ITC 1/55
Information Technology and Control
Vol. 55 / No. 1/ 2026
pp. 323-339
DOI 10.5755/j01.itc.55.1.42892

Unsupervised CycleGAN-Based Model for Optimizing Depth-of-Field Effects in Photographic Images

Received 2025/09/26 | Accepted after revision 2026/01/07

HOW TO CITE: Sun, G., Liu, Y. (2026). Unsupervised CycleGAN-Based Model for Optimizing Depth-of-Field Effects in Photographic Images. *Information Technology and Control*, 55(1), 323-339. https://doi.org/10.5755/j01.itc.55.1.42892

Unsupervised CycleGAN-Based Model for Optimizing Depth-of-Field Effects in Photographic Images

Guopeng Sun
Image Art College, Luxun Academy of Fine Arts, Shenyang, 110004, China

Yin Liu*
School of Visual Communication Design, Luxun Academy of Fine Arts, Dalian, 116650, China

Corresponding author: YinlylyLiu@outlook.com

To address issues such as blurred foreground-background separation, low detail fidelity, and hazy background blur in traditional methods for optimizing depth-of-field effects in photographic images, this study proposes an improved unsupervised cyclic generative adversarial networks model. First, the study builds upon unsupervised cyclic generative adversarial networks, enhancing the generator through deep separable convolutions and coordinate attention modules. It replaces the dual-discriminator architecture with a class activation mapping module within the discriminator, while integrating the Retinex algorithm into post-processing to address background haze effects. Algorithm performance test results indicated that in the depth-of-field information processing scene, structural similarity reached 30.4. In the edge information processing scene, structural similarity reached 28.1. All scenes maintained optimal performance. In depth-of-field optimization tests for bow-and-net images, the model's output images achieved an average information entropy of 7.22 (compared to 7.12 for original images). It effectively mitigated typical depth-of-field defects such as background overexposure and motion blur. The model achieves a balance between depth-of-field effects, efficiency, and visual quality, generating high-resolution shallow depth-of-field images. It provides a solution for unsupervised depth-of-field optimization, applicable to post-production photography and industrial image analysis. Future performance enhancements can be achieved through multi-scene dataset expansion and end-to-end optimization.

KEYWORDS: Unsupervised; CycleGAN; Photographic images; Depth-of-field; Retinex algorithm

1. Introduction

In the realms of photographic creation and image post-processing, depth-of-field (DOF) serves as a core visual element for shaping layered compositions and emphasizing focal subjects, directly influencing both artistic expression and the efficiency of information transmission in images [29]. Shallow DOF imagery effectively guides viewers' attention toward core areas through the visual contrast of sharp foregrounds and blurred backgrounds. This technique finds widespread application in portrait photography, product displays, and industrial inspection scenes [31]. Traditional DOF effects rely on either professional optical equipment or manual post-processing software adjustments. The former is constrained by hardware costs and shooting scenes, while the latter requires manual annotation of foreground areas, resulting in complex operations, low efficiency, and highly subjective outcomes. These methods struggle to meet the demands for large-scale, automated DOF optimization in photographic imagery [28].

With the advancement of deep learning technologies, generative adversarial networks (GANs) have opened new avenues for image style transfer and cross-domain conversion. Among these, cyclic generative adversarial networks (CycleGAN) have emerged as a significant research direction for cross-domain visual tasks such as depth effect optimization and image style transfer. Leveraging their unsupervised learning capability that eliminates the need for paired samples, CycleGANs have overcome the reliance on labeled datasets inherent in traditional supervised learning approaches [19, 23]. CycleGAN achieves bidirectional cross domain mapping through the constraint of cyclic consistency loss, which makes it more stable than earlier GAN models in unpaired scenarios and lays the methodological foundation for subsequent related research. In recent years, the academic community has conducted extensive exploration on the structural improvement and performance enhancement of CycleGAN, forming optimization branches for different task scenarios. In quality optimization tasks such as low light image enhancement, G. Tang et al. proposed an improved CycleGAN model to address the problem of detail loss in traditional CycleGAN generated images. By optimizing the residual structure of the generator and the discriminator loss function, the model effectively improves the detail fidelity and color reproduction of images in low light environments [21]. In the direction of model lightweighting, He's team [9] proposed the CycleGAN* architecture, which introduces collaborative learning mechanisms and improved adversarial networks to enhance multimodal data adaptation capabilities while reducing computational complexity, providing key technical support for the deployment of CycleGAN in resource constrained scenarios such as mobile devices. In addition, the fusion innovation of GAN and Transformer also provides important reference for the structural upgrade of CycleGAN. The hybrid Transformer GAN model proposed by Grigas et al. [7] significantly improves the accuracy of medical image preprocessing by integrating the global feature capture ability of Transformer with the generation advantage of GAN in structural MRI preprocessing tasks. Its cross domain feature alignment approach provides a new perspective for optimizing the feature extraction module of CycleGAN [7]. At the level of task expansion, CycleGAN's cross domain mapping capability is constantly being explored and applied to non style transfer scenarios. Lin et al. [14] applied CycleGAN to the field of image watermarking, utilizing its cross domain mapping ability to achieve hidden embedding and extraction of watermarks, and verified the scalability of CycleGAN in non style transfer visual tasks. Similarly, the Hemigen model proposed by Dirvanauskas et al., as a GAN based human embryo image generator, achieves high-quality generation of specific biomedical images through an unsupervised generation mechanism. Its generator's detail control strategy shares similarities with CycleGAN's requirement for preserving texture details in depth optimization, providing practical reference for CycleGAN's detail optimization in professional fields [6]. Because its generator only uses one residual structure and ordinary convolutions, it is challenging to tell foreground from background parts. This often results in blurred foreground details or uneven background blurring [3-32]. Furthermore, existing GAN-based DOF optimization models primarily address general-purpose scenes. These models lack the adaptability necessary for specialized applications, such as industrial inspection, which limits their scope [33].

To address the aforementioned issues, this study proposes an optimized model for enhancing the DOF effect in photographic images by integrating unsupervised CycleGAN. Through lightweight improvements to the network architecture and innovative design of the post-processing workflow, the model aims to achieve synergistic improvements in DOF optimization, computational efficiency, and VQ. The innovation of the research lies in the proposal of a lightweight convolution coordinate attention (CA) embedded fusion mechanism, which breaks the limitations of their independent applications and achieves a synergistic improvement in feature extraction efficiency and accuracy. Simultaneously establish a semantic focusing mechanism for single branch discriminators to solve the problem of redundant feature fusion in dual discriminators. In addition, constructing a dynamic adaptation logic between the generated results and Retinex breaks through the limitations of traditional fixed parameter calibration and provides a new paradigm for cross domain generation quality calibration.

2. Methodology

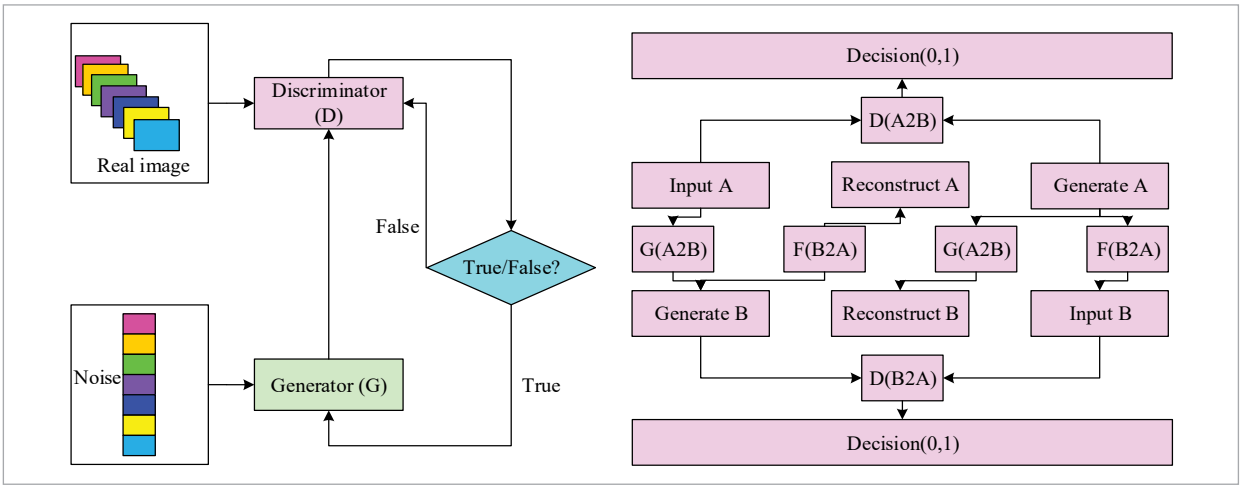
2.1. Unsupervised CycleGAN Algorithm Integrating an Improved Generator with Attention Mechanisms

Unsupervised learning, as one of the core branches of machine learning, fundamentally relies on training samples without human annotation. By autonomously learning the inherent distribution patterns within input data, it uncovers latent structural features or pattern information. Typical application scenes encompass clustering analysis, dimensionality reduction, and generative modeling [2]. In the field of generative modeling, GANs have emerged as a key technology for achieving high-quality image generation and cross-domain translation within unsupervised learning frameworks, leveraging their dynamic generator-discriminator game mechanism. As a significant derivative of GANs, CycleGAN further overcomes the traditional reliance on paired samples by introducing a cyclic consistency constraint. As a result, bidirectional cross-domain mapping relationships can be built, making it easier to translate images across two different domains without the need for paired data [4]. The schematic diagrams of GAN and CycleGAN network training are shown in Figure 1.

