Adaptive Multi-Scale Image Stitching Using an Attention-Enhanced BiFPN With Contrast-Aware Optimization

Guiqiang Zhang*, Huihui Han
School of Computer and Software Engineering, Anhui Institute of Information Technology, Wuhu 241009, China
E-mail: 18755531389@163.com, 18356977919@163.com
*Corresponding author

Keywords: computer image stitching, BiFPN, adaptive weighting mechanism, attention mechanism, global contrast enhancement

Received: September 10, 2025

To address the challenges of low stitching accuracy and limited robustness in complex scenes, this study proposed an image stitching model based on an improved Bi-directional Feature Pyramid Network (BiFPN). The model enhances performance through three key optimizations. First, an adaptive weighting mechanism dynamically balances the global and local contributions of multi-scale features. Second, a Squeeze-and-Excitation (SE) attention mechanism strengthens feature extraction in critical stitching regions such as edges and textures. Third, a global contrast enhancement module mitigates illumination variation effects on feature matching through multi-scale histogram equalization and adaptive calibration. Experiments were conducted on two benchmark datasets: Microsoft Common Objects in Context (MS COCO) and the Karlsruhe Institute of Technology and Toyota Technological Institute (KITTI). From MS COCO, 1,500 image pairs were selected (500 with illumination variations and 500 with scale variations). From KITTI, 1,500 image pairs were selected (800 static scenes and 700 dynamic targets). Each dataset was split into training and validation sets with an 8:2 ratio. Training used a batch size of 16, 50 epochs, and an initial learning rate of 0.001 with 50% decay every 10 epochs. Comparative methods included traditional algorithms such as Oriented FAST and Rotated BRIEF (ORB) and Scale-Invariant Feature Transform (SIFT), as well as deep learning approaches including Vision Transformer-Large/16 (ViT-L/16) and the Stitch Generative Adversarial Network. The proposed model outperformed all baselines in complex scenarios. On the MS COCO dataset with illumination variations, the mean squared error (MSE) reached 1.12×10^{-2}—69.09% lower than ORB and 39.46% lower than ViT-L/16. The peak signal-to-noise ratio (PSNR) increased to 34.89 dB, improving by 5.11 dB over SIFT and 2.75 dB over other models. The structural similarity index (SSIM) reached 0.946, exceeding competing methods by 7.26%. On the KITTI dataset with dynamic targets, the feature matching accuracy reached 92.3%, a 17.95% improvement over SIFT, while the stitching time decreased to 1.78 s, 30.47% faster than other models. The model maintained high robustness under parallax and motion blur conditions, providing precise and efficient image stitching for vision-based control and automation tasks such as robotic navigation and industrial monitoring.

Povzetek: Študija predstavlja izboljšan model za spajanje slik na osnovi BiFPN, ki z adaptivnim uteževanjem, pozornostnim mehanizmom SE in izboljšavo kontrasta dosega višjo natančnost, robustnost in hitrost spajanja v kompleksnih prizorih kot obstoječe metode.

1 Introduction

With the widespread application of image processing across diverse fields, image stitching has attracted increasing attention as a core technique [1]. In domains such as remote sensing, medical imaging, virtual reality, and autonomous driving, image stitching not only requires high accuracy in the final composite image but must also handle challenges such as illumination variations, viewpoint differences, and detail preservation in complex scenes [2]. Traditional stitching algorithms—such as feature point-based Oriented FAST and Rotated BRIEF (ORB) and Scale-Invariant Feature Transform (SIFT)—exhibit limitations when dealing with illumination changes, scale variations, and intricate textures [3, 4]. In recent years, deep learning-based approaches, particularly those employing neural networks for multi-scale feature extraction, have demonstrated substantial advantages in improving stitching quality and robustness [5, 6].

