Computers, Materials & Continua

Tech Science Press

https://doi.org/10.32604/cmc.2026.082622

ARTICLE

Multi-Scale Supervised Dual-Layer Generative Adversarial Network: A Method for Region Restoration of LCM Images Degraded by Exposure Issues

Longhu Huang and Sheng Zheng*

School of Electrical and Information Engineering, Wuhan Institute of Technology, Wuhan, China

*Corresponding Author: Sheng Zheng. Email: 08090401@wit.edu.cn

Received: 19 March 2026; Accepted: 22 May 2026; Published: 23 July 2026

ABSTRACT: In the field of online automated defect inspection for small-size liquid crystal display modules (LCMs), the accuracy of module loading is crucial for the subsequent lighting inspection. However, due to the physical characteristics of the module's flexible ribbon cable, the ribbon often exhibits varying degrees of curling, causing conventional monocular vision systems to frequently encounter local underexposure or overexposure when positioning the workpiece, resulting in loss of local details and significantly affecting subsequent positioning and loading. To address the problem of local image degradation caused by abnormal exposure, this study proposes a regional image generation method based on a dual-layer generative adversarial network (GAN) with multi-scale supervision. This method first uses a mask localization module to restrict the region for image generation, then employs multi-scale local generation and adversarial learning to produce high-fidelity images in areas with local exposure anomalies, and finally uses a newly added global discriminator to regulate the edges of the generated images, allowing the generated images to smoothly connect with the original images, thereby achieving local image repair. Compared to single-layer GAN models, the repaired overall image achieved a peak signal-to-noise ratio (PSNR) improvement of 3.01% and a structural similarity (SSIM) improvement of 11.7%, while the PSNR of the repaired images in exposure-anomalous regions increased by 49.04% and SSIM increased by 23.18%. In addition, not only does the model produce restored images with excellent visual effects, but through verification by deployment on an actual factory production line, the error between the restored images and normal samples is within 0.08 mm, validating the model's value in practical industrial applications.

KEYWORDS: Image restoration; multi-scale supervision; GAN; LCM

1 Introduction

With the rapid iteration of consumer electronics, small liquid crystal display modules (LCMs) have been widely used in wearable devices as core display components, such as smart watches, smart glasses, and health monitoring devices [1]. Due to the extremely complex process of LCM manufacturing, it is inevitable that modules with defects will be produced during production, such as surface scratches, point defects, line defects, leakage defects, etc., [2], so defect detection of the output modules is an indispensable step in the entire production process. In the context of "black light factories" becoming the transformation goal of the manufacturing industry, LCM production lines are accelerating towards full-process automation, online automated defect detection has become a rigid demand for LCM manufacturing, and vision-guided robotic arm feeding system is one of the keys to achieving flexible, accurate, and high-speed loading of production lines [3]. The LCM online automated inspection system is shown in Fig. 1, and the loading part is composed of an industrial camera, a robotic arm, a conveyor belt and a unloading area.

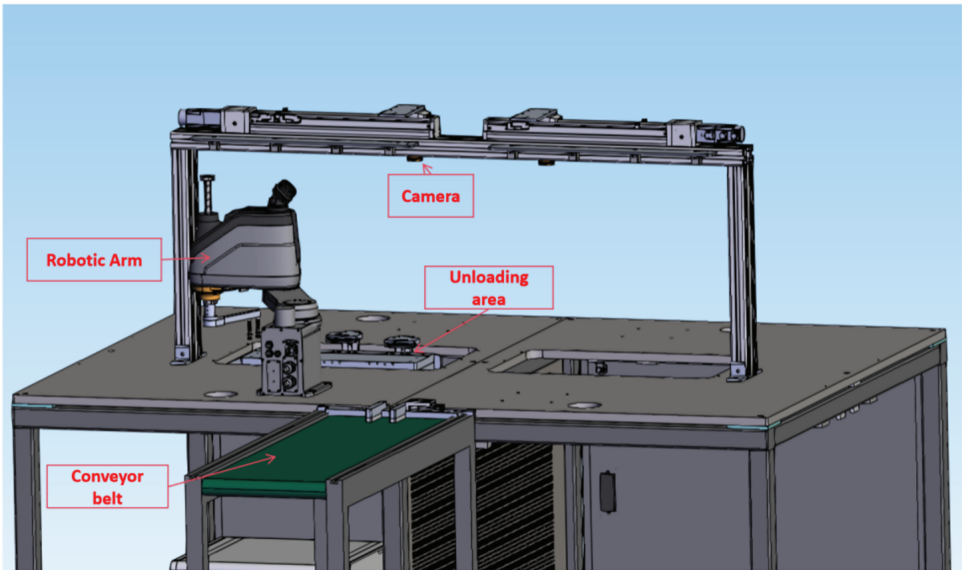


Figure 1: 3D diagram of the loading section of the LCM online automation inspection system.

However, for online automated inspection systems, the diversification of LCM geometries is a key factor affecting production line flexibility. As a highly integrated optoelectronic component, the core of an LCM consists of a liquid crystal panel, backlight, driver IC, and flexible circuit board (FPC). Although the display area and overall dimensions vary significantly according to end-product requirements, the gold finger interface at the end of the FPC exhibits a high degree of consistency, as shown in Fig. 2. Therefore, the gold finger at the end of the FPC becomes a key feature for identification by the loading system.

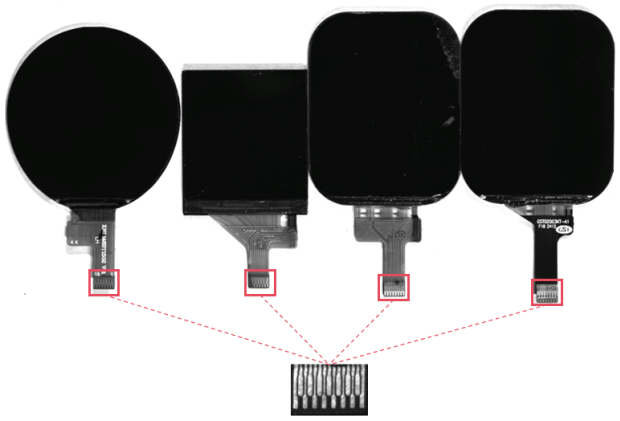


Figure 2: Images of four different LCMs and comparison of FPC ends.

However, FPCs are prone to curling and deformation, and their material has reflective properties. Under fixed ambient lighting, the contrast of images captured from FPCs with different curling conditions varies. In cases of severe exposure abnormalities, the contrast of local images can be significantly reduced, resulting in loss of texture details and posing obstacles to subsequent positioning algorithms, as shown in Fig. 3.

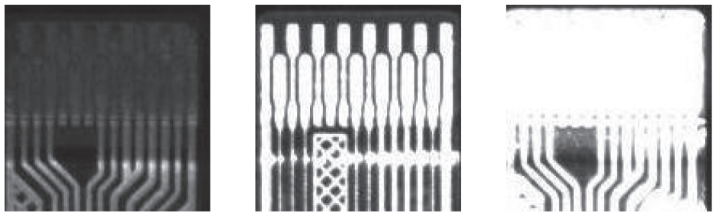


Figure 3: Comparison images of three exposure scenarios: (**a**) Underexposed. (**b**) Normal. (**c**) Overexposed.

To address this issue, this study proposes a dual-layer generative adversarial network (GAN) model based on multi-scale supervision. The model integrates a mask attention module, which obtains the location of the area to be repaired in the original image through a masked image, then generates the image in that area using a generator, and evaluates the generated image through both local and global discriminators. Multi-scale supervision is introduced in the generator and local discriminator, where generated images of different resolutions are supervised separately, which helps to learn coarse structure, texture, and fine details respectively, improving generation quality and training stability. This study makes the following contributions:

1. By conducting on-site visits to industrial production environments, we collected, cleaned, and screened data. This process enabled us to build a reliable dataset of LCM loading images, providing high-quality public resources for subsequent researchers.
2. To address the special requirement for strict geometric alignment between the region to be repaired and the background in LCM feeding scenarios, this study proposes a dual-layer collaborative generation framework that decouples local generation from global constraints through a hard fusion mechanism: the local generator focuses on content reconstruction of the masked region, while the global discriminator only evaluates the overall consistency after fusion, avoiding optimization conflicts between local realism and global coherence in the generator.
3. This study proposes a dynamic weight learning strategy for multi-scale generation to address the training instability caused by static weighting in existing multi-scale supervision methods. Existing methods typically apply fixed weights to each scale, causing the generator to face high-frequency detail optimization pressure from early training stages, which easily leads to mode collapse. This study designs a three-stage progressive weight scheduling: early stage prioritizes coarse-scale structure modeling, middle stage achieves balanced transition, and late stage focuses on fine-scale texture refinement. This approach allows the generator to gradually adapt from easy to difficult, significantly improving convergence stability.

2 Related Works

This section discusses topics related to this study, including traditional methods and deep learning-based methods in the field of recovering improperly exposed images, the development of GAN-based image restoration methods, and region-controllable image generation.

2.1 Image Restoration Method for Exposure Abnormalities

To address image degradation caused by exposure abnormalities, mainstream methods are divided into traditional image processing and deep learning-based approaches. Traditional image processing methods mainly rely on manually designed features and fixed mathematical transformations, offering advantages such

as high computational efficiency and strong interpretability. For example, Xiang et al. [4] used histogram-based equalization techniques, introducing quadruple histogram segmentation and clipping techniques to process the channels and brightness of underwater low-contrast and blurry images, increasing the brightness in dark areas while preserving the brightness in bright areas, and enhancing the edges and richness of the images; Deng and Xie [5] incorporated adaptive gamma correction into adaptive histograms, suppressing artifacts while maintaining overall image brightness. However, these methods rely on preset parameters or prior assumptions, making it difficult to adapt to complex lighting changes, and their ability to recover from extreme anomalies (such as reflective highlights) is limited. Deep learning-based methods learn an end-to-end mapping from degraded images to normally exposed images in a data-driven manner, achieving significant progress in recent years. For example, Liu et al. [6] proposed a learning-based method for optical degradation image restoration, which can adaptively recognize different types of non-ideal imaging conditions and learn to restore images accordingly; Chen et al. [7] constructed a dual-network cascade model, consisting of an exposure prediction network and an exposure fusion network, to recover details lost in underexposed or overexposed regions. However, these models focus on recovering images based on the causes of optical image degradation, involve a large number of parameters and high computational complexity, and require a large number of images with different exposures, making them limited in small-sample industrial scenarios. The physical characteristics of FPC determine the complexity of its image degradation, and the core task of the feeding system is not to pursue visual aesthetics for the human eye, but to ensure the precise extractability of finger edge contours and pin array features. This requires extremely high quality for restored images, with the pin error in the restored image needing to be within 0.08 mm, that is, the width of one pin.