Figure 1(a) illustrates the GAN schematic. The classical GAN employs a dual-network adversarial training architecture, with its core comprising two deep learning models: the Generator (G) and the Discriminator (D). These two models achieve collaborative optimization through a zero-sum game [13]. The value function $\min_G \max_D V(D,G)$ of the minimax game is displayed in Equation (1).

$$\min_G \max_D V(D,G) = E_{x\sim pdata}(x)\left[\log D(x)\right] + E_{z\sim pz}(z)\left[\log 1 - D(G(z))\right] \quad (1)$$

Figure 1
Schematic diagram of GANs and CycleGAN training.



In Equation (1), $D(x)$ denotes the discriminator D's estimate of the probability that the real data x is true. $G(z)$ represents the fake data generated by the generator G mapping random noise z. $E_{x\sim pdata}(x)$ is the expectation over all real data distributions. $E_{z\sim pz}(z)$ is the expectation over all noise prior distributions [8]. Figure 1(b) depicts the training schematic of the CycleGAN network. The input layer consists of unpaired images from domains A and B. The generator comprises G (A→B, which converts inputA into generateB mimicking domain B (D-B)) and F (B→A, which converts inputB into generateA mimicking domain A (D-A)). The discriminator includes D(A→B) (determining whether generate B is a real-world B image) and D(B→A) (determining whether generateA is a real-world A image) [30]. Equation (2) defines the discriminator's LF for the mapping direction from D-A to D-B (generator G and discriminator D(A→B)).

$$L(G,D_B,A,B)=E_{b\sim pdata}(b)\left[\log D_B(b)\right]+E_{a\sim pdata}(a)\left[\log 1-D_B(G(a))\right] \quad (2)$$

Equation (3) defines the discriminator's LF for the mapping direction from D-B to D-A (generator F and discriminator D(B→A)).

$$L(G,D_A,A,B)=E_{a\sim pdata}(a)\left[\log D_A(b)\right]+E_{a\sim pdata}(b)\left[\log 1-D_A(G(b))\right] \quad (3)$$

Among these, reconstructA reverses generateB from domain F to produce an image in D-A. The reconstructB reverses generateA from domain G to produce an image in D-B. The fundamental mechanism of CycleGAN for unsupervised cross-domain translation is formed by reconstructA and reconstructB, which both measure the difference between the reconstructed images (RIs) and the original inputs to establish cycle constraints [22]. The L1 norm is employed to quantify the difference between the RI and the original image, as shown in Equation (4).

$$L(G,F)=E_{a\sim pdata}(a)\left[\|F(G(a))-a\|1\right]+E_{b\sim pdata}(b)\left[\|G(F(b))-b\|1\right] \quad (4)$$

The generator architecture in the CycleGAN network is shown in Figure 2.

In Figure 2, the input stage receives a single image from the source domain (SD). The encoding stage extracts multi-scale features (MSFs) through three subsampling operations, completing the transfor-

Figure 2

Generator network structure diagram.

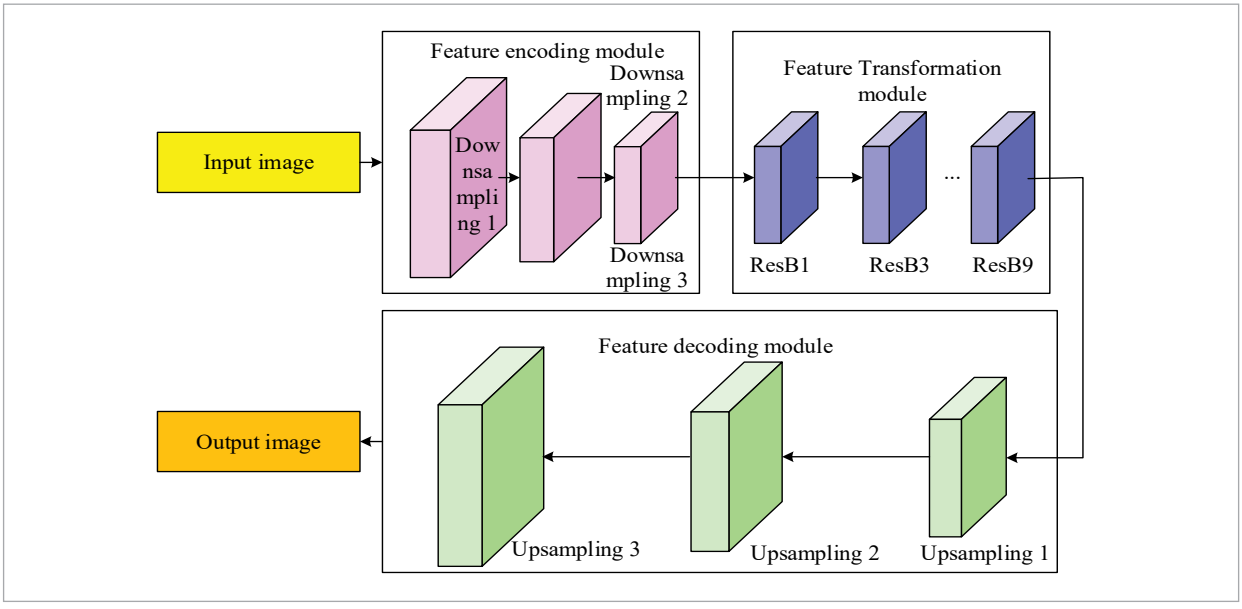
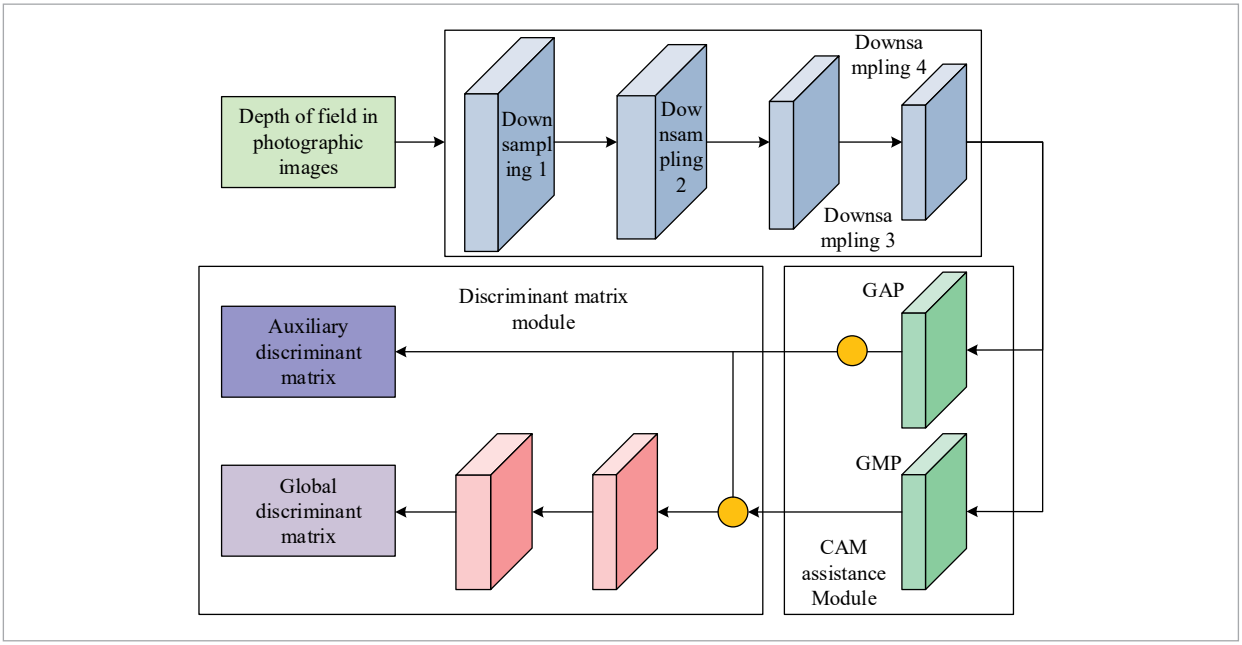


Figure 3
Discriminator network structure diagram.