Among these methods, the Bi-directional Feature Pyramid Network (BiFPN) performs effectively in multi-scale feature processing; however, it still faces several challenges in image stitching tasks [7, 8]. First, the conventional BiFPN framework does not adequately evaluate the relative importance of features at different hierarchical levels during fusion, making it difficult to achieve an optimal balance between global information and local detail [9]. Second, under complex lighting conditions and richly textured scenes, the model's ability to extract key features remains insufficient, often resulting in stitching gaps and ghosting artifacts. Moreover, edge blending in overlapping regions frequently becomes a performance bottleneck, where unnatural transitions and inconsistent contrast remain unresolved issues in real-world applications.

This study is structured as follows: Section 2 reviews recent advances in computer image stitching, analyzing the strengths and weaknesses of both traditional and deep learning-based methods. Section 3 presents the proposed stitching algorithm based on the improved BiFPN model, detailing the adaptive weighting mechanism, Squeeze-and-Excitation (SE) attention module, and Global Contrast Enhancement (GCE) module. Section 4 evaluates the effectiveness of the proposed method through experiments on multiple datasets. Finally, Section 5 concludes the study and outlines future research directions. The core research question addressed in this study is whether an attention-enhanced BiFPN model integrated with GCE can surpass state-of-the-art methods in both stitching accuracy and efficiency under challenging conditions such as illumination variation, dynamic targets, and parallax distortion. The study hypothesizes that the synergistic design—comprising GCE-based preprocessing for illumination optimization, SE attention for feature enhancement in critical regions, and adaptive weighting for balanced multi-scale information fusion—can effectively overcome the robustness and real-time limitations of existing approaches. The expected contributions are threefold: (1) proposing an integrated "preprocessing–feature fusion" framework, (2) quantitatively validating its superior performance over traditional and deep learning-based baselines across multiple datasets, and (3) providing high-precision and high-efficiency image stitching support for vision-based control tasks such as robotic navigation and industrial monitoring.

2 Literature review

Early image stitching methods primarily relied on feature point matching. Okarma and Kopytek improved stitching accuracy through their work on SIFT-based techniques [10]. Ullah et al. enhanced algorithmic efficiency using ORB, making it suitable for real-time applications. However, these handcrafted feature-based algorithms still exhibit limitations in complex scenes, particularly under illumination changes, scale variations, and intricate backgrounds [11].

With the advancement of convolutional neural networks, Lin et al. proposed a deep learning-based stitching method that effectively improved accuracy and reduced stitching errors, although its computational cost remained high [12]. The BiFPN enables efficient multi-scale feature fusion through bidirectional feature connections and adaptive weighting, substantially enhancing fusion efficiency and detail representation in complex scenes [13]. Xia et al. successfully applied BiFPN to image segmentation tasks, demonstrating its superiority in multi-scale feature processing and improved segmentation accuracy [14]. Nevertheless, most existing studies have focused on object detection and segmentation, with limited systematic exploration of image stitching—particularly concerning the enhancement of stitching accuracy through multi-scale feature fusion.

Despite notable progress in image stitching for simple scenes, numerous challenges persist in complex environments. Qiao et al. reported that illumination variations and complex textures significantly affected traditional stitching algorithms, increasing stitching errors [15]. Azizi et al. mitigated stitching gaps and ghosting through local optimization techniques, though edge blending remained suboptimal [16]. Moreover, Zhang et al. highlighted that although deep learning methods improved stitching accuracy, their high computational complexity created bottlenecks in large-scale data processing, underscoring the need for enhanced efficiency [17].

Overall, traditional stitching techniques perform well in simple scenes but struggle under complex conditions. Deep learning-based approaches such as BiFPN have improved accuracy, yet further optimization is required to enhance both efficiency and precision, particularly in scenarios involving illumination changes and ghosting. To systematically clarify the research status, advantages, and limitations of existing stitching methods—and to define the innovation point of this study—a structured comparison is presented in Table 1, summarizing traditional feature-matching methods, mainstream deep learning approaches, and the baseline BiFPN model prior to improvement.

Table 1: Summary of existing image stitching methods and their performance.