2.2 Image Restoration Method Based on GAN Networks

In 2014, Goodfellow et al. [8] first proposed GAN, which is based on the adversarial game between the generator and the discriminator. The generator tries to produce realistic images, while the discriminator strives to distinguish between real and fake images, laying the foundation for subsequent image generation technologies. In 2019, Chen et al. [9] proposed GAN-RS, using a multi-branch discriminator to remove underwater noise while preserving image content. In 2022, Lin et al. [10] proposed DGAN, adding an additional loss function through structural similarity to enhance the clarity and color correction of underwater images. In 2023, Hu et al. [11] proposed DARE-GAN, using an unsupervised degradation representation learning strategy and introducing perceptual degradation feature interpolation, achieving excellent performance in facial restoration. In 2024, Chahi et al. [12] proposed MFGAN, introducing multiple CNN streams in the GAN generator and considering the strength of traditional layers at different scale levels through multi-kernel filtering. In 2025, Nguyen et al. [13] proposed UFR-GAN, using a lightweight and multi-degradation GAN to achieve state-of-the-art performance under various degraded scenarios. In the same year, Balaji [14] proposed PMPGAN, introducing a multi-dimensional generation network along with a multi-dimensional pyramid aggregation module and attention module, providing advantages and efficiency in eliminating image degradation and generating visually realistic fake images. The model in this article also starts from the GAN network, relying on the idea of generative and adversarial learning to gradually standardize the quality of generated regional images and achieve end-to-end image restoration without requiring prior knowledge of image degradation due to abnormal exposure.

2.3 Region-Based Image Generation Method

Region-controllable image generation refers to generating the desired image within a specified area of an image while ensuring the consistency of features at the connection between the area and the entire

image, thus avoiding image distortion caused by the random generation of image by the model. To this end, many researchers have increased constraints on generative models to improve the controllability of image generation in specific regions. Niu et al. [15] proposed a region- and intensity-controllable GAN, which controls the region of generated defects using defect masks and designs a defect attention loss to force the generative model to focus on the defective areas; Zhang et al. [16] proposed ControlNet, a technique that adds additional input condition controls to pre-trained large diffusion models, enhancing the model's precise control over spatial layouts in image generation tasks; Li et al. [17] proposed GLIGEN, which guides the generative model to generate specific objects in particular locations, controlling the content and layout of generated images based on entity concepts in the open world, achieving controllable generation capabilities for generating specific entities in specific positions or combining various layout control conditions; BarTal et al. [18] proposed MultiDiffusion, which achieves flexible control over the image generation process without additional training by integrating multiple diffusion paths, performing excellently in panoramic image generation and region-based text-to-image generation tasks. The region-controllable design in this study is similar to that proposed by Shuanlong in that it also introduces a mask module to restrict the image generation area; the difference is that the model in this study, through nesting the GAN model, designs an additional global discriminator to ensure the consistency of the generated image with the overall image. However, methods such as ControlNet and region-controllable GAN have limitations in industrial anomaly repair scenarios: ControlNet relies on additional conditional control networks, resulting in high computational overhead and requiring retraining; region-controllable GAN focuses on attribute editing rather than precise region inpainting. The region controllability designed in this study restricts the image generation region by introducing a mask module, and ensures the consistency between the generated image and the global context through nesting of the GAN model and designing an additional global discriminator.

3 Methodology

The multi-scale supervised dual-layer generative adversarial model proposed in this study is derived from the GAN model. As shown in Fig. 4, two modifications are made: (1) a new global discriminator D_2 is added after the initial discriminator D (hereafter referred to as D_1); and (2) a multi-scale supervision module is incorporated into both the generator and the local discriminator.

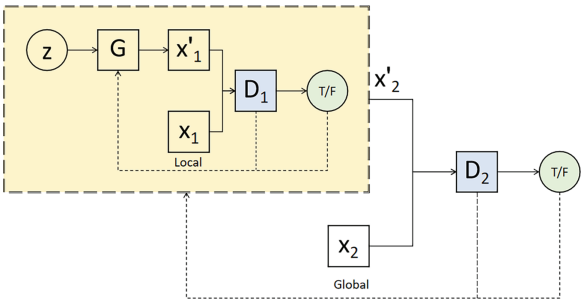


Figure 4: Architecture diagram of a dual-layer generative adversarial model with multi-scale supervision.

Specifically, the first layer maintains the same architecture as the basic GAN model. However, the generated fake data (x_1') and the input real data (x_1) are only partial regions of the target image, i.e., images within a specified region. D_1 only determines the authenticity of the content generated by G.

Subsequently, the image generated by G replaces the corresponding region of the real image. This replaced complete image then serves as the input for the generator of the next-layer GAN model. Meanwhile,

it is fed together with the real complete image into D_2 for discrimination, to determine whether the complete image after fusing the G-generated content appears realistic.

3.1 Multi-Scale Supervised Generation and Adversarial

The core idea of multi-scale supervised generation and adversarial networks is: to introduce a multi-scale supervision module into the generator G and the local discriminator D_1, allowing the generator to better learn local contextual relationships and improve repair quality. The process of multi-scale supervision is shown in Fig. 5. The model inputs include the original image (256 × 256) and the corresponding mask (256 × 256), both of which are fed into the multi-scale local generator G. The generator G adopts an encoder-decoder architecture: the encoder progressively downsamples the input to 16 × 16 and injects noise (z = 100), while the decoder progressively upsamples to restore to 128 × 128, preserving detailed information through skip connections. The generator is also equipped with two auxiliary heads (Aux Heads), which output auxiliary images of 32 × 32 and 64 × 64, respectively, for multi-scale supervision. The final output of the generator is a 128 × 128 generated image, which together with the two auxiliary outputs is fed into the multi-scale discriminator D_1. D_1 discriminates the generation quality at each scale through three scale paths (128 × 128, 64 × 64, 32 × 32), and obtains the final discrimination result through weighted fusion. Meanwhile, the generated image is embedded into the corresponding region of the original image through the fusion module, forming a complete 256 × 256 fused image. This fused image is then fed into the global discriminator D_2 together with the real complete image. D_2 discriminates whether the overall fused image is realistic and natural, thereby ensuring the consistency between the generated content and the global context.

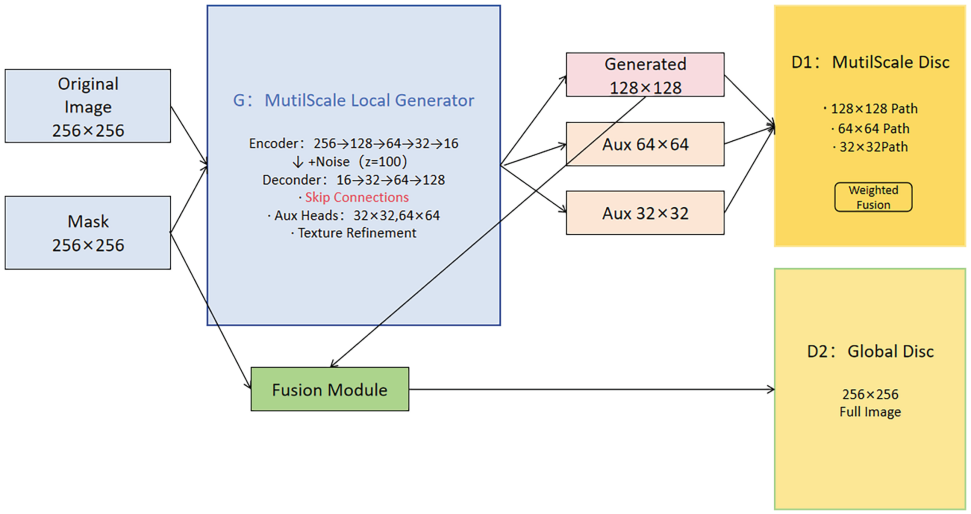


Figure 5: Flowchart of the multi-scale supervised dual-layer generative adversarial model.

The multi-scale generator uses a U-Net encoder-decoder structure and explicitly generates outputs at multiple scales during the decoding process. Specifically: in generator G, after inputting a 256 × 256 image along with a mask, the image is gradually downsampled to a 16 × 16 feature map. During the upsampling process, the decoder generates outputs at three scales: 32 × 32 (coarse structure), 64 × 64 (medium details), and 128 × 128 (fine details). Each scale has an independent output head, realized through head_32, head_64, and the final dec1, as shown in Fig. 6.

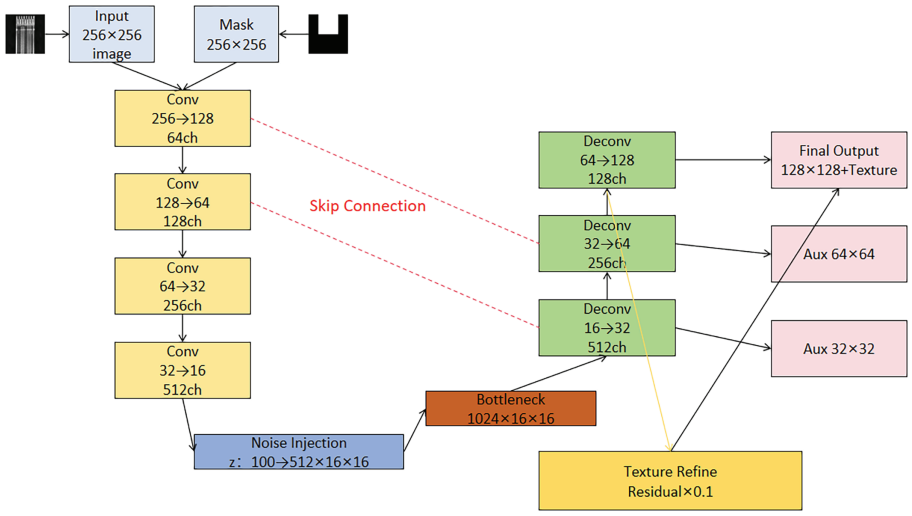


Figure 6: Multi-scale supervised generator structure diagram.

The multi-scale discriminator simultaneously handles inputs at three scales: D_32 processes 32 × 32 inputs, D_64 processes 64 × 64 inputs, and D_128 processes 128 × 128 inputs. Each scale has an independent discrimination path but shares underlying features, as shown in Fig. 7.