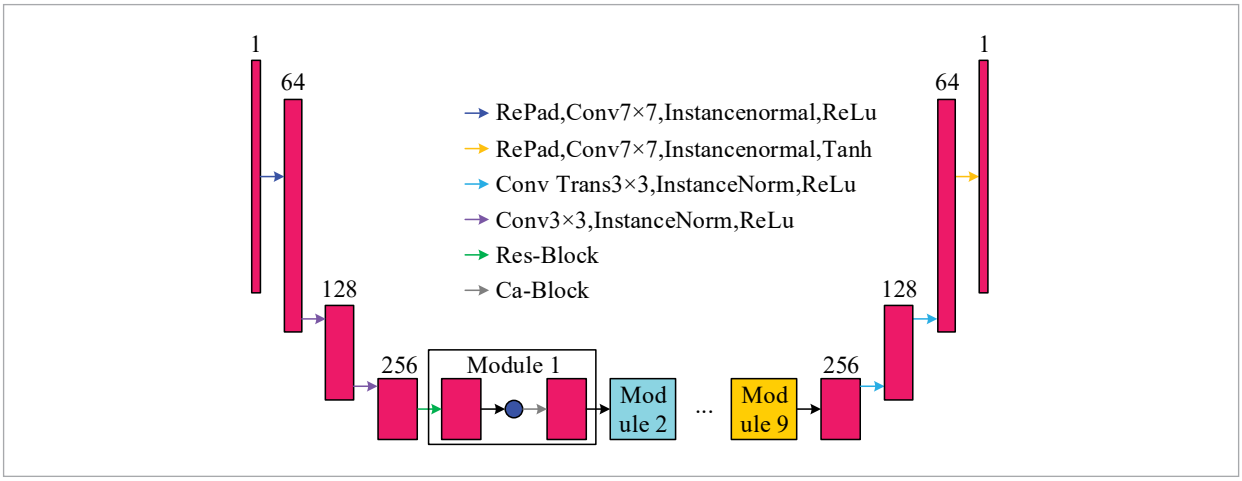
mation from image to high-dimensional features. The transformation stage performs deep processing on the high-dimensional features through nine residual blocks, achieving cross-domain mapping from SD features to target domain features. The decoding stage restores the feature space dimension through three upscaling operations, completing the reconstruction from high-dimensional features to target domain images. The output stage generates an image matching the input image dimensions and conforming to the target domain style [27,31]. Generator G aims to minimize this value function. In practice, the generator's objective is typically transformed into a maximization task, which yields stronger early gradients [12]. Consequently, the LF for the generator is displayed as shown in Equation (5).

$$\min_{G} V(G) = -E_{z\sim pz}(z)\left[\log D(G(z))\right] \tag{5}$$

This study incorporates the CAM attention mechanism as an auxiliary discriminative module within the global discriminator to enhance the CycleGAN discriminator's ability to distinguish between front and rear scenes in cross-domain generated images. This approach avoids the exponential increase in model parameters caused by introducing local discriminators. By utilizing the auxiliary discriminative weight matrix output by CAM, the approach focuses on discriminating key regions within the global feature map (FM) [1]. The discriminator network architecture is illustrated in Figure 3.

As shown in Figure 3, the discriminator features a serial architecture of three modules: a parameter-sharing feature extractor, a CAM assistance module, and a discriminative matrix. The core feature extractor uses four downsampling layers (4×4 Conv + InstanceNorm + LeakyReLU) to capture multi-scale features (MSF), reducing spatial resolution while increasing channels to 512. The subsequent CAM module receives these deep feature maps (FMs) to generate a foreground-focused weight matrix. This matrix guides the final discriminative matrix module to emphasize key regions during its classification, outputting a probability and providing gradient signals for adversarial training. To enhance detail fidelity and efficiency, the standard CycleGAN generator is improved with three key components (Figure 4): depthwise separable convolutions (DSCs), a lightweight coordinate attention block (CA_Block), and an MSF enhancement mechanism.

Figure 4
Overall architecture diagram of the improved generator.



In Figure 4, during the input stage, the original image undergoes reflection padding first. This process expands the image's edge regions through mirroring, preventing edge detail loss caused by subsequent convolutional operations. Subsequently, the image enters a 7×7 DSC layer. This layer captures global shallow features while reducing the model's parameter count to one-ninth that of traditional convolutions. Following instance normalization and ReLU activation, the initial features undergo nonlinear mapping [10,18,32]. The feature encoding stage employs two cascaded subsampling units, each centered around a DSC with a stride of 2. This process progressively halves the spatial resolution of FMs while increasing the quantity of channels to 256 dimensions. Residual modules employ 3×3 separable convolutions coupled with shortcut connections, mitigating gradient vanishing while preserving channel independence. CA_Blocks enhance key semantic region features through horizontal and vertical CA weight allocation. The attention mechanism comparison is illustrated in Figure 5.

Figure 5
Comparison Chart of Attention mechanisms.

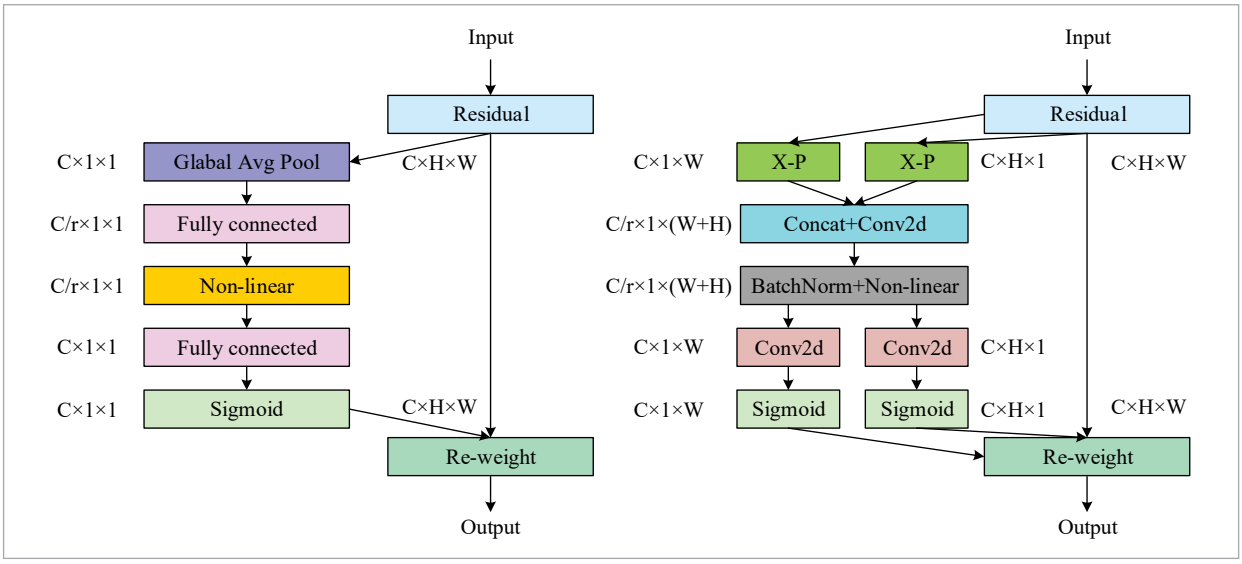


Figure 5(a) depicts the self-attention (SA) module. This module generates three feature spaces, namely Query, Key, and Value, by applying three linear transformations along the channel dimension to the input FM. Figure 5(b) depicts the CA module. This module encodes two-dimensional global spatial information into a pair of direction-aware feature vectors by performing global pooling on the input FM along both the height and width spatial dimensions [16]. Subsequently, using a shared parameter transformation function and nonlinear activation, these two feature vectors are transformed into attention weights carrying spatial position information. To accomplish position-based channel weight recalibration and improve the network's perception of the target's spatial location, these weights are finally multiplied by the input FM. To collaboratively optimize the content conversion, structural consistency, and foreground focusing ability in the image translation process, a multi task comprehensive loss function is constructed as shown in Equation (6).

$$L_t = \lambda_1 \cdot (L_G + L_{D(A\to B)} + L_{D(B\to A)}) + \lambda_2 \cdot L_{cyc} + \lambda_3 \cdot L_{CAM} \quad (6)$$