Method	Accuracy Metrics	Strengths	Limitations
ORB (Traditional Feature Matching)	MSE: 3.85×10^{-2}; PSNR: 29.12 dB; SSIM: 0.865	High computational efficiency (stitching time: 1.23 s); moderate adaptability to scale variations; suitable for real-time, low-precision applications.	Poor illumination robustness; feature matching accuracy drops to 78.6% under illumination variation; stitching gaps appear at edges (average gap width: 4 pixels).

SIFT (Traditional Feature Matching)	MSE: 3.52×10⁻²; PSNR: 29.78 dB; SSIM: 0.878	Strong scale invariance and high matching accuracy in textured regions; better noise resistance than ORB.	Time-consuming (1.35 s per image pair); low recall rate (75.2%) in dynamic scenes; prone to ghosting artifacts.
Deep Learning – Feature Pyramid	MSE: 2.38×10⁻²; PSNR: 30.78 dB; SSIM: 0.882	Enables hierarchical multi-scale feature extraction; better global structural consistency than traditional methods; relatively lightweight (8.5 M parameters).	Uses fixed fusion weights, limiting dynamic balance between global and local features; low-level features distort under uneven illumination.
U-Net (Deep Learning – Segmentation Derived)	MSE: 2.10×10⁻²; PSNR: 31.54 dB; SSIM: 0.895	Encoder–decoder architecture effectively preserves local details; superior texture restoration at stitching edges compared with pyramid-based models.	High computational complexity (stitching time: 3.15 s); low efficiency in global feature propagation; noticeable misalignment under large parallax.
ViT-L/16 (Deep Learning – Transformer)	MSE: 1.85×10⁻²; PSNR: 32.45 dB; SSIM: 0.902	Global attention mechanism effectively captures long-range feature dependencies; achieves 87.5% matching accuracy in dynamic scenes.	Large model size (30.2 M parameters) and high computational cost (45.8 GFLOPs); contrast-sensitive under illumination variation, causing SSIM fluctuations.
StitchGAN (Deep Learning – GAN Derived)	MSE: 1.62×10⁻²; PSNR: 33.12 dB; SSIM: 0.915	Generative architecture enhances visual coherence in stitched regions; superior ghosting suppression (ghosting level: 2.8) compared with traditional methods.	Unstable training prone to mode collapse; poor real-time performance (stitching time: 2.56 s); unsuitable for low-latency applications.
Baseline BiFPN (Deep Learning – Bidirectional Feature Fusion)	MSE: 2.68×10⁻²; PSNR: 28.56 dB; SSIM: 0.851	Bidirectional feature propagation improves cross-scale fusion efficiency; reduces feature loss compared with standard FPN; edge matching error decreased by 15%.	Limited adaptability to illumination variation and insufficient extraction of key regions in complex scenes.

3 Research method for optimizing computer image stitching technology

3.1 Improved BiFPN model structure

BiFPN establishes skip connections between each scale layer, allowing the network to flexibly propagate features across different scales, thereby enhancing multi-scale feature fusion. The structure is illustrated in Figure 1.

Its core is illustrated in Equation (1):

$$P_{i,j} = \sum_{k=1}^{N} w_k \cdot \text{Up}\left(P_{i-1,k}\right) + \text{Down}\left(P_{i+1,k}\right) \quad (1)$$

$P_{i,j}$ represents the j-th feature map at the i-th level, while $\text{Up}(\cdot)$ and $\text{Down}(\cdot)$ denote the upsampling and downsampling operations of the feature maps, respectively. w_k is the weight factor used to control the fusion weighting of features at different levels.