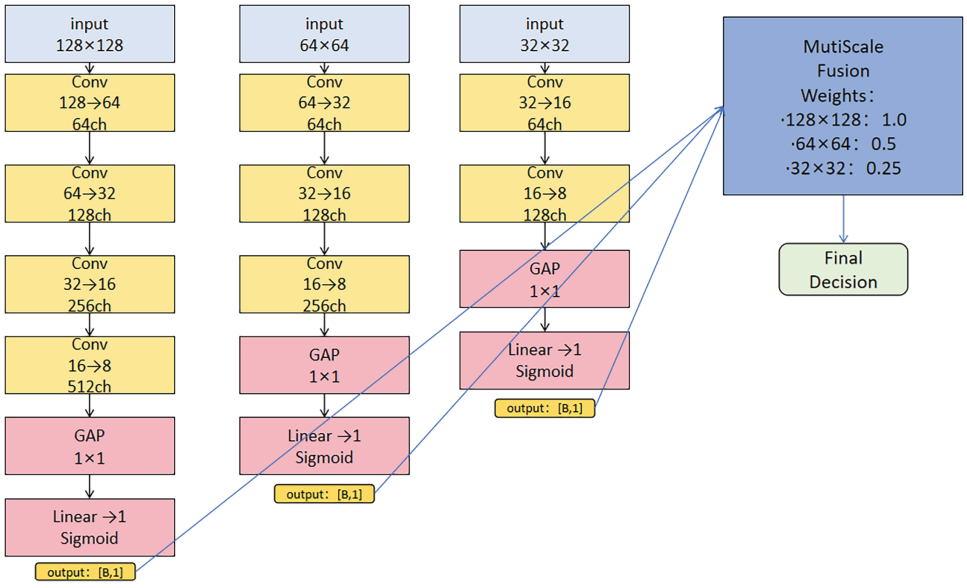


Figure 7: Diagram of the multi-scale supervised local discriminator structure.

During the training process, calculate the loss for the three scales separately:

1. Discriminator loss: Calculate the BCE loss of real samples and generated samples separately for each scale; use dynamic weights to balance the contributions of different scales.
2. Generator loss: Adversarial loss: make the generated samples fool the discriminators at three scales; Reconstruction loss: calculate the L_1 difference between the generated image and the real image at three scales.

The generator achieves feature fusion across different scales through skip connections. The advantage of multi-scale supervision lies in using explicit supervision over visual features at different levels, ensuring effective discrimination of real and fake samples at various scales, enhancing discriminative robustness, and making the training process more stable through progressive training. The generated results have better structural consistency and detail quality, while intermediate supervision also alleviates the gradient vanishing problem.

In order to achieve multi-scale optimization from coarse to fine, this study designs a dynamic weight adjustment mechanism based on training progress. Let e be the current training epoch and E be the total number of epochs. Define the normalized training progress $t = e/E \in [0,1]$. The multi-scale weight $w_s^{(t)}$ adopts a three-stage piecewise constant strategy:

$$w_s(t) = \begin{cases} w_s^{(1)}, & 0 \le t < 0.3 \\ w_s^{(2)}, & 0.3 \le t < 0.7 \\ w_s^{(3)}, & 0.7 \le t \le 1 \end{cases} \tag{1}$$

where $s \in \{32, 64, 128\}$ denotes the scale index. The weight parameters for each stage are shown in the following Table 1:

Table 1: Progressive weight parameters for different scales and stages.

Scale s	$w_s^{(1)}$	$w_s^{(2)}$	$w_s^{(3)}$
32 × 32	1.0	0.5	0.2
64 × 64	0.5	1.0	0.5
128 × 128	0.3	0.7	1.0

This progressive weight scheduling enables the generator to optimize gradually from easy to difficult, avoiding training instability caused by directly learning high-frequency details in the early stages. The training weight variation curves are shown in the Fig. 8.

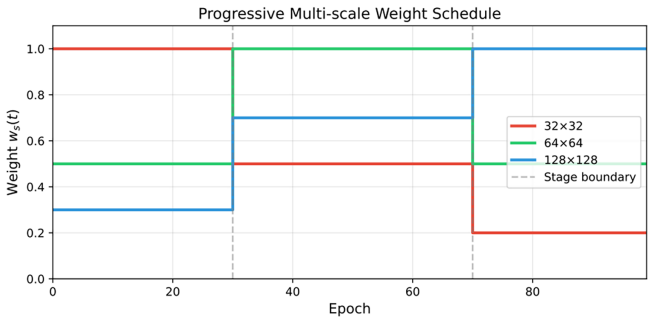


Figure 8: Schematic diagram of training weight variation.

3.2 Loss Function

The loss function of the multi-scale discriminator D_1 is as follows:

$$L_{D_1} = \frac{1}{2} \sum_{s \in \{32,64,128\}} w_s^{(t)} \cdot \left[L_{BCE}\left(D_s\left(x_s\right), 1\right) + L_{BCE}\left(D_s\left(G_s\left(z\right)\right), 0\right) \right] \tag{2}$$

where s represents the scale index, taking values of 32, 64, and 128; $w_s^{(t)}$ is the dynamic weight of scale s at training step t as defined in the previous section; L_{BCE} is the binary cross-entropy loss function; D_s is the branch of the discriminator at scale s; x_s is the local patch of the real image at scale s; $G_s(z)$ is the output of the generator at scale s; 1 is the true label, 0 is the false label, and z is the random noise vector.

D_2 Loss Function:

$$L_{D_2} = \frac{1}{2}\left[L_{BCE}\left(D_2\left(x_{global}\right),1\right) + L_{BCE}\left(D_2\left(Fusion\left(x, G\left(z\right), mask\right)\right),0\right)\right] \tag{3}$$

where x_{global} is the complete real image, $Fusion$ is the fusion module function, x is the original real image after masking, $G(z)$ is the 128 × 128 local patch output by the generator, $mask$ is the mask module, and $Fusion(x, G(z), mask)$ is the result of fusing the generated patch into the complete image.

Local adversarial loss:

$$L_G^{adv\_local} = \sum_{s\in\{32,64,128\}} w_s^{(t)} \cdot L_{BCE}\left(D_s\left(G_s\left(z\right)\right),1\right) \tag{4}$$

where $D_s(G_s(z))$ is the discriminator's evaluation of the generated samples, and 1 is the target label for true. Where $D_2()$ is the global discriminator's judgment on the fused fake image, where 1 represents the target being real, indicating that it is desired that no repair traces are visible globally.

Global Adversarial Loss:

$$L_G^{adv\_global} = L_{BCE}\left(D_2\left(Fusion\left(x, G\left(z\right), mask\right)\right),1\right) \tag{5}$$

Reconstruction loss (L_1):

$$L_G^{rec} = \sum_{s\in\{32,64,128\}} w_s^{(t)} \cdot \lambda_s \left\| G_s\left(z\right) - x_s^{real} \right\|_1 \tag{6}$$

Here, λ_s is the basic reconstruction coefficient for each scale ($\lambda_{128} = 10$, $\lambda_{64} = 5$, $\lambda_{32} = 2.5$), $||...||_1$ is the L_1 norm, $G_s(z)$ is the output of the generator at scale s, and x_s^{real} is the local patch of the real image at scale s. The L_1 loss is calculated as:

$$\left\| G_s\left(z\right) - x_s^{real} \right\|_1 = \sum_{i,j} \left| G_s\left(z\right)_{i,j} - x_{s,i,j}^{real} \right| \tag{7}$$

Generator total loss:

$$L_G = \left(L_G^{adv\_local} + L_G^{adv\_global}\right) + L_G^{rec} \tag{8}$$

where L_G^{adv} is the total adversarial loss, and L_G^{rec} is the total reconstruction loss. The global adversarial loss $L_G^{adv\_global}$ and the global discriminator loss L_{D_2} do not change with the training progress and always maintain a constant weight of 1.0. This design ensures the stability of the global consistency constraint, while the dynamic weight $w_s^{(t)}$ only applies to multi-scale local supervision, enabling progressive optimization of local detail generation.

3.3 Parameter Design

To meet the training requirements of model, the following designs were made for images and data parameters: the resolution of input and output images is 256 × 256 to ensure basic image processing tasks; the

generator's repair region size is 128 to achieve local image restoration; the random noise vector dimension is 100, balancing the diversity and stability of model training; the number of input channels of the model is 2 to ensure a one-to-one correspondence between the original image and the mask image, and the number of output channels is 1, so that the model finally outputs a complete image.

Model hyperparameter design: The total amount of data is 1500, the total number of training epochs is 100, the batch size is 16, with 70% for the training set, 15% for the test set, and 15% for the validation set; the learning rates for the generator and discriminator are 0.0001, making the model training more robust and preventing the model from not converging; the Adam optimizer's β_1 and β_2 are 0.5 and 0.999, respectively, allowing the generator to more flexibly cope with updates from the discriminator while providing long-term stable gradient scale estimation.

The total number of parameters of the model is 47.9M (approximately 192 MB of storage), with a memory requirement of about 2.3 GB. Training on an RTX 3090 takes about 5 h per 100 epochs.

4 Experiment and Analysis

This section introduces the datasets used for experiments, compares the performance of currently popular methods for recovering exposure-abnormal images, and demonstrates the effectiveness of the modules in our proposed network through ablation studies.

4.1 Datasets and Evaluation Metrics

This article mainly discusses the issue of image degradation caused by uneven lighting in the machine vision part of the automated feeding process in the field of LCM defect detection. The data comes from actual automated feeding equipment on production lines and is captured using a high-resolution (5472 × 3648) industrial camera. Since images with too high a resolution significantly increase the burden on model training and cause resource waste, it is necessary to crop the original images. First, the target object (i.e., the LCM) is cropped from the original image. Then, a 256 × 256 image containing the feature area identified by the feeding positioning algorithm is cropped from this image. Next, a 128 × 128 image of the feature area is cropped from that image. The final obtained image is the image that the generator in the first layer of the model is expected to learn to generate. The final 256 × 256 and 128 × 128 images are chosen as the model input, with moderate pixel size, a low training burden, and the ability to obtain training results in a relatively short time.