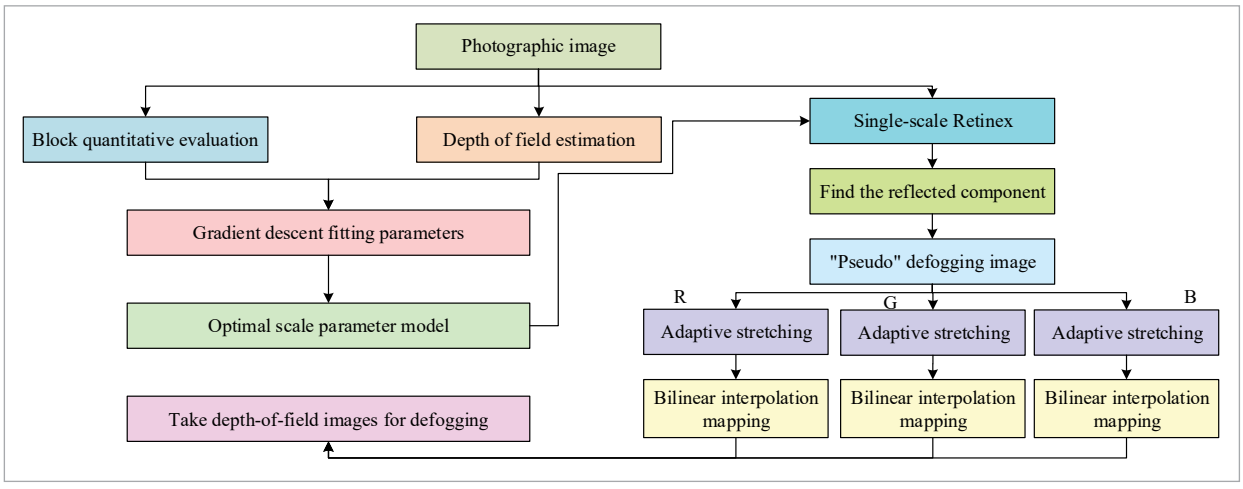
In Equation (6), L_{cyc} represents the cyclic consistency loss, and L_{CAM} represents the CAM guidance loss. During training, the generator generates semantically clear images of the foreground and background through joint optimization of L_G and L_{cyc}; After the parameter sharing module in the discriminator extracts features, the CAM module optimizes the weight matrix based on L_{CAM}, focuses on the features of the foreground region, and weights the input to the discrimination matrix module; The discriminative matrix module combines weighted features to calculate adversarial losses, and updates the discriminator and generator in reverse. The introduction of L_{CAM} enables the CAM module to learn foreground concepts independently without relying on adversarial loss, and ensures attention reliability through mean difference constraints.

2.2. Construction of an Optimization Model for DOF Effects in Photographic Images Based on an Improved Unsupervised CycleGAN

To further mitigate potential defects in generated images such as color bias, background blurring, and overall fuzziness, image enhancement techniques are incorporated. Among these, the Retinex algorithm, based on human visual mechanisms, separates illumination component (IC) and reflection component (RC) through multi-scale Gaussian blurring and corrects color distortions. It effectively eliminates haziness while preserving textures, demonstrating significant advantages in defogging and color correction [11, 16]. The Retinex-based defogging workflow for photographic DOF is illustrated in Figure 6.

Figure 6

A DOF defogging process for photographic images based on Retinex.



In Figure 6, first, to enhance local feature representation without computational burden, the original image is adaptively segmented into blocks. Quantitative metrics then assess the clarity and fog concentration within each block. Second, based on this block-wise assessment, a DOF (Depth of Field) estimation algorithm calculates regional DOF values. This delineates the image's spatial hierarchy (foreground, midground, background), providing a basis for spatially-adaptive defogging. Third, using the DOF map as a constraint, a mapping model between defogging performance and algorithm parameters is constructed. Gradient descent iteratively optimizes this model to minimize the error between theoretical and practical defogging quality, yielding initial optimal parameters. Finally, these initial parameters are combined with the DOF hierarchy for a secondary optimization, constructing a globally optimal scale parameter model. This model is applied in subsequent Retinex-based defogging stages, which separate an image into its illumination and reflectance components to address fog effects [26]. Assuming the observed image is displayed as $I(x, y)$, the Retinex definition is expressed as Equation (7).

$$I(x,y) = L(x,y) \cdot R(x,y) . \tag{7}$$

In Equation (7), $L(x, y)$ represents the IC, and $R(x, y)$ denotes the RC. By applying the single-scale Retinex (SSR) algorithm to the input image and separating its IC and RC through logarithmic transformation, the global haze effect's interference on the RC is preliminarily eliminated. Through logarithmic separation and iterative computation by the SSR algorithm, the RC $R(x, y)$, representing the scene's intrinsic properties, is extracted from the foggy image. Based on the obtained RC $R(x, y)$, preliminary adjustments are made to the RC to generate a "pseudo" defogged image, as shown in Equation (8).

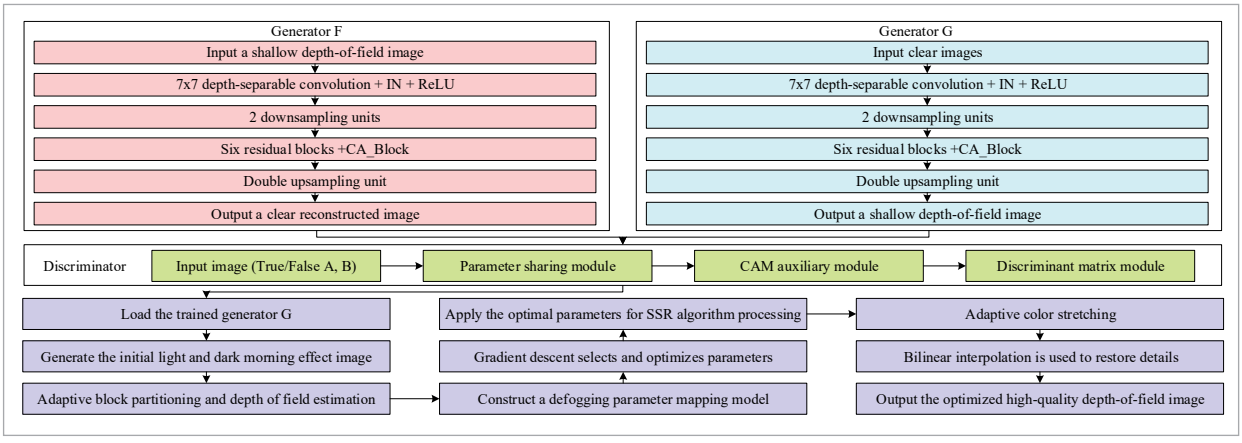
$$r_{SSR}(x,y) = I(x,y) - I(x,y) * G(x,y) . \tag{8}$$

In Equation (8), $r_{SSR}(x, y)$ is the estimated logarithmic RC. * denotes the convolution operation. $G(x, y)$ is the centering function, typically implemented using a two-dimensional Gaussian kernel. To correct color imbalance in the "pseudo" defogged images, an adaptive stretching algorithm is applied independently to each RGB channel. This enhances contrast while preserving inter-channel color harmony. Subsequently, a bilinear interpolation mapping algorithm smooths local pixel distortions caused by stretching. It calculates a weighted average from the four nearest neighboring pixels to restore fine textures and edge details. Together, these steps ensure the final defogged output achieves the clarity and detail fidelity required for practical applications. The complete flowchart of the improved unsupervised CycleGAN-based model is presented in Figure 7.

In Figure 7, the unlabeled source domain image is input first, and the edge information is optimized through preprocessing. Further, by improving the CycleGAN generator to encode and extract multi-scale features, combined with CA-Block to enhance the semantic distinction between foreground and background; Subse-

Figure 7

An optimization model for the DOF effect of photographic images based on an improved unsupervised CycleGAN.



quently, the residual module completes cross domain feature transformation and synchronizes the dynamic game optimization of the discriminator enhanced by CAM. Finally, decode and reconstruct the image, correct the blurring deviation through cyclic consistency constraints, and output the depth of field optimization result. To address potential color deviations and blurred details in the initial image, the process further incorporates a post-processing module based on Retinex theory. By adaptively adjusting defogging parameters through depth estimation, combined with color correction and detail restoration techniques, the VQ is optimized. The final output is a high-quality shallow DOF image featuring distinct foreground and background layers with natural color reproduction. To accurately evaluate model performance, the study employs image information entropy (IE) to quantify the average information content or uncertainty within an image, as shown in Equation (9).

$$H(X) = -\sum_{i=1}^{n} p(x_i)\log_2 p(x_i). \tag{9}$$

In Equation (9), X represents the image. n denotes the total number of possible pixel values (PVs) in the image. x_i represents the i-th gray level (GL). $p(x_i)$ denotes the probability of GL x_i occurring in the entire image. Blind/referenceless image spatial quality evaluator (BRISQUE) is a typical reference-free image quality (IQ) metric that enables objective evaluation of digital image VQ without relying on original reference images. This algorithm extracts natural scene statistics (NSS) features from images and employs a support vector machine regression model for quality prediction. It is primarily used to quantitatively assess key visual characteristics such as image clarity, sharpness, contrast, and noise levels, ultimately generating a quantitative metric reflecting the overall VQ of the image. BRISQUE feature extraction is based on mean subtracted contrast normalized (MSCN) coefficients, which effectively capture image distortion information through normalization of local statistical features. The local mean describes the average brightness level within the neighborhood centered on the target pixel, as shown in Equation (10).