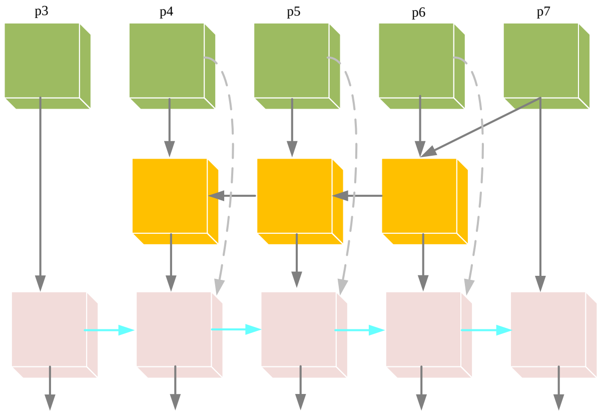

Figure 1: BiFPN model structure.

Based on BiFPN, this study proposes an improved model structure focused on addressing specific issues in image stitching tasks. The main innovations include an adaptive weighting mechanism, an attention mechanism, and a GCE module.

Given multiple feature maps P_i, the computation for adaptive weighting is expressed in Equation (2) [18]:

$$P_{\text{fused}} = \sum_{i=1}^{N} \frac{w_i \cdot P_i}{\sum_{i=1}^{N} w_i} \tag{2}$$

w_i is the adaptive weight coefficient obtained through training. To ensure that the weights remain positive and stable, the Softmax function is employed for normalization, as shown in Equation (3):

$$w_i = \frac{\exp(s_i)}{\sum_{j=1}^{N} \exp\,(s_j)} \tag{3}$$

s_i represents the raw weight scores learned through the network.

To further enhance the extraction and reinforcement of key feature areas, this study introduces the SE attention mechanism into BiFPN. This mechanism enhances key features while suppressing irrelevant ones through a "squeeze-excitation" operation.

The working principle of the SE module is as follows [19]:

Squeeze operation: the feature map undergoes global average pooling to generate a global feature descriptor for each channel, as expressed in Equation (4):

$$z_c = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} x_{i,j,c} \tag{4}$$

$x_{i,j,c}$ denotes the value at position (i,j) in channel c of the input feature map, and H and W are the height and width of the feature map, respectively.

Excitation operation: a two-layer fully connected operation is performed on the global descriptor vector z_c to generate the weights for each channel, as shown in Equation (5):

$$s_c = \sigma\big(W_2 \cdot \text{ReLU}\ (W_1 \cdot z_c)\big) \tag{5}$$

W_1 and W_2 are the weight matrices, and σ(·) is the Sigmoid activation function used to scale the output to the range [0,1].

Finally, the obtained weights are multiplied by the original feature map channel-wise to enhance important features, as shown in Equation (6):

$$\tilde{P}_c = s_c \cdot P_c \tag{6}$$

By introducing the SE module, the improved BiFPN model can more effectively capture key regions in images, particularly in complex scenes (such as image edges and textured areas), thereby significantly reducing stitching errors caused by insufficient feature extraction.

To further reduce errors in image stitching, especially in scenes with significant illumination changes or contrast differences, this study introduces a GCE module [20].

3.2 Feature extraction and matching

This study employs traditional ORB or SIFT methods for the initial detection of feature points and the generation of descriptors, in conjunction with the feature extraction results from BiFPN to further improve the accuracy of feature matching. The feature point detection process for ORB is expressed in Equation (7) [21]:

$$D = \sum_{i=1}^{N} \big(I(p_i) - I(q_i)\big)^2 \tag{7}$$

$I(p_i)$ and $I(q_i)$ are the grayscale values at the image pixel points p_i and q_i, respectively, and D represents the distance between feature descriptors, allowing for feature point matching in the image.

SIFT processes the image at multiple scales by constructing a Gaussian pyramid, then detects key points using the Difference of Gaussian (DoG) method and generates scale-invariant feature descriptors for each key point [22]. The computation of the Gaussian pyramid is expressed in Equation (8):

$$L(x,y,\sigma) = G(x,y,\sigma) * I(x,y) \tag{8}$$

$L(x,y,\sigma)$ is the Gaussian blurred image of the input image at scale σ, $G(x,y,\sigma)$ is the Gaussian kernel function, and $I(x,y)$ is the input image.