The above dataset preparation process is only a reference for a single image. In fact, due to the production process of LCMs and the characteristics of FPC circuits, the position of the feature regions in the original image is not fixed, so fixed pixel coordinate areas cannot be used to crop and divide them. To solve this problem, during dataset preparation, we first use an edge gradient-based template matching method to perform step-by-step matching and cropping on the original high-resolution images in a hierarchical manner. This method uses edge gradient features as the basis for matching:

$$G(x, y) = \sqrt{G_x^2 + G_y^2}, \quad \theta = \arctan\left(\frac{G_y}{G_x}\right) \tag{9}$$

where G_x and G_y are the gradient components computed by the improved Sobel operator, and θ is the gradient direction. A multi-scale pyramid architecture is employed, where the velocity scale controls the number of pyramid levels, and the feature scale controls the sparsity of edge points. Finally, the Normalized Cross-Correlation (NCC) is adopted:

$$R = \frac{\sum_{u,v}(G_I - \overline{G}_I)(G_T - \overline{G}_T)}{\sqrt{\sum(G_I - \overline{G}_I)^2 \sum(G_T - \overline{G}_T)^2}} \tag{10}$$

where R is the matching score, which is set to 0.5 in this study. In the LCM image dataset with rounded-rectangle straight-arranged wire type, the overall workpiece contour including the module and wires is first set as the primary matching template to identify the workpiece position in the high-resolution image. Taking the center of the matched rectangular box as the reference point, the slender front part of the wires is then used as the secondary matching template to match the entire wire portion. With the midpoint of the wire top as the reference, the image is expanded by 128 pixels to the left and right, and 256 pixels downward. The expanded image is then cropped to obtain a 256 × 256 sized image. Finally, the wire-end gold fingers are used as the ultimate matching template to locate the gold finger position. With the midpoint of the gold finger top as the reference, the image is expanded by 64 pixels to the left and right, and 128 pixels downward. The expanded image is cropped to obtain a 128 × 128 sized image. Due to the complex textures in the original high-resolution images, to avoid matching unintended regions with only a single-level matching, each level of matching uses the center of the rectangular box matched at the previous level as the reference. By setting the matching region, the precise matching at the next level is completed, as shown in Fig. 9. The pixel error at each level is within 10 pixels, ensuring that the final cropped region completely contains the feature region of the image. At the same time, for the final cropped 128 × 128 image, a mask image is created based on the size of the previous-level image (256 × 256). The mask image ensures a one-to-one correspondence between the original image and the feature region image. The purpose of creating the mask image is to guide the generator in learning the content to be generated, preventing it from learning unintended content, and to inform the generator of the location of this content in the original image. Through D_2, the fusion effect between the generated image and the original image is regulated, achieving region-controllable image restoration.

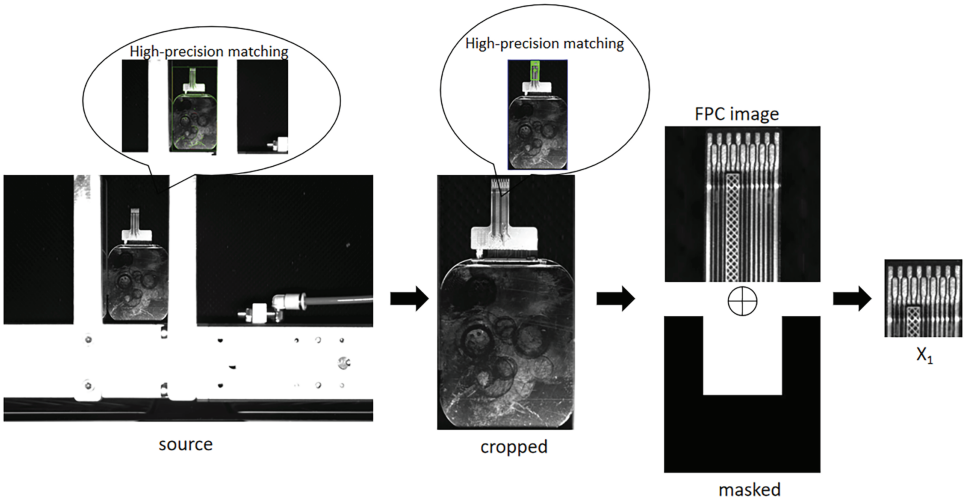


Figure 9: Flowchart of the multi-scale supervised two-level generative adversarial model.

Through data acquisition on the factory feeding equipment, a total of 1500 different production line images of rounded-rectangle straight-arranged wire type LCMs were collected. According to the dataset production process, they were divided into original image sets and mask image sets. These LCMs were continuously produced by the production line within a specific time period, and each module has an independent ID number to ensure that no duplicate data exists during image organization. The datasets were

divided according to the creation process into a set of original images and a set of mask images. These two datasets were randomly divided into training set, test set, and validation set in proportions of 70%, 15%, and 15%, ensuring that each original image corresponded to a mask image. The number of images in each paired dataset after division was 1125, 225, and 225 respectively.

This study employs three metrics to analyze model performance: Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), and Learned Perceptual Image Patch Similarity (LPIPS). PSNR is an objective metric used to measure the difference between two images [19]. A higher PSNR value indicates that the two images are more similar and that the quality loss is smaller. The core idea is to measure the degree of distortion by calculating the mean squared error (MSE) between the original image and the distorted image, and then quantify it through the ratio of the maximum signal power to the noise power. MSE is the average of the differences in pixel values between the two images, and its calculation formula is:

$$\mathrm{MSE} = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\left[I_1(i,j) - I_2(i,j)\right]^2 \tag{11}$$

where I_1 and I_2 are two images, M and N are the height and width of the images, respectively, and i and j are the pixel position indices. With MSE, PSNR can be calculated using the following formula:

$$\mathrm{PSNR} = 10 \cdot \log_{10}\left(\frac{\mathrm{MAX}^2}{\mathrm{MSE}}\right) \tag{12}$$

where MAX is the maximum possible pixel value in the image.

SSIM is a perceptual model based on the human visual system (HVS) used to measure the similarity between two images in terms of luminance, contrast, and structure [20]. Compared to PSNR, SSIM is more aligned with the perception of the HVS and can more accurately reflect image quality. Its core idea is to treat an image as a collection of luminance, contrast, and structure, and evaluate overall similarity by comparing these three aspects. The complete formula for SSIM is:

$$\mathrm{SSIM}(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{13}$$

where μ_x and μ_y are the average values of images x and y within a local window, representing the brightness levels of the images; ${\sigma_x}^2$ and ${\sigma_y}^2$ are the variances of images x and y within the local window, representing the contrast of the images; σ_{xy} is the covariance of images x and y within the local window, representing the structural similarity of the images.

To verify the effectiveness of the algorithm proposed in this study, we established an experimental platform to ensure that all tests are conducted under consistent conditions. We use Windows 10 as the operating system, PyTorch as the deep learning framework, and GAN as the benchmark network model. The experimental configuration includes an RTX 3090 GPU (24 GB), and programming is done using Python 3.10.

4.2 Comparative Experiment

To comprehensively evaluate the performance of the multi-scale supervised dual-layer adversarial generative network proposed in this study for repairing exposure anomaly images, a hierarchical comparison strategy is adopted in the comparative experiments. First, comparisons are made with traditional non-learning methods, followed by comparisons with modern inpainting models. All experimental data are sourced from 1500 images collected from the factory. Since the modules on the production line do not pass

through repeatedly, it is impossible to obtain real abnormal and normal sample pairs. Therefore, this study uses synthetic exposure anomalies to simulate different exposure conditions on the production line. Linear contrast stretching is employed to adjust the contrast of local images, with the contrast factor α varying within the range of $[0.1, 3.0]$.

1. Traditional histogram equalization (HE) and non-learning exposure fusion network (EF) are selected for comparison experiments. For HE, we adopt adaptive equalization, optimizing parameters on the validation set through grid search to ensure fairness of comparison. The processing range is explicitly specified, similar to the masked image, where a grid range is designated on the original image for equalization. For EF, simulated multi-exposure images are input, and fusion weights are optimized through grid search on the 1500 images to ensure its best performance on this dataset. The average PSNR, average SSIM, and average LPIPS between the repaired images and the original images for the overall and repaired regions of each method are shown in Tables 2 and 3.

Table 2: Comparison of PSNR, SSIM, and LPIPS scores for images restored by three methods on the overall image.

	HE			Exposure Fusion			OURS		
	PSNR	**SSIM**	**LPIPS**	**PSNR**	**SSIM**	**LPIPS**	**PSNR**	**SSIM**	**LPIPS**
TOTAL	33.32	0.92	0.11	32.90	0.92	0.10	35.03	0.95	0.05
Overexposed	33.67	0.86	0.18	33.44	0.91	0.12	35.18	0.95	0.04
Underexposed	32.97	0.97	0.05	32.36	0.92	0.08	34.88	0.95	0.06

Table 3: Comparison of PSNR, SSIM, and LPIPS scores for images restored by three methods on local regions.

	HE			Exposure Fusion			OURS		
	PSNR	**SSIM**	**LPIPS**	**PSNR**	**SSIM**	**LPIPS**	**PSNR**	**SSIM**	**LPIPS**
TOTAL	28.94	0.71	0.24	28.36	0.73	0.19	29.14	0.84	0.16
Overexposed	28.17	0.51	0.41	29.34	0.72	0.17	29.29	0.82	0.15
Underexposed	29.70	0.91	0.07	27.38	0.75	0.11	28.99	0.85	0.16

Table 2 presents the comparison between the overall restored images from the three methods and the original overall images. Since we focus on regional image restoration, and this region only occupies one-fourth of the original image, the PSNR, SSIM, and LPIPS scores of all three models on the overall image are relatively high. Among them, the nested-GAN proposed in this study achieves the highest overall scores. Although HE obtains the lowest LPIPS score under underexposure conditions, it is overall inferior to OURS.

Table 3 presents the comparison between the locally restored exposure anomaly regions from the three methods and the corresponding regions in the original images. EF repairs exposure anomaly regions by fusing multiple images with different exposures. Although it achieves the lowest PSNR scores for both overexposed and underexposed image restoration, its overall SSIM score is higher than HE, while its LPIPS value is lower than HE. This method demonstrates relatively stable performance in image restoration without particularly outstanding aspects. HE performs best in repairing underexposed images, because texture details still exist in underexposed images, merely with low contrast, resulting in a visually blurry appearance. After histogram equalization, the contrast is enhanced and texture clarity is improved, making the quality closest to the original image. However, for overexposed images, this method yields the worst restoration effect, because overexposed images have lost texture details, and reducing contrast cannot recover these details.

Although the proposed model is inferior to histogram equalization in repairing overexposed images, it achieves the highest overall PSNR and SSIM scores and the lowest LPIPS score, with values of 29.14, 0.84, and 0.16, respectively. Since this method is not based on optical principles or image processing principles for restoration, but rather on an image generation approach, its performance on overexposed and underexposed images is highly stable.