$$\mu(i,j) = \sum_{k=-K}^{K}\sum_{l=-L}^{L} w_{k,l} I_{k,l}(i,j). \tag{10}$$

In Equation (10), w denotes a two-dimensional Gaussian weighting function. K and L display the size of the neighborhood window. $I_{k,l}(i,j)$ is the PV located at $(i+k,j+l)$. Local variance is used to measure the dispersion of PVs within the neighborhood, as displayed in Equation (11).

$$\sigma(i,j) = \sqrt{\sum_{k=-K}^{K}\sum_{l=-L}^{L} w_{k,l}(I_{k,l}(i,j) - \mu(i,j)^2}. \tag{11}$$

In Equation (11), $\sigma(i,j)$ represents the local variance. The MSCN coefficient eliminates the influence of illumination variations and enhances local texture features by performing mean subtraction and contrast normalization on the original PVs, as shown in Equation (12).

$$\hat{I}(i,j) = \frac{I(i,j) - \mu(i,j)}{\sigma(i,j) + C}. \tag{12}$$

In Equation (12), C denotes a constant, set as an extremely small value to prevent the denominator from becoming zero. By extracting MSCN coefficients, the BRISQUE algorithm effectively quantifies image distortion features such as blur and noise, providing critical input for subsequent quality assessment. A popular unreferenced IQ measure that evaluates IQ by examining its naturalness and distortion levels is the naturalness IQ evaluator (NIQE). The NIQE score is calculated as shown in Equation (13).

$$D(f_{test}; v, \Sigma) = \sqrt{(f_{test} - v)^T \Sigma^{-1} (f_{test} - v)}. \tag{13}$$

In Equation (13), f_{test} denotes a column vector. v denotes the mean vector.

3. Results and Analysis

3.1. Performance Analysis of Improved Unsupervised CycleGAN Algorithms

The configuration of network parameters in the unsupervised CycleGAN algorithm directly impacts the effectiveness of DOF feature transfer and IQ optimization. Table 1 displays the model's main parameter values and structural design information.

To comprehensively evaluate the proposed algorithm across various image processing tasks, three

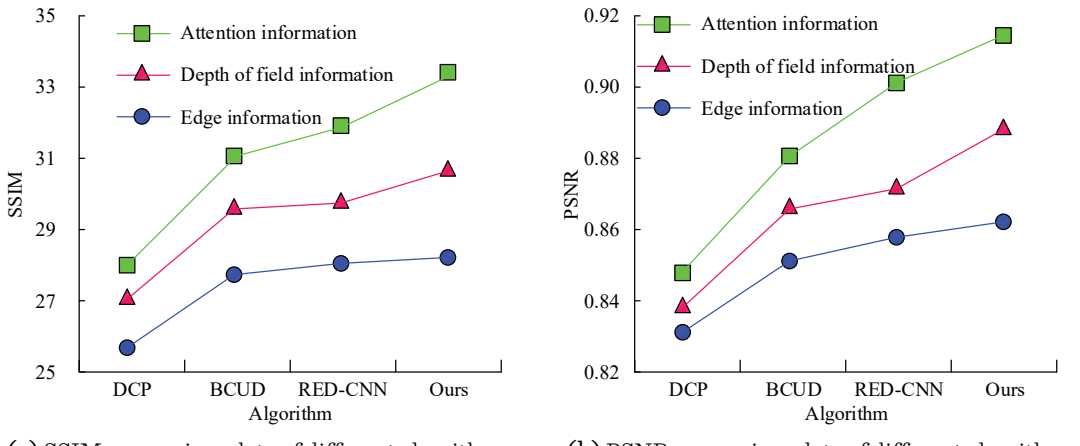
Table 1
Parameter of the improved unsupervised CycleGAN algorithm.

Project	Network layer	Convolutional layer		
Feature encoding network parameters	Convolution kernel	7*7	3*3	3*3
	The number of filters	32	64	128
	The number of convolutional sliding steps	1	2	2
	Normalization method	Instance	Instance	Batch
Feature decoding network parameters	Convolution kernel	3*3	7*7	7*7
	The number of filters	64	32	3
	The number of convolutional sliding steps	1/2	1/2	1
	Normalization method	Instance	Instance	/

benchmark algorithms were selected for comparison: Dark Channel Prior (DCP), Bidirectional Contrastive Unsupervised Denoising (BCUD), and Residual Encoder-Decoder Convolutional Neural Network (RED-CNN). Experiments were conducted in three key scenarios: attention information enhancement, depth-of-field optimization, and edge feature extraction. Performance was measured using Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM). The publicly available AAPM-Mayo dataset was used to ensure objective and reproducible evaluation. PSNR is a widely used objective image quality metric, though it has limitations and is best suited for applications involving known maximum signal values and additive noise. SSIM assesses perceptual similarity between two images. A comparative analysis of SSIM and PSNR across all algorithms is presented in Figure 8.

Figure 8 compares the SSIM and PSNR results of different algorithms. The proposed algorithm achieves the highest SSIM values across multiple scenes—33.6 (attention information processing), 30.4 (DOF information processing), and 28.1 (edge information processing)—owing to the integration of the CA module and CAM mechanism, which enhance key spatial features and foreground optimization while maintaining structural consistency through residual connections. In PSNR, the proposed method also attains a high value of 0.913 in the attention information processing scene.

Figure 8
Comparison data of SSIM and PSNR for different algorithms.



(a) SSIM comparison data of different algorithms (b) PSNR comparison data of different algorithms

To further evaluate the algorithm's capability in preserving visual information integrity, Visual Information Fidelity (VIF) and Information Fidelity Criterion (IFC) are adopted as core metrics. These were tested across four typical image categories: portraits, landscapes, architecture, and food. The experimental outcomes are presented in Figure 9.

Figure 9 presents the IFC and VIF comparison results. The proposed algorithm achieves the highest average IFC (2.85) and VIF (0.391) across all image categories, outperforming the DCP algorithm (IFC: 2.01; VIF: 0.292). Specifically, the proposed method shows approximately 6.2% and 5.1% improvement in VIF over the RED-CNN and BUCD algorithms, respectively. This enhancement is attributed to the integration of CAM and CA modules for precise focus on key visual regions, combined with Retinex-based post-processing that adapts to human visual characteristics, thereby reducing lighting and noise interference and better preserving visual fidelity. For further validation, the SOTS-Indoor dataset was used as a benchmark to evaluate the algorithm's performance in image quality optimization using no-reference metrics BRISQUE and NIQE. The corresponding results are illustrated in Figure 10.

Figure 9

Comparison results of IFC and VIF values of different algorithms.

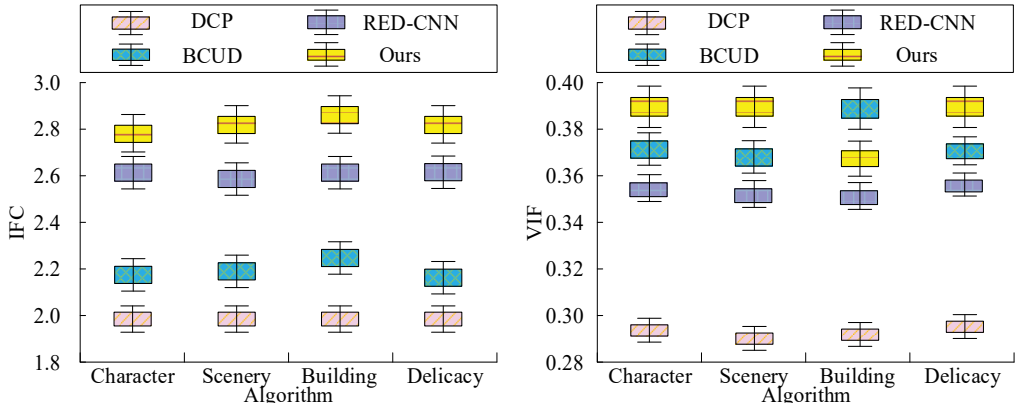
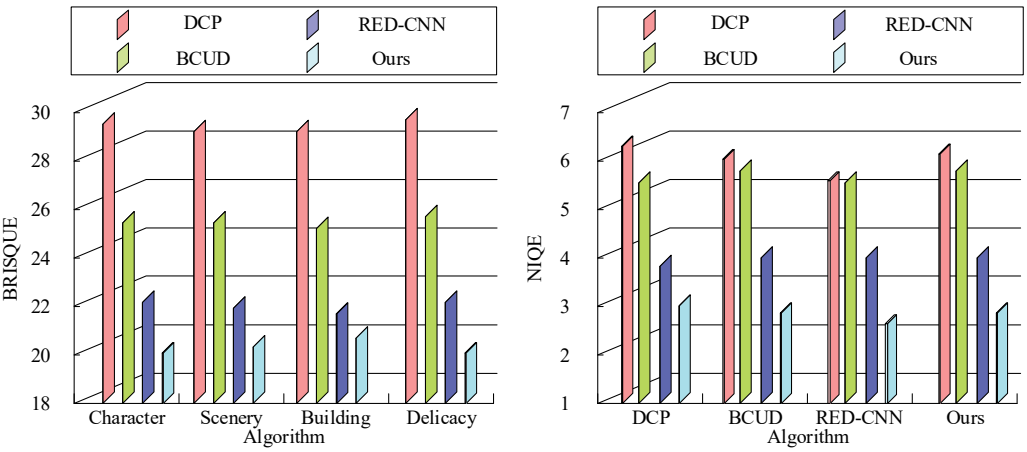


Figure 10

Comparison results between BRISQUE and NIQE.



