for Accurate Detection of Weld Surface Defects. *Sensors* **2025**, *25*, 6314. [CrossRef]
5. Stavridis, J.; Papacharalampopoulos, A.; Stavropoulos, P. Quality assessment in laser welding: A critical review. *Int. J. Adv. Manuf. Technol.* **2018**, *94*, 1825–1847. [CrossRef]
6. Zheng, X.; Zheng, S.; Kong, Y.; Chen, J. Recent advances in surface defect inspection of industrial products using deep learning techniques. *Int. J. Adv. Manuf. Technol.* **2021**, *113*, 35–58. [CrossRef]
7. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [CrossRef]
8. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 779–788. [CrossRef]

9. Li, L.; Jiang, Z.; Li, Y. Surface Defect Detection Algorithm of Aluminum Based on Improved Faster RCNN. In Proceedings of the 2021 IEEE 9th International Conference on Information, Communication and Networks (ICICN), Xi'an, China, 25–28 November 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 527–531. [CrossRef]
10. Oh, S.-J.; Jung, M.-J.; Lim, C.; Shin, S.-C. Automatic Detection of Welding Defects Using Faster R-CNN. *Appl. Sci.* **2020**, *10*, 8629. [CrossRef]
11. Hu, T.; Huang, X.; Yang, Z.; Liu, Z.; Zhao, J.; Xu, Z. Vision-based welding quality detection of steel bridge components in complex construction environments. *Urban Lifeline* **2025**, *3*, 5. [CrossRef]
12. Zhu, S.; Zhang, L.; Zhao, J.; Hu, X.; Luo, X. Research on steel structure weld seam recognition algorithm based on improved YOLOv5. *Sci. Rep.* **2025**, *15*, 34276. [CrossRef]
13. Yang, Y.; Zhou, Y.; Din, N.U.; Li, J.; He, Y.; Zhang, L. An Improved YOLOv5 Model for Detecting Laser Welding Defects of Lithium Battery Pole. *Appl. Sci.* **2023**, *13*, 2402. [CrossRef]
14. Liu, L.; Zhang, M.; Zhou, X.; Zhai, R.; Wang, D.; Zhu, S. A weld defect detection algorithm designed based on YOLOv9. *Measurement* **2026**, *257*, 118630. [CrossRef]
15. Wang, Y.; Shang, L.; Li, B.; Liu, Y.; Ji, Y.; Hao, L.; Zhang, Y.; Li, Y.; Tian, M. Artificial-Weld-Crack Detection Network, YOLOv6-NW, Based on Target Recognition Technology. *Materials* **2024**, *17*, 6102. [CrossRef] [PubMed]
16. Xu, X.; Hou, W.; Li, X. Detection method of small size defects on pipeline weld surface based on improved YOLOv7. *PLoS ONE* **2024**, *19*, e0313348. [CrossRef]
17. Wang, J.; Li, J.; Wei, S. Surface Defect Detection of Transparent Parts Based on Improved YOLOv4 Model. In Proceedings of the 2022 5th International Conference on Machine Learning and Machine Intelligence, Hangzhou, China, 23–25 September 2022; Association for Computing Machinery: New York, NY, USA, 2023.
18. Çelik, M.T.; Arslankaya, S.; Yildiz, A. Real-time detection of plastic part surface defects using deep learning- based object detection model. *Measurement* **2024**, *235*, 114975. [CrossRef]
19. Cengil, E. Weld Defect Detection with YOLOv10. *Naturengs* **2024**, *5*, 77–81.
20. Hütten, N.; Alves Gomes, M.; Hölken, F.; Andricevic, K.; Meyes, R.; Meisen, T. Deep Learning for Automated Visual Inspection in Manufacturing and Maintenance: A Survey of Open- Access Papers. *Appl. Syst. Innov.* **2024**, *7*, 11. [CrossRef]
21. Shorten, C.; Khoshgoftaar, T.M. A survey on Image Data Augmentation for Deep Learning. *J. Big Data* **2019**, *6*, 60. [CrossRef]
22. Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. In *Advances in Neural Information Processing Systems 27 (NIPS 2014), Montreal, QC, Canada, 8–13 December 2014*; Curran Associates, Inc.: Red Hook, NY, USA, 2014; pp. 2672–2680. [CrossRef]
23. Wang, Y.; Hu, W.; Wen, L.; Gao, L. A New Foreground-Perception Cycle-Consistent Adversarial Network for Surface Defect Detection With Limited High-Noise Samples. *IEEE Trans. Ind. Inform.* **2023**, *19*, 11742–11751. [CrossRef]
24. Huang, M.; Sheng, C.; Rao, X. Conditional Variational Autoencoder and generative adversarial network-based approach for long-tailed fault diagnosis for the motor system. *Measurement* **2025**, *242*, 116116. [CrossRef]
25. Niu, S.; Li, B.; Wang, X.; He, S.; Peng, Y. Defect attention template generation cycleGAN for weakly supervised surface defect segmentation. *Pattern Recognit.* **2022**, *123*, 108396. [CrossRef]
26. Xie, X.; Li, C.; Qing, R.; Zhou, C.; Zhang, Z. High-quality matched transfer generation adversarial network for synthetic cross-material surface defect images. *Digit. Signal Process.* **2024**, *148*, 104441. [CrossRef]
27. Liang, B.; Sun, Z.; Yang, Z.; Tao, W.; Huo, C.; Xiao, S. Data augmentation and optimization method based on conditional generative adversarial network and convolutional neural network for transformer fault diagnosis. *Measurement* **2025**, *254*, 117872. [CrossRef]
28. Karras, T.; Aittala, M.; Hellsten, J.; Laine, S.; Lehtinen, J.; Aila, T. Training generative adversarial networks with limited data. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 12104–12114.
29. Everingham, M.; Van Gool, L.; Williams, C.K.I.; Winn, J.; Zisserman, A. The Pascal Visual Object Classes (VOC) Challenge. *Int. J. Comput. Vis.* **2010**, *88*, 303–338. [CrossRef]
30. Heusel, M.; Ramsauer, H.; Unterthiner, T.; Nessler, B.; Hochreiter, S. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In *Advances in Neural Information Processing Systems 30 (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017*; Curran Associates, Inc.: Red Hook, NY, USA, 2017; pp. 6629–6640. [CrossRef]
31. Kynkäänniemi, T.; Karras, T.; Laine, S.; Lehtinen, J.; Aila, T. Improved precision and recall metric for assessing generative models. In *Advances in Neural Information Processing Systems 32 (NeurIPS 2019), Vancouver, BC, Canada, 8–14 December 2019*; Curran Associates, Inc.: Red Hook, NY, USA, 2019; pp. 3927–3936. [CrossRef]
32. Triantafyllidis, G.A.; Tzovaras, D.; Sampson, D.; Strintzis, M.G. Combined Frequency and Spatial Domain Algorithm for the Removal of Blocking Artifacts. *EURASIP J. Adv. Signal Process.* **2002**, *2002*, 971438. [CrossRef]
33. Zhang, X.; Yang, W.; Hu, Y.; Liu, J. DMCNN: Dual-domain multi-scale convolutional neural network for compression artifacts removal. In Proceedings of the 2018 25th IEEE International Conference on Image Processing (ICIP), Athens, Greece, 7–10 October 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 390–394.