2. The LaMa and MAT models are selected for performance comparison with the proposed model. LaMa [21], proposed in 2021, is an image inpainting method based on Fast Fourier Convolution (FFC), which achieves excellent performance in regular mask inpainting tasks by capturing global context information in large receptive fields. MAT [22], proposed in 2022, is the first Transformer-based inpainting method that directly processes high-resolution images, effectively modeling global context through a mask-aware mechanism. For both methods, the same region mask definition as the proposed method is adopted. They are retrained on the same dataset with identical dataset splitting and hyperparameter search strategies. The average PSNR, average SSIM, and average LPIPS between the repaired images and the original images for the overall and repaired regions of each method are shown in Tables 4 and 5.

Table 4: Comparison of PSNR, SSIM, and LPIPS scores for images restored by three methods on the overall image.

	LaMa			MAT			OURS		
	PSNR	**SSIM**	**LPIPS**	**PSNR**	**SSIM**	**LPIPS**	**PSNR**	**SSIM**	**LPIPS**
TOTAL	34.25	0.92	0.07	34.68	0.94	0.06	35.03	0.95	0.05
Overexposed	34.58	0.93	0.06	35.02	0.94	0.05	35.18	0.95	0.04
Underexposed	33.92	0.95	0.08	33.34	0.95	0.07	34.88	0.95	0.06

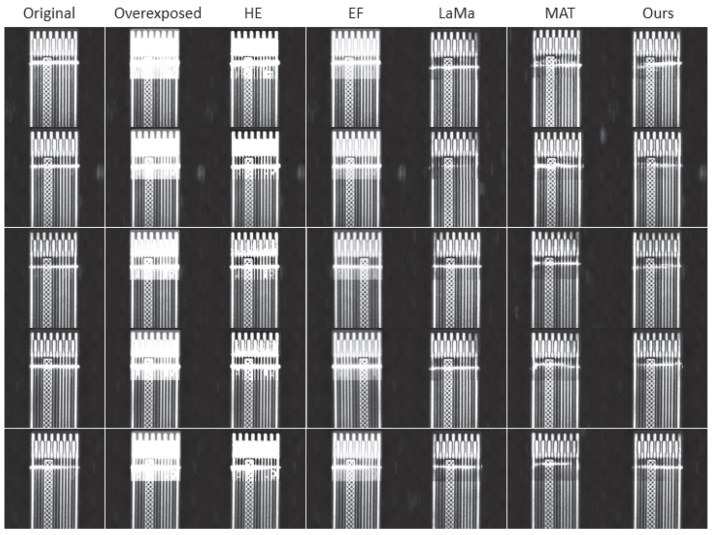
Table 5: Comparison of PSNR, SSIM, and LPIPS scores for images restored by three methods on local regions.

	LaMa			MAT			OURS		
	PSNR	**SSIM**	**LPIPS**	**PSNR**	**SSIM**	**LPIPS**	**PSNR**	**SSIM**	**LPIPS**
TOTAL	28.85	0.79	0.14	28.95	0.81	0.13	29.14	0.84	0.16
Overexposed	28.92	0.76	0.15	29.08	0.79	0.12	28.99	0.82	0.15
Underexposed	28.78	0.82	0.13	28.82	0.83	0.14	28.99	0.85	0.16

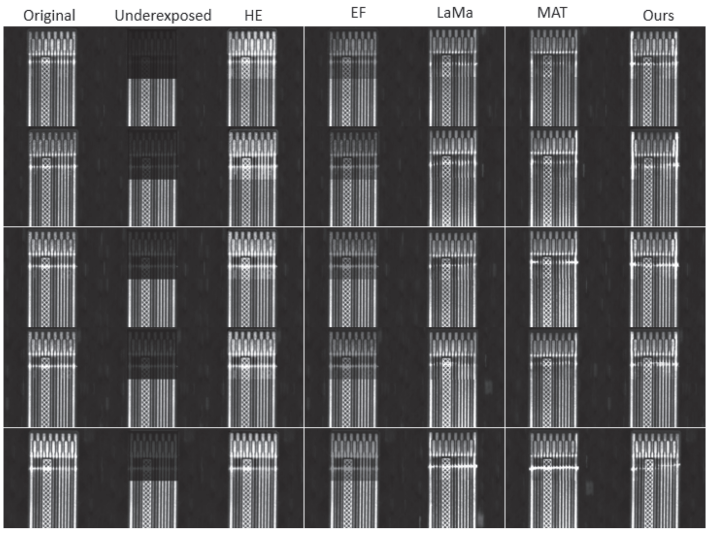
In Table 4, the proposed model achieves optimal performance across all three metrics overall, demonstrating its advantage in structural consistency after image restoration. In Table 5 for local regions, LaMa and MAT, as modern deep inpainting methods, outperform traditional methods in both PSNR and LPIPS metrics, validating the effectiveness of deep learning in image inpainting tasks. However, both methods are designed for general mask inpainting without fully considering the particularity of exposure anomalies. Although LaMa's Fourier convolution enlarges the receptive field, its global context aggregation mechanism lacks specificity for the drastic local contrast changes and grayscale continuity destruction caused by exposure anomalies. In overexposed regions, LaMa's SSIM is only 0.76, lower than 0.82 of the proposed method, indicating significant deficiency in structure preservation; in underexposed regions, the PSNR gap between LaMa and the proposed method is relatively small, but the SSIM gap remains, suggesting its limited contrast restoration capability. The proposed model improves SSIM by 6.3% compared to LaMa, validating the effectiveness of the multi-scale supervision mechanism in exposure anomaly repair. Although MAT's Transformer architecture can model long-range dependencies, its self-attention mechanism shows insufficient response to local features of exposure anomalies. Due to MAT's global modeling through

Transformer, it achieves optimal performance in LPIPS metric, indicating high perceptual quality of generated images, but its SSIM is significantly lower than the proposed method. In overexposed regions, MAT's SSIM is only 0.79, while the proposed method reaches 0.82, validating the advantage of the proposed model in maintaining structural consistency in exposure transition regions. Furthermore, MAT's LPIPS for underexposure is higher than for overexposure, indicating that it produces more perceptual distortion when handling underexposure contrast enhancement, while the proposed method performs more balanced in overexposure and underexposure scenarios, reflecting the advantage of multi-scale supervision in being insensitive to exposure anomaly types.

The actual restored images of the above five methods are shown in Fig. 10.



(a)



(b)

Figure 10: Comparison of the five methods after repairing images with exposure anomalies: (**a**) Overexposure transformation. (**b**) Underexposure transformation.

Fig. 10 respectively shows the restoration effects of ten original images after exposure transformation through five methods, including five overexposed cases and five underexposed cases. Consistent with the evaluation metric analysis, histogram equalization can only barely recover the general image for overexposed images, with lost texture details unrecoverable, while the repair effect for underexposed images is better. The exposure fusion network demonstrates relatively balanced restoration effects, capable of recovering images close to the original level, but with noticeable boundaries at the connection between the region and the original image. This is because, to simulate exposure anomalies in actual scenarios, the exposure transformation of the original image is only applied within a region of interest, and the transformed image itself has obvious grayscale discontinuities, which persist after repair by the exposure fusion network. LaMa still exhibits obvious brightness distortion in overexposed regions after repair, failing to effectively recover lost texture details, with slight color shifts appearing at exposure transition boundaries, while underexposed regions show insufficient contrast enhancement, with the overall image appearing dark and failing to fully restore shadow details. The MAT method demonstrates good perceptual quality in the repaired images, with overall visual effects being relatively natural, but some detailed textures appear slightly blurred, edge contours in overexposed regions are not as clear as the original image, and there are certain differences in light and shadow levels in underexposed regions compared to the original image. The images restored by the multi-scale supervised dual-layer generative adversarial network proposed in this study are visually closest to the original images, with repaired regional images being complete and richer in details, while connecting more smoothly with the overall image. Overall, LaMa and MAT, as representative methods for general image inpainting, both have limitations in this specific task of exposure anomaly repair: LaMa is limited by its insufficient processing capability for local contrast changes, while MAT still has room for improvement in fine structure preservation in exposure anomaly regions. The proposed model specifically addresses the problems of contrast recovery, detail reconstruction, and boundary smoothing in exposure anomaly repair, achieving optimal performance in both subjective visual effects and objective evaluation metrics.

4.3 Ablation Experiment

The multi-scale supervised dual-layer adversarial generative network proposed in this study adopts a progressive training strategy, dynamically adjusting the training weights: focusing on coarse scales in the early stage, balancing in the middle stage, and focusing on fine scales in the later stage. The specific training results are shown in Fig. 11. The first row contains five original images randomly selected from the dataset, the second row shows the corresponding mask images, and the third to fifth rows display the images generated by the generator at each scale (32×32, 64×64, 128×128).

To validate the effectiveness of each component of the proposed multi-scale supervised dual-layer model, the following comparative experiments are designed:

1. **Single-layer GAN**: Serves as the most basic baseline. This model contains only a single generator and a single discriminator (local). The input is a 256×256 mask image, and it directly generates the 128×128 local region without the multi-scale supervision module, global discriminator, or progressive weighting strategy. The generator adopts a standard U-Net structure, and the discriminator is a single-scale PatchGAN.
2. **Dual-layer model without multi-scale supervision**: Retains the dual-layer architecture (local generator + global discriminator) but removes the multi-scale supervision module. G only outputs a single scale of 128×128, and D_1 only discriminates at the 128×128 scale, without 32×32 and 64×64 auxiliary supervision.
3. **Multi-scale supervised dual-layer model without progressive training**.
4. The proposed model.