(a) The BRISQUE comparison results of each algorithm (b) The NIQE comparison results of each algorithm

Figure 10(a) compares the BRISQUE scores of different algorithms. The proposed algorithm achieves the lowest average score of 19.3, which is about 10.2% lower than that of RED-CNN, thanks to its use of depthwise separable convolution, CAM-based foreground focus, and Retinex-based post-processing to reduce distortion. Figure 10(b) shows the NIQE results, where the proposed algorithm also performs best with an average score of 2.1. To evaluate computational efficiency, training parameters, FLOPs, and execution latency were measured under consistent settings, with input images standardized to (3,256,256). The comparison results are shown in Figure 11.

Figure 11 presents computational complexity comparisons: the proposed algorithm achieves balanced efficiency (3.4M parameters, 27G FLOPs, 24ms latency) via its streamlined CAM single-branch discriminator, outperforming RED-CNN (4.4M, 43G, 35ms) while surpassing DCP (0.2M, 1G, 4ms) in capability. Ablation studies on the AAPM Mayo dataset and bow net samples, evaluating PSNR, SSIM, foreground accuracy, and BRISQUE, quantify each module's contribution to DOF optimization (Table 2).

As shown in Table 2, incorporating DSC improves PSNR by 0.7 and SSIM by 0.013 by decoupling spatial and channel convolutions to reduce computational complexity. Adding CA further increases PSNR by 1.8 and SSIM by 0.042 while reducing BRISQUE to 23.6, leveraging coordinate attention to enhance foreground feature response. The CAM module boosts foreground recognition accuracy from 64.5% to 89.7%, with PSNR and SSIM gains of 1.6 and 0.046 respectively, by guiding discriminator focus toward foreground regions via feature weight

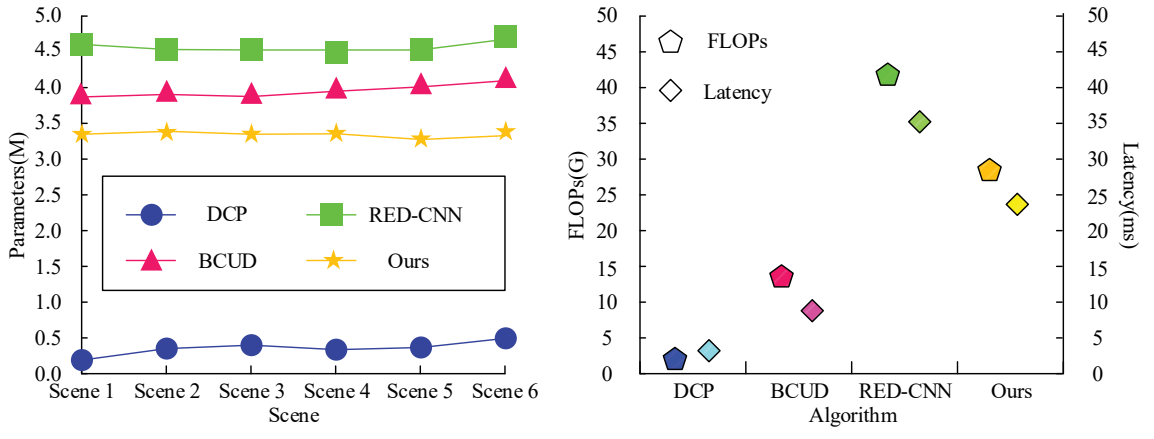
Table 2

Ablation Experiment Results.

Model Configuration	PSNR	SSIM	Prospect recognition accuracy (%)	BRISQUE
Baseline: Original CycleGAN	26.8	0.812	62.3	28.7
Baseline + DSC	27.5	0.825	63.1	27.4
Baseline + DSC + CA	29.3	0.867	64.5	23.6
Baseline + DSC + CA + CAM (i.e., the proposed generator + discrimator)	30.9	0.913	89.7	20.1
Full Model (the above + Retinex post-processing)	31.5	0.921	90.2	19.3

Figure 11

Comparison results of computational efficiency of different algorithms.



(a) Parameters values of different algorithms (b) The FLOPs values and Latency of different algorithms

heatmaps to preserve detail fidelity. Finally, Retinex enhancement achieves a PSNR improvement of 0.6, reduces BRISQUE to 19.3, and increases foreground accuracy to 90.2% by dynamically adjusting illumination components to mitigate background haze and color deviation.

3.2. Performance Evaluation of a DOF Effect Optimization Model for Photographic Images Integrating Unsupervised CycleGAN

To ensure the scientific rigor and reproducibility of performance testing for models optimizing DOF effects in photographic images by integrating unsupervised CycleGAN, experiments are conducted under standardized hardware environments and software frameworks. Table 3 provides specific configuration parameters for the experimental setup.

To comprehensively evaluate the training stability, convergence, and generalization of the CycleGAN-based DOF optimization model while monitoring overfitting, this study establishes a quantitative evaluation system tracking training loss and test PSNR. A dynamic weighting strategy balances cycle consistency, adversarial, and CAM-guided losses: initially set at 10:1:2 for cross-domain consistency, adjusted to 8:1:3 after 50 iterations to enhance foreground-background discrimination, and fixed at 7:1:3 post-100 iterations until convergence. DeepLab v3+ serves as the baseline for comparison, with convergence results shown in Figure 12.

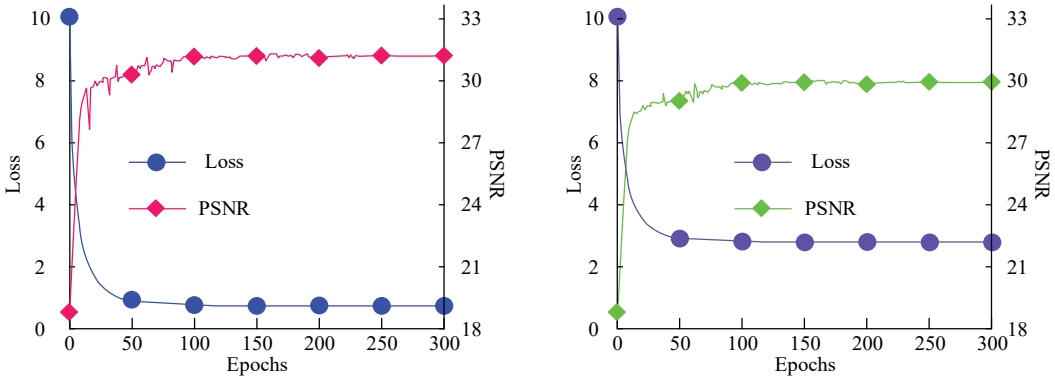
Table 3

Experimental environment configuration.

Configuration item	Model specification
Central processing unit	Intel Xeon Gold 6248R/AMD EPYC 7742
Graphics processor	NVIDIA Tesla V100/NVIDIA A100 (32GB/40GB HBM2)
Memory	128GB / 256GB DDR4 ECC
Hard disk storage	1TB NVMe SSD + 4TB SATA HDD
Operating system	Ubuntu 18.04 LTS/20.04 LTS
Development language	Python 3.8
Deep learning framework	PyTorch 1.9.0/1.10.0 + CUDA 11.1
Unsupervised learning library	CycleGAN

Figure 12

Comparison of loss and PSNR curves between two types of models.



(a) Study the loss value and PSNR value of the research model

(b) Loss and PSNR values of DeepLab v3+model

Figure 12(a) shows the research model achieves training stability at iteration 50 (loss ≈0.8) with PSNR converging to 30.9 by iteration 100 without degradation. In contrast, DeepLab v3+ (Figure 12(b)) stabilizes at higher loss (3) and lower PSNR (22). To evaluate DOF enhancement across diverse defects, eight representative samples (Figures A–H) from the Bow Network dataset—covering overexposure, motion blur, color casts, and low-light scenarios—are tested using Information Entropy (IE) metrics, with optimization gains shown in Figure 13.