34. Tancik, M.; Srinivasan, P.; Mildenhall, B.; Fridovich-Keil, S.; Raghavan, N.; Singhal, U.; Ramamoorthi, R.; Barron, J.; Ng, R. Fourier features let networks learn high frequency functions in low dimensional domains. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 7537–7547.
35. Khanam, R.; Hussain, M. Yolov11: An overview of the key architectural enhancements. *arXiv* **2024**, arXiv:2410.17725. [CrossRef]
36. Wang, C.-Y.; Yeh, I.-H.; Mark Liao, H.-Y. Yolov9: Learning what you want to learn using programmable gradient information. In *European Conference on Computer Vision*; Springer: Cham, Switzerland, 2024; pp. 1–21.
37. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. ECA-Net: Efficient channel attention for deep convolutional neural networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 11534–11542.
38. Perona, P.; Malik, J. Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. Pattern Anal. Mach. Intell.* **1990**, *12*, 629–639. [CrossRef]
39. Liu, D.; Kababji, S.E.; Mitsakakis, N.; Pilgram, L.; Walters, T.; Clemons, M.; Pond, G.; El-Hussuna, A.; Emam, K.E. Synthetic Data Generation for Augmenting Small Samples. *arXiv* **2025**, arXiv:2501.18741. [CrossRef]
40. Seddik, M.E.A.; Chen, S.-W.; Hayou, S.; Youssef, P.; Debbah, M. How bad is training on synthetic data? a statistical analysis of language model collapse. *arXiv* **2024**, arXiv:2404.05090. [CrossRef]

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.