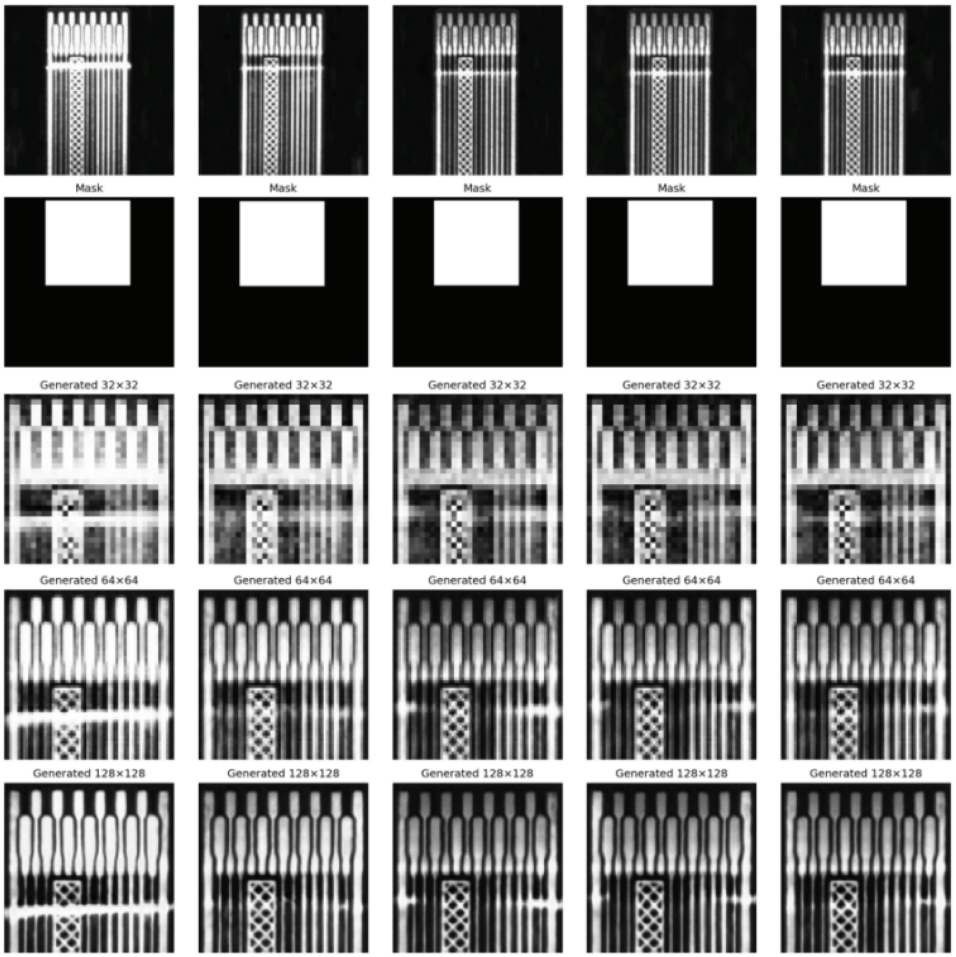


Figure 11: Multi-scale supervised dual-layer generative adversarial model training result image.

The quantitative results of the ablation study are summarized in Table 6, where PSNR (in dB), SSIM, and LPIPS are reported for both the entire image and the local restored region. The proposed complete model (OURS) consistently achieves the highest scores across all evaluation criteria: PSNR of 35.03 dB (entire) and 29.14 dB (local), SSIM of 0.95 and 0.84, and LPIPS of 0.05 and 0.16, respectively. These results validate the incremental contribution of each proposed component.

Table 6: Ablation study on different components.

Model	Architecture			PSNR		SSIM		LPIPS	
	Dual	MS	Prog	Total	Local	Total	Local	Total	Local
GAN	×	×	×	33.86	19.45	0.75	0.69	0.15	0.30
Dual-GAN	✓	×	×	34.21	27.32	0.88	0.81	0.10	0.22
MS-Dual-GAN	✓	✓	×	34.65	28.45	0.91	0.83	0.07	0.19
OURS	✓	✓	✓	35.03	29.14	0.95	0.84	0.05	0.16

Several key observations emerge from the ablation results. First, the comparison between the single-layer GAN and Dual-GAN validates the effectiveness of the dual-layer architecture: introducing the global

discriminator alongside the local generator yields substantial improvements across all metrics, with the full-image PSNR rising from 33.86 to 34.21 dB. Second, equipping the dual-layer model with multi-scale supervision (MS-Dual-GAN) further boosts performance, raising the PSNR to 34.65 dB and markedly improving the SSIM from 0.88 to 0.91. Third, incorporating the progressive training strategy (Prog) into the full model produces additional gains, particularly for the local region, where the PSNR improves from 28.45 to 29.14 dB and the SSIM increases from 0.83 to 0.84. Notably, the LPIPS metric decreases consistently as each component is added, indicating progressively enhanced perceptual quality.

The actually restored images of the four experiments are shown in Fig. 12.

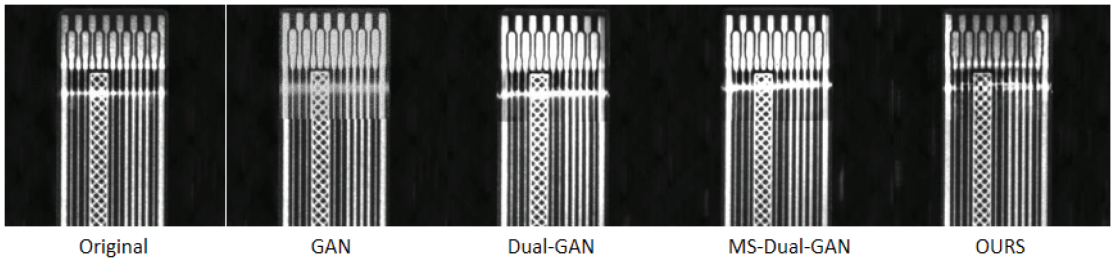


Figure 12: Comparison images of four experiments.

From Fig. 12, it can be intuitively observed that the region images generated by the single-layer GAN are slightly blurry, and when fused into the original image, the image texture is discontinuous at the connection. This is because the single-layer GAN only learns the image generation method for exposure anomaly regions without considering the matching degree of the generated image with the original image. Compared with the single-layer GAN, the dual-layer GAN introducing the global discriminator effectively solves the problem of discontinuous fusion boundaries, but the quality of locally generated images still has room for improvement. The model training of the dual-layer GAN is shown in Fig. 13. In the generator loss plot, the training loss is unstable in the early stage with extremely poor generalization, and the validation loss oscillates violently without a stable convergence trend. In the discriminator loss plot, the validation loss surges in the middle stage, indicating discriminator generalization collapse. In the generator adversarial loss decomposition plot, G_adv2 is significantly higher than G_adv1 in the later stage, indicating that global consistency is more difficult to optimize, the discriminator is too strong, and the generator is difficult to deceive. In the generator reconstruction loss plot, the curve converges fast in the early stage, but is interfered by adversarial loss in the later stage, with reconstruction loss rising and the difference between generated images and targets increasing. This figure shows that the model training is unstable and the discriminator dominance is severe.

After introducing the multi-scale supervision module, the detail reconstruction of local images is improved, with local PSNR increasing by 1.13%, while the overall structural consistency is slightly enhanced. However, the model training curves are shown in Fig. 14. The generator total loss oscillates violently on the validation set, indicating unstable training. The global discriminator is too strong, with D_2 loss significantly lower than D_1, making it difficult for the generator to simultaneously satisfy local and global constraints. Under fixed weights, the reconstruction quality of fine scales is obviously worse than that of coarse scales, and the generator fails to effectively learn high-frequency details.

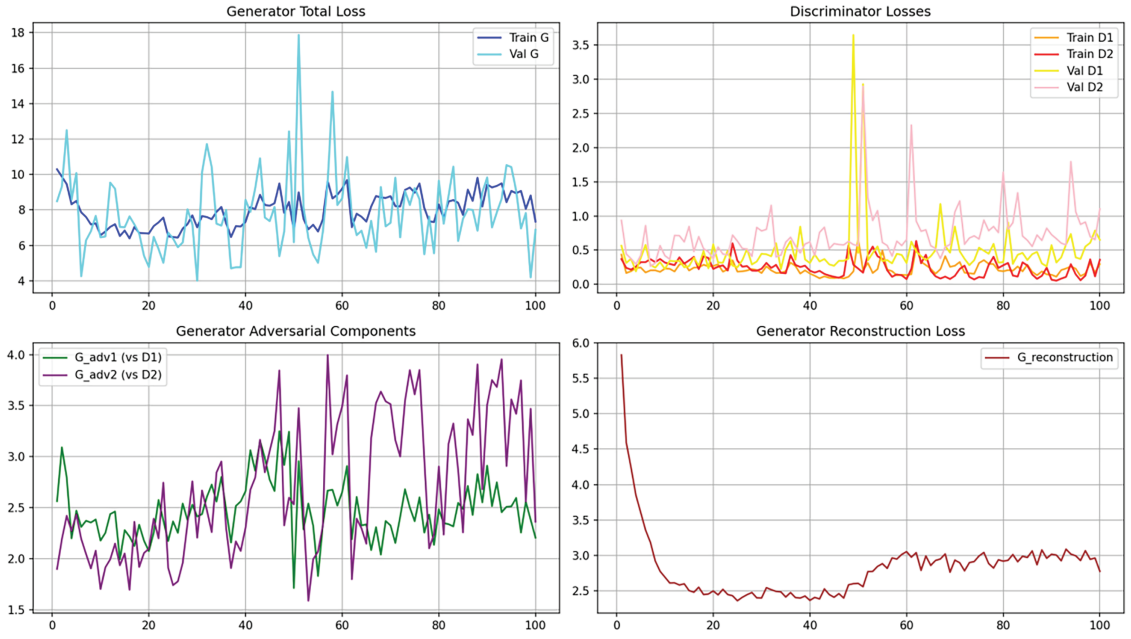


Figure 13: Performance chart of Dual-GAN.

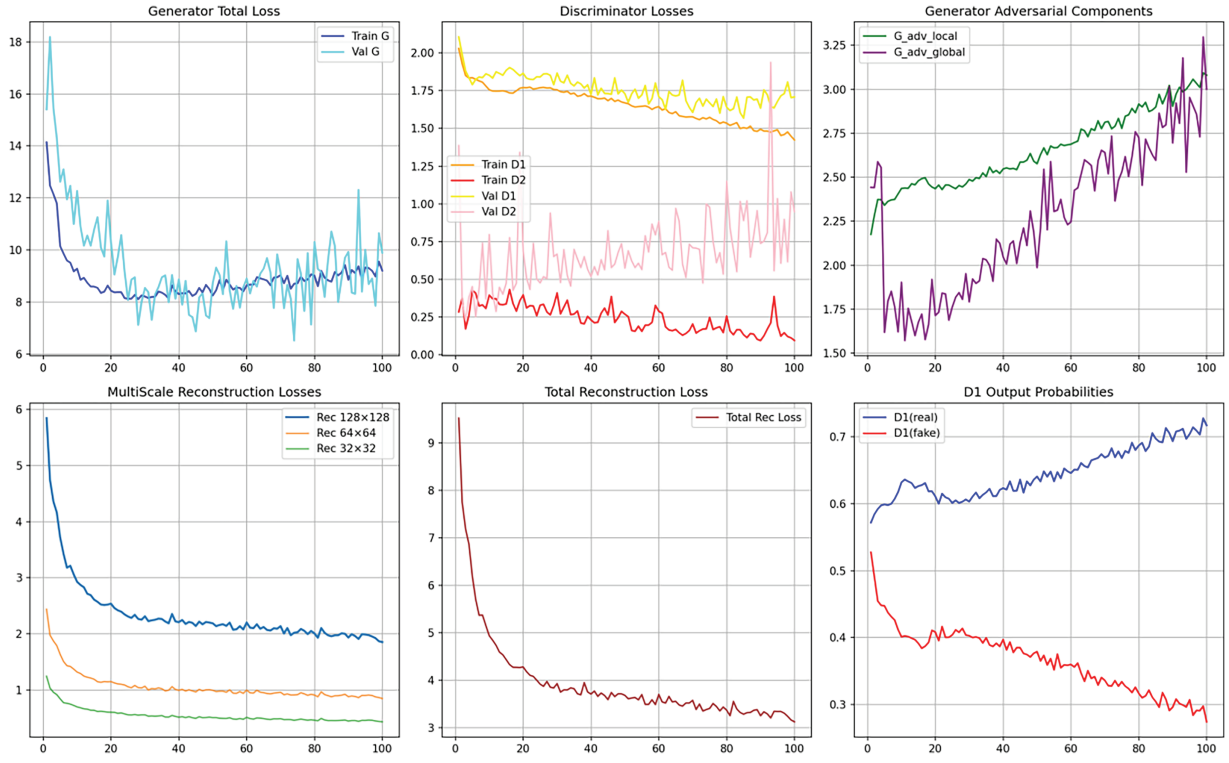


Figure 14: Performance chart of MS-Dual-GAN.