The original image's IE is displayed in Figure 13(a). The information entropies of Figures A to H are 7.19, 7.15, 7.22, 7.03, 6.94, 7.17, 7.29, and 6.92, respectively. The average IE is 7.12. Figure 13(b) shows the IE of the DOF image. The information entropies for Figures A to H are 7.26, 7.18, 7.28, 7.08, 7.26, 7.22, 7.30, and 7.18, respectively. The average IE is 7.22. Figure H demonstrates that adjusting the foreground-background contrast through the model increases IE from 6.92 to 7.18, validating the model's ability to optimize global DOF features. In summary, the depth-optimized model integrating unsupervised CycleGAN effectively enhances the information richness of bow-and-net photography images, providing higher-quality input data for subsequent bow-and-net defect detection tasks. In order to comprehensively evaluate the practical engineering application potential of the model and compensate for the lack of efficiency evaluation, based on the standardized experimental environment mentioned above, this study compared the training and inference costs of the proposed model with the baseline model DeepLab v3+from four core dimensions: parameter quantity, training time, inference delay, and video memory occupation. The results are shown in Table 4.

Table 4

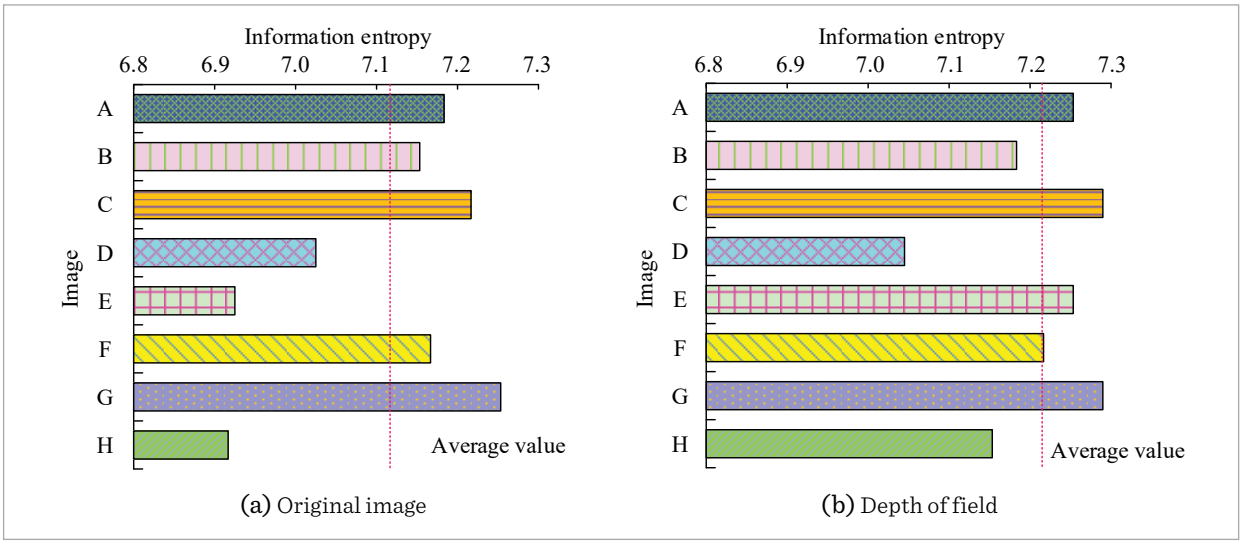
Comparison of Training and Inference Costs Between Models.

/	Ours	DeepLab v3+
Parameter quantity (M)	3.4	28.2
Total training time (h)	23.1	43.6
Single round iteration time (s)	8.3	15.7
Inference delay (ms)	24	58
Inference video memory usage (GB)	1.8	4.5

According to Table 4, the depth of field optimization model for photographic images fused with unsupervised CycleGAN achieves significant cost optimization while maintaining performance advantages. The parameter count is only 12.1% of DeepLab v3+, significantly reducing the pressure on model storage and deployment; During the training phase, both

Figure 13

Information entropy of the original image and the DOF transformation image



the single round iteration time and total time were reduced by about 47%, improving the efficiency of model iteration optimization; The inference delay and video memory usage have been reduced by 58.6% and 60.0% respectively, meeting the real-time processing and hardware resource adaptation requirements in industrial inspection and other scenarios. The comparative results indicate that the proposed model achieves a better balance between performance and efficiency, with stronger scalability and engineering implementation value.

4. Discussion

Experimental results on the AAPM-Mayo dataset show that the proposed algorithm outperforms DCP, BCUD, and RED-CNN across PSNR, SSIM, VIF, and IFC metrics. In attention information processing, it achieves an SSIM of 33.6 and a PSNR of 0.913, highlighting its strength in structural preservation. The average VIF reaches 0.391—a 6.2% improvement over RED-CNN—while the average IFC is 2.85, indicating better information retention. These findings are consistent with prior work in medical image segmentation [5]. On the SOTS-Indoor dataset, the algorithm reduces BRISQUE by 10.2% and achieves a lower NIQE than competitors, demonstrating enhanced sharpness and naturalness, particularly in depth-of-field optimization, as supported by related studies [24]. Computationally, the algorithm maintains high efficiency with only 3.4M parameters — far fewer than RED-CNN — while achieving 27 GFLOPS and 24 ms latency, significantly better than RED-CNN (43 GFLOPS, 35 ms). This lightweight design aligns with approaches like FreqGAN [25], which also balances performance and efficiency through streamlined network architecture.

Compared to DeepLab v3+, the proposed model incorporates a CAM module that accelerates convergence by 50 epochs via foreground probability weight matrices delivering focused gradient signals, enabling rapid key feature capture within the first 50 epochs. Evaluated on eight typical DOF defect samples from the BowNet dataset, the model migrates high-DOF features from landscape images via unsupervised learning while learning color correction from augmented references, improving average IE from 7.12 to 7.22. Specific gains include low-light samples (Figure E: 6.94→7.26) with enhanced carbon fiber texture and background depth, and blurred foreground samples (Figure H: 6.92→7.18) with optimized contrast, validating robust adaptability to diverse DOF defects.

5. Summary and future work

To address foreground-background blurring, low fidelity, and inefficiency in conventional CycleGAN, an optimized unsupervised DOF effect model is proposed. Experiments on AAPM-Mayo and SOTS-Indoor datasets demonstrate superior performance over DCP, BCUD, and RED-CNN across PSNR, SSIM, VIF, and IFC metrics. The model achieves SSIM of 33.6, average VIF of 0.391, and BRISQUE of 19.3, while maintaining computational efficiency (3.4M parameters, 27G FLOPs, 24ms latency) suitable for real-time applications. Training converges at 150 iterations with test PSNR stabilizing at 30.9, effectively optimizing typical DOF defects with average IE improved from 7.12 to 7.22. This model effectively balances DOF enhancement, efficiency, and visual quality for photographic post-processing. However, limitations persist in handling severe noise or ultra-high-resolution images, necessitating future improvements in robustness and adaptability.

References

1. Ahmed, M. F., Pervin, N., Bhowmik, S., Rashid, M. A., Kuwana, A., Kobayashi, H. Design and Implementation of Remote Controlling System Using GAN In Optical Camera Communication. IEEE Photonics Journal, 2023, 15(2), 1-10. https://doi.org/10.1109/JPHOT.2023.3250455
2. Bhosle, K., Musande, V. Evaluation of Deep Learning CNN Model for Recognition of Devanagari Digit. Artificial Intelligence Applications, 2023, 1(2), 114-118. https://doi.org/10.47852/bonviewAIA3202441
3. Chen, G., Xie, H., Rao, X., Liu, X., Otikovs, M., Fryd-

man, L., Liu, C. MRI Motion Correction Through Disentangled CycleGAN Based on Multi-Mask K-Space Subsampling. IEEE Transactions on Medical Imaging, 2025, 44(4), 1907-1921. https://doi.org/10.1109/TMI.2024.3523949