After introducing the progressive training weights, all metrics reach the optimal values, while the training process achieves stable convergence. As seen in Fig. 15 Several loss curves exhibit significant

fluctuations at the 30th and 70th epochs due to changes in the training weights assigned to each scale: early stages emphasize coarse-scale learning, middle stages focus on mid-scale learning, and later stages prioritize fine-scale learning. The validation losses of the generator and discriminator are slightly higher than their training losses. The loss of discriminator D_2 is lower than that of D_1 because D_2 is a single-scale global discriminator, which is relatively simple, while D_1 is a multi-scale local discriminator, which is more complex. The losses of discriminators at each scale show a downward trend and gradually converge. The difference in D_1's output probabilities for real and generated images gradually decreases, eventually approaching 0.5, indicating that the quality of the generated images is nearly sufficient to fool the discriminator.

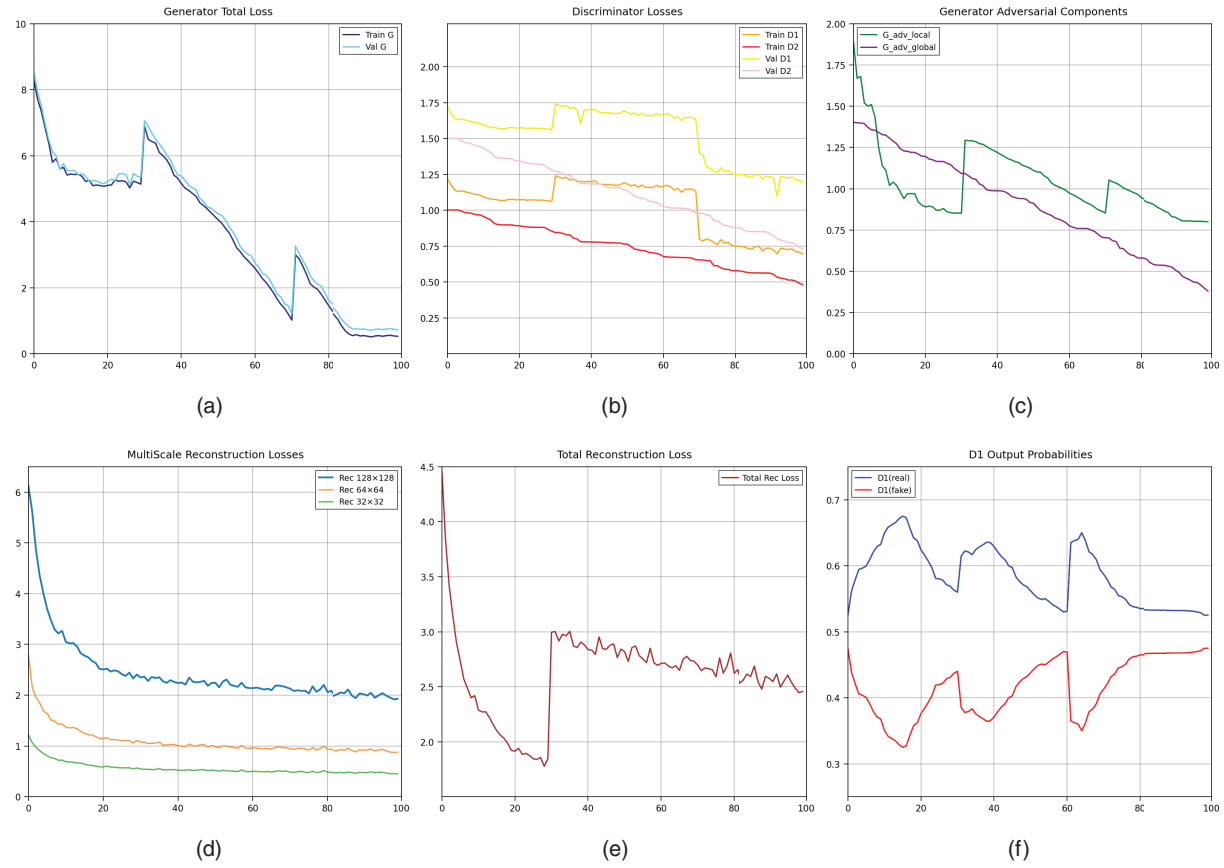


Figure 15: Performance graph of the dual-layer generative adversarial model with multi-scale supervision: (**a**) Generator total loss. (**b**) Discriminator losses. (**c**) Generator adversarial components. (**d**) MultiScale reconstruction losses. (**e**) Total reconstruction loss. (**f**) D_1 Output probabilities.

4.4 Field Experiments

In order to verify the reliability of the model in actual industrial production environments, we carried out an actual production line deployment in an intelligent manufacturing factory in Jiangxi, China, to evaluate the performance of our proposed model, as shown in Fig. 16.

(a) (b)

Figure 16: Actual production line diagram: (**a**) Deployed equipment. (**b**) Connection between the equipment and the upstream production line.

This production line first uses an industrial camera to photograph small-sized LCMs on the conveyor belt. Under normal circumstances, the material feeding positioning algorithm identifies the FPC end gold finger feature and instructs the robotic arm to perform the feeding operation. When there is an image exposure anomaly and the feature still cannot be recognized after multiple runs of the feature matching algorithm, the deployed model described in this study is activated. The abnormal image is input into the model, which first crops the image according to the process used during dataset preparation, then restores the abnormal region image. The restored image is sent to the feeding positioning algorithm for feature matching to obtain the module's pixel coordinates.

The production line employs an industrial camera from Hikrobot with model MV-CS200-10GM. The camera resolution is 5472 × 3648, the working distance is 700 mm, the target width is 130 mm, and the target height is 87 mm. The single-pixel precision formula of the camera is calculated as:

$$P_h = \frac{W_t}{R_h}, \quad P_v = \frac{H_t}{R_v} \tag{14}$$

where P_h is the single-pixel precision in the horizontal direction, P_v is the single-pixel precision in the vertical direction, W_t is the target width, H_t is the target height, R_h is the horizontal resolution of the camera, and R_v is the vertical resolution of the camera. The calculation yields that the single-pixel precision in both horizontal and vertical directions is approximately 0.02 mm. Due to multi-level target cropping, the image processed by the model is only a portion of the actual scene image, roughly one-twenty-fifth of the original image. The actual coordinate error should theoretically be magnified by 25 times, and the actual error is calculated as follows:

$$E_X = E_x \cdot S, \quad E_Y = E_y \cdot S \tag{15}$$

where E_X and E_Y are the errors of the actual x and y coordinates, respectively, and E_x and E_y are the errors of the cropped x and y coordinates, respectively. S is the scaling factor. The calculation yields that the average deviation of the actual x coordinate is 1.075 pixels, and the average deviation of the y coordinate is 3.75 pixels. The conversion to the world coordinate system error is calculated as follows:

$$\delta_x = E_X \times P, \quad \delta_y = E_Y \times P \tag{16}$$

where δ_x and δ_y are the actual horizontal and vertical errors of the module, respectively. The calculation yields a horizontal error of 0.02 mm and a vertical error of 0.075 mm. The allowable feeding error for LCM is 0.08 mm, which is verified through subsequent feeding and lighting steps to meet the actual requirements.

This study collected data from 500 instances of abnormal image exposure, recording and comparing the midpoint coordinates of feature areas in the anomaly images after physical restoration (by manually straightening the FPC to normalize the exposure) with the midpoint coordinates of feature areas after exposure repair. Five groups of coordinate data are extracted as shown in Table 7. The average error of 500 sets of coordinate data is calculated as 0.09 pixels, with a standard deviation of 0.06 pixels and a maximum error of 0.38 pixels. The coordinate deviation vector diagram is shown in Fig. 17, where blue arrows represent normal cases, red arrows represent abnormal cases, and the orange dashed circle indicates the threshold boundary. Among the repaired coordinates, 25 sets exceed the error range, resulting in a failure rate of 5%.

Table 7: Comparison of module coordinates of physically restored images with coordinates of generated repaired images.

	Original		Generated	
	X	**Y**	**X**	**Y**
1	125.535	46.992	125.757	47.472
2	124.603	50.728	124.911	50.959
3	124.078	50.556	123.890	51.704
4	125.779	46.743	125.880	47.703
5	124.706	46.931	124.978	47.776

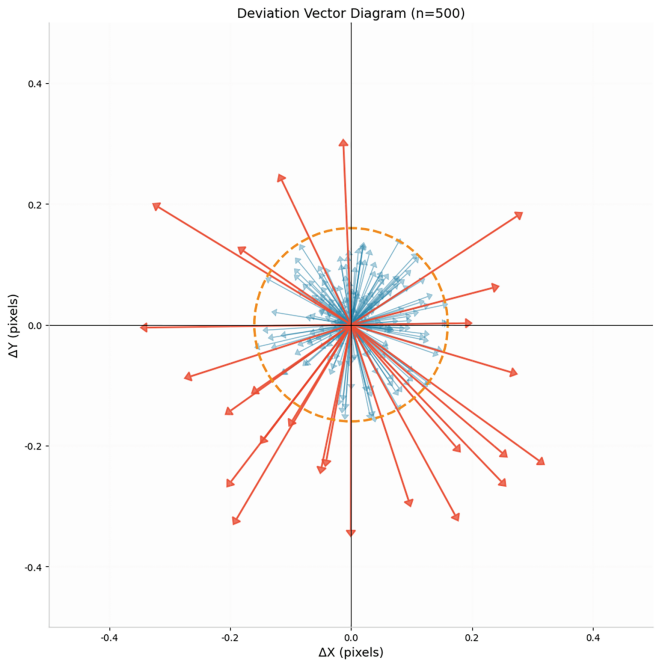


Figure 17: Vector diagram of coordinate data deviation.

To verify the generalization capability of the model, this study additionally collected feeding images of three different specifications of LCM modules from the production line for contrast loss restoration experiments. The restoration effects are shown in Fig. 18. One hundred experiments were conducted for each model, and the calculated average errors for the three models are 0.08, 0.09, and 0.09 pixels, respectively, with failure rates of 3%, 7%, and 5%, respectively. Model A performs the best, possibly because its appearance is similar to the model used during training. Model B has the highest failure rate, as its FPC differs significantly from that used during training. The error vector diagrams of the three models are shown in Fig. 19.

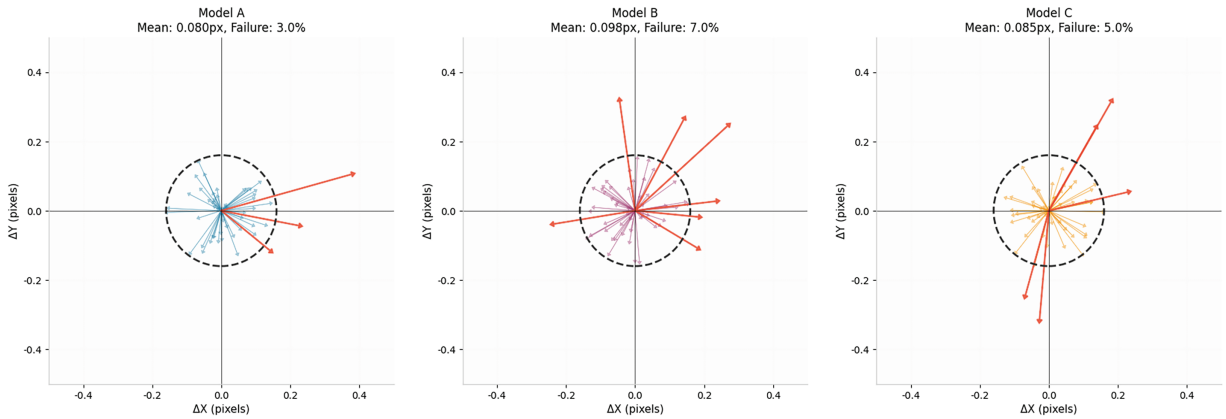
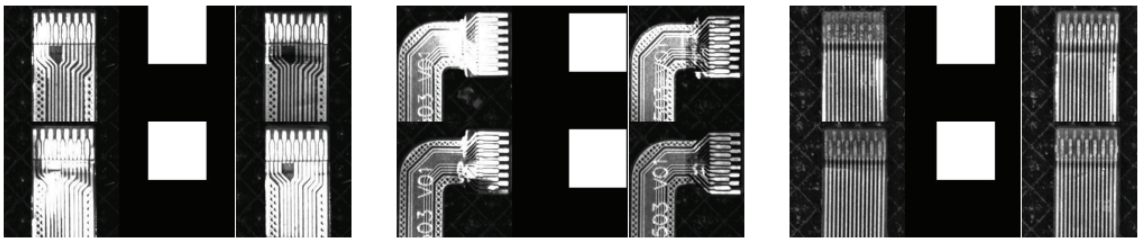


Figure 18: Generalization experiment results on different LCM.



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

Figure 19: Original images, masks, and restored images of different LCMs: (**a**) ModelA. (**b**) ModelB. (**c**) ModelC.

5 Conclusion

This study proposes a multi-scale supervised dual-layer generative adversarial network to address the contrast degradation problem caused by uneven illumination during image acquisition in the field of LCD module defect detection. Through the region-of-interest mask input module, controllable local image generation is performed for exposure anomaly regions. A multi-scale supervision module is introduced, which improves the quality of generated images and the fusion matching degree with the original image through three-scale image generation and two-layer discriminator networks. A progressive training strategy is adopted to dynamically adjust the training focus and improve model convergence.

Experimental results demonstrate that the proposed model can accomplish end-to-end image restoration tasks without prior knowledge of the image optical degradation model. In deployment tests in actual industrial production environments, the model controls the feeding positioning error after repairing

exposure anomaly images within 0.08 mm (corresponding to 0.16 pixels), meeting the precision requirements of practical scenarios. Generalization experiments further verify the model's adaptability to different LCM models.

The current limitation of the model is that it only processes grayscale images. Future work will extend to color image restoration and adapt to image degradation problems in more industrial fields.

Acknowledgement: Not applicable.

Funding Statement: The authors received no specific funding for this study.

Author Contributions: The authors confirm contribution to the study as follows: study conception and design: Longhu Huang; data collection: Longhu Huang; analysis and interpretation of results: Longhu Huang; draft manuscript preparation: Longhu Huang, Sheng Zheng. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: The data that support the findings of this study are available from the corresponding author.

Ethics Approval: Not applicable.

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

References

1. Zhu H, Huang J, Liu H, Zhou Q, Zhu J, Li B. Deep-learning-enabled automatic optical inspection for module-level defects in LCD. IEEE Internet Things J. 2022;9(2):1122–35. doi:10.1109/JIOT.2021.3079440.
2. Luo S, Zheng S, Zhao Y. YOLO-DEI: enhanced information fusion model for defect detection in LCD. CMC. 2024;81(3):3881–901. doi:10.32604/cmc.2024.056773.
3. Kim S, Nguyen TP, Yoon J. A systematic review of machine vision applications in factory and manufacturing processes: from quality control to predictive diagnostics. Int J Precis Eng Manuf Green Technol. 2026;13(2):745–75. doi:10.1007/s40684-025-00769-2.
4. Xiang D, Wang H, He D, Zhai C. Research on histogram equalization algorithm based on optimized adaptive quadruple segmentation and cropping of underwater image (AQSCHE). IEEE Access. 2023;11(2):69356–65. doi:10.1109/ACCESS.2023.3290201.
5. Deng W, Xie G. Image contrast enhancement and brightness preservation based on an adaptive histogram correction framework. Appl Opt. 2025;64(13):3502. doi:10.1364/AO.557280.
6. Liu H, Wang F, Huang Z, Yang Y, Situ G. AdaptiveNet: a learning-based method for the restoration of optically degraded images. J Opt. 2025;27(4):045703. doi:10.1088/2040-8986/adbd50.
7. Chen Y, Yu M, Jiang G, Peng Z, Chen F. End-to-end single image enhancement based on a dual network cascade model. J Visual Commun Image Represent. 2019;61(5):284–95. doi:10.1016/j.jvcir.2019.04.008.
8. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial networks. Commun ACM. 2020;63(11):139–44. doi:10.1145/3422622.
9. Chen X, Yu J, Kong S, Wu Z, Fang X, Wen L. Towards real-time advancement of underwater visual quality with GAN. IEEE Trans Ind Electron. 2019;66(12):9350–9. doi:10.1109/TIE.2019.2893840.
10. Lin JC, Hsu CB, Lee JC, Chen CH, Tu TM. Dilated generative adversarial networks for underwater image restoration. J Mar Sci Eng. 2022;10(4):500. doi:10.3390/jmse10040500.
11. Hu Y, Wang Y, Zhang J. DEAR-GAN: degradation-aware face restoration with GAN prior. IEEE Trans Circuits Syst Video Technol. 2023;33(9):4603–15. doi:10.1109/TCSVT.2023.3244786.
12. Chahi A, Kas M, Kajo I, Ruichek Y. MFGAN: towards a generic multi-kernel filter based adversarial generator for image restoration. Int J Mach Learn Cybern. 2024;15(3):1113–36. doi:10.1007/s13042-023-01959-7.
13. Nguyen BA, Kha MB, Dao DM, Nguyen HK, Nguyen MD, Nguyen TV, et al. UFR-GAN: a lightweight multi-degradation image restoration model. Pattern Recognit Lett. 2025;197(12):282–7. doi:10.1016/j.patrec.2025.08.008.

14. Balaji K. Pervasive multifaceted process based generative adversarial network for image quality enhancement. Appl Soft Comput. 2025;184(7):113780. doi:10.1016/j.asoc.2025.113780.
15. Niu S, Li B, Wang X, Peng Y. Region- and strength-controllable GAN for defect generation and segmentation in industrial images. IEEE Trans Ind Inf. 2022;18(7):4531–41. doi:10.1109/TII.2021.3127188.
16. Zhang L, Rao A, Agrawala M. Adding conditional control to text-to-image diffusion models. In: Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision (ICCV); 2023 Oct 1–6; Paris, France. p. 3813–24.
17. Li Y, Liu H, Wu Q, Mu F, Yang J, Gao J, et al. GLIGEN: open-set grounded text-to-image generation. In: Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2023 Jun 17–24; Vancouver, BC, Canada. p. 22511–21.
18. Bar-Tal O, Yariv L, Lipman Y, Dekel T. MultiDiffusion: fusing diffusion paths for controlled image generation. In: Proceedings of the 40th International Conference on Machine Learning (ICML); 2023 Jul 23–29; Honolulu, HI, USA. p. 1737–52.
19. Snell J, Ridgeway K, Liao R, Roads BD, Mozer MC, Zemel RS. Learning to generate images with perceptual similarity metrics. In: Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP); 2017 Sep 17–20; Beijing, China. p. 4277–81.
20. Zheng S, Zhao Y, Chen X, Luo S. An LCD defect image generation model integrating attention mechanism and perceptual loss. Symmetry. 2025;17(6):833. doi:10.3390/sym17060833.
21. Suvorov R, Logacheva E, Mashikhin A, Remizova A, Ashukha A, Silvestrov A, et al. Resolution-robust large mask inpainting with fourier convolutions. arXiv:2109.07161. 2021.
22. Li W, Lin Z, Zhou K, Qi L, Wang Y, Jia JMAT. Mask-aware transformer for large hole image inpainting. In: Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2022 Jun 18–24; New Orleans, LA, USA. p. 10758–68. doi:10.1109/CVPR52688.2022.01048.