4. Chen, Y., Shi, J., Shen, C., Huang, W., Zhu, Z. CycleGAN with Momentum Control for Intelligent Bearing Fault Diagnosis Driven by Dynamic Models. IEEE Sensors Journal, 2024, 24(23), 39315-39324. https://doi.org/10.1109/JSEN.2024.3471789
5. Chen, Z., Bian, Y., Shen, E., Fan, L., Zhu, W., Shi, F. Moment-Consistent Contrastive CycleGAN for Cross-Domain Pancreatic Image Segmentation. IEEE Transactions on Medical Imaging, 2025, 44(1), 422-435. https://doi.org/10.1109/TMI.2024.3447071
6. Dirvanauskas, D., Maskeliūnas, R., Raudonis, V., Damaševičius, R., Scherer, R. Hemigen: Human Embryo Image Generator Based on Generative Adversarial Networks. Sensors, 2019, 19(16), 3578. https://doi.org/10.3390/s19163578
7. Grigas, O., Maskeliūnas, R., Damaševičius, R. Improving Structural MRI Preprocessing with Hybrid Transformer GANs. Life, 2023, 13(9), 1893. https://doi.org/10.3390/life13091893
8. Guo, L., Huang, S., Liu, H., Wen, B. Toward Robust Image Denoising via Flow-Based Joint Image and Noise Model. IEEE Transactions on Circuits and Systems for Video Technology, 2024, 34(7), 6105-6115. https://doi.org/10.1109/TCSVT.2023.3345667
9. He, Y., Seng, K. P., Ang, L. M. CycleGAN*: Collaborative AI Learning with Improved Adversarial Neural Networks for Multimodalities Data. IEEE Transactions on Artificial Intelligence, 2024, 5(11), 5616-5629. https://doi.org/10.1109/TAI.2024.3432856
10. He, Z., Ran, W., Liu, S., Li, K., Lu, J., Xie, C. Low-Light Image Enhancement with Multi-Scale Attention and Frequency-Domain Optimization. IEEE Transactions on Circuits and Systems for Video Technology, 2024, 34(4), 2861-2875. https://doi.org/10.1109/TCSVT.2023.3313348
11. Jiang, K., Wang, Q., An, Z., Wang, Z., Zhang, C., Lin, C.-W. Mutual Retinex: Combining Transformer and CNN for Image Enhancement. IEEE Transactions on Emerging Topics in Computational Intelligence, 2024, 8(3), 2240-2252. https://doi.org/10.1109/TETCI.2024.3369321
12. Lauenburg, L., Lin, Z., Zhang, R., dos Santos, M., Huang, S., Arganda-Carreras, I. 3D Domain Adaptive Instance Segmentation via Cyclic Segmentation GANs. IEEE Journal of Biomedical and Health Informatics, 2023, 27(8), 4018-4027. https://doi.org/10.1109/JBHI.2023.3281332
13. Li, J., Yan, S., Liu, X., Zhang, J., Zhuo, L. CycFormer: Unsupervised Rain Removal Network Based on CycleGAN and Transformer. IEEE Transactions on Intelligent Transportation Systems, 2025, 26(7), 10596-10611. https://doi.org/10.1109/TITS.2025.3548157
14. Lin, D., Tondi, B., Li, B., Barni, M. A CycleGAN Watermarking Method for Ownership Verification. IEEE Transactions on Dependable and Secure Computing, 2025, 22(2), 1040-1054. https://doi.org/10.1109/TDSC.2024.3424900
15. Lin, L., Wang, T., Liu, H., Zhu, C., Chen, J. Toward Quantifiable Face Age Transformation Under Attribute Unbias. IEEE Transactions on Circuits and Systems for Video Technology, 2024, 34(11), 11768-11782. https://doi.org/10.1109/TCSVT.2024.3422661
16. Liu, C., Wang, Z., Birch, P., Wang, X. Efficient Retinex-Based Framework for Low-Light Image Enhancement Without Additional Networks. IEEE Transactions on Circuits and Systems for Video Technology, 2025, 35(5), 4896-4909. https://doi.org/10.1109/TCSVT.2024.3520802
17. Liu, D., Chen, Y., Wu, Z. Digital Twin (DT)-CycleGAN: Enabling Zero-Shot Sim-to-Real Transfer of Visual Grasping Models. IEEE Robotics and Automation Letters, 2023, 8(5), 2421-2428. https://doi.org/10.1109/LRA.2023.3254460
18. Liu, Y., Li, H., Hu, C., Luo, S., Luo, Y., Chen, C. W. Learning to Aggregate Multi-Scale Context for Instance Segmentation in Remote Sensing Images. IEEE Transactions on Neural Networks and Learning Systems, 2025, 36(1), 595-609. https://doi.org/10.1109/TNNLS.2023.3336563
19. Martin, B., Edwards, K., Jeffrey, I., Gilmore, C. Experimental Microwave Imaging System Cal-

ibration via Cycle-GAN. IEEE Transactions on Antennas and Propagation, 2023, 71(9), 7491-7503. https://doi.org/10.1109/TAP.2023.3296915

20. Que, Y., Xiong, L., Wan, W., Xia, X., Liu, Z. Denoising Diffusion Probabilistic Model for Face Sketch-to-Photo Synthesis. IEEE Transactions on Circuits and Systems for Video Technology, 2024, 34(10), 10424-10436. https://doi.org/10.1109/TCSVT.2024.3409184
21. Tang, G., Ni, J., Chen, Y., Cao, W., Yang, S. X. An Improved CycleGAN-Based Model for Low-Light Image Enhancement. IEEE Sensors Journal, 2024, 24(14), 21879-21892. https://doi.org/10.1109/JSEN.2023.3296167
22. Wang, H., Che, Z., Yang, Y., Wang, M., Xu, Z., Qiao, X. RDFC-GAN: RGB-Depth Fusion CycleGAN for Indoor Depth Completion. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024, 46(11), 7088-7101. https://doi.org/10.1109/TPAMI.2024.3388004
23. Wang, W., Cui, Z. X., Cheng, G., Cao, C., Xu, X., Liu, Z., Zhu, Y. A Two-Stage Generative Model with CycleGAN and Joint Diffusion for MRI-Based Brain Tumor Detection. IEEE Journal of Biomedical and Health Informatics, 2024, 28(6), 3534-3544. https://doi.org/10.1109/JBHI.2024.3373018
24. 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 Transactions on Industrial Informatics, 2023, 19(12), 11742-11751. https://doi.org/10.1109/TII.2023.3252410
25. Wang, Z., Zhang, Z., Qi, W., Yang, F., Xu, J. FreqGAN: Infrared and Visible Image Fusion via Unified Frequency Adversarial Learning. IEEE Transactions on Circuits and Systems for Video Technology, 2025, 35(1), 728-740. https://doi.org/10.1109/TCSVT.2024.3460172
26. Wu, T., Wu, W., Yang, Y., Fan, F.-L., Zeng, T. Retinex Image Enhancement Based on Sequential Decomposition with a Plug-and-Play Framework. IEEE Transactions on Neural Networks and Learning Systems, 2024, 35(10), 14559-14572. https://doi.org/10.1109/TNNLS.2023.3280037
27. Xue, B., Zheng, Q., Li, Z., Wang, J., Mu, C., Yang, J., Li, X. Perturbation Defence Ultra High-Speed Weak Target Recognition. Engineering Applications of Artificial Intelligence, 2024, 138, 109420. https://doi.org/10.1016/j.engappai.2024.109420
28. Yang, X., Fu, Q., Elhoseiny, M., Heidrich, W. Aberration-Aware Depth-from-Focus. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2025, 47(9), 7268-7278. https://doi.org/10.1109/TPAMI.2023.3301931
29. Zhang, Z., Liu, C., Wang, X., Han, Z., Yang, G., Wang, C. DLP-Fusion: Depth of Field, Light Source, and Polarization Fusion Toward Intelligent Optical Imaging for Complex Scenes. IEEE Transactions on Circuits and Systems for Video Technology, 2024, 34(9), 8266-8280. https://doi.org/10.1109/TCSVT.2024.3393608
30. Zhou, J., Li, Y., Qin, H., Dai, P., Zhao, Z., Hu, M. Sonar Image Generation by MFA-CycleGAN for Boosting Underwater Object Detection of AUVs. IEEE Journal of Oceanic Engineering, 2024, 49(3), 905-919. https://doi.org/10.1109/JOE.2024.3350746
31. Zhou, P., Shi, L., Liu, X., Jin, J., Zhang, Y., Hou, J. Light Field Depth Estimation via Stitched Epipolar Plane Images. IEEE Transactions on Visualization and Computer Graphics, 2024, 30(10), 6866-6879. https://doi.org/10.1109/TVCG.2023.3344132
32. Zhou, Z., Chen, Y., He, A., Que, X., Wang, K., Yao, R. NKUT: Dataset and Benchmark for Pediatric Mandibular Wisdom Teeth Segmentation. IEEE Journal of Biomedical and Health Informatics, 2024, 28(6), 3523-3533. https://doi.org/10.1109/JBHI.2024.3383222
33. Zhu, H., Guo, Y., Xu, W., You, X. Semi-Supervised MIMO Detection Using Cycle-Consistent Generative Adversarial Network. IEEE Transactions on Cognitive Communications and Networking, 2023, 9(5), 1226-1240. https://doi.org/10.1109/TCCN.2023.3279260