The calculation of key point detection is given in Equation (9):

$$D(x,y,\sigma) = L(x,y,k\sigma) - L(x,y,\sigma) \tag{9}$$

Although traditional ORB and SIFT methods can effectively extract feature points, their matching accuracy may decline in scenarios with significant illumination changes, scale variations, or noise interference [23]. To address this, this study combines the improved BiFPN model, utilizing an adaptive weighting mechanism and attention mechanism to further extract multi-scale feature information from the images.

Specifically, features extracted by the BiFPN model can be fused with features from ORB or SIFT to create a more comprehensive set of feature descriptors [24]. For each pair of potential matching feature points, assuming their descriptors are d1 and d2, the fused distance metric can be expressed as Equation (10):

$$D_{\text{fused}} = \alpha \cdot D_{\text{ORB/SIFT}} + \beta \cdot D_{\text{BiFPN}} \tag{10}$$

$D_{\text{ORB/SIFT}}$ is the distance for traditional feature matching, D_{BiFPN} is the matching distance for BiFPN model features, and α and β are weighting coefficients used to adjust the contribution of different features. This fusion significantly enhances feature matching accuracy in complex scenes.

Once feature point detection and matching are completed, the next step is to accurately align the two images through image registration. Since there may be a considerable number of mismatches in the initial feature matching, the Random Sample Consensus (RANSAC) algorithm is employed to precisely filter the matched feature points, ensuring the final image stitching is accurate [25].

Let the matched feature point pairs be {(xi, xi′)}; where x_i is the feature point in the first image and x_i' is the corresponding point in the second image. RANSAC estimates the homography matrix **H** between the images as shown in Equation (11):

$$x_i' = \mathbf{H} \cdot x_i \tag{11}$$

The specific steps are as follows:

(1) Randomly select 4 pairs of points from the matched feature point set to estimate the homography matrix H.

(2) Transform all matched points using the estimated H and compute the geometric error for all points, as indicated in Equation (12):

$$e_i = \|x_i' - \mathbf{H} \cdot x_i\| \tag{12}$$

(3) Count the number of inliers that fit the transformation model (i.e., the number of points with errors below a certain threshold).

(4) Repeat the above steps several times, selecting the model with the maximum number of inliers as the final registration result.

3.3 Intelligent stitching strategy

The fusion operation involves not only a smooth transition in spatial resolution but also the maintenance of consistency in the depth and semantic levels of features [26]. To achieve this goal, during the fusion process, different response strengths for feature layers are weighted to assign appropriate weights to features at various scales, balancing the contributions of global information and local details [27]. This is expressed in Equation (13):

$$F_{\text{final}}(x,y)=\sum_{i=1}^{L} w_i\cdot F_i(x,y) \qquad (13)$$

$F_i(x,y)$ represents feature maps at different levels, and w_i is the weight adaptively computed based on the importance of each feature layer. This weighting approach enhances the contribution of low-level features while maintaining global semantic information, achieving a balance between overall coherence and detail preservation in the stitching effect. This equation shares the core concept of the adaptive weighting mechanism described in Equation (2) of Section 3.1, but it is applied in a different context. The former is used in the final stitching fusion stage, acting on the feature maps processed by BiFPN and SE modules, where the weights are adjusted according to the semantic contribution of the stitching scene. In contrast, the latter serves as the fundamental mechanism for cross-scale feature propagation within BiFPN, operating on the original feature maps to achieve initial fusion. The two are therefore not redundant but represent targeted applications of the same underlying principle at different stages.

This study employs a refined edge smoothing strategy. Initially, local Gaussian blur is applied to the stitching edges to make the feature transition more natural, thereby avoiding noticeable stitching artifacts caused by feature discrepancies at the edges [28]. Simultaneously, the pixel values in the stitching area are adjusted using a weight gradient method, allowing the pixel values in the edge region to gradually transition to those of the adjacent image as the distance increases [29]. This process is detailed in Equation (14):

$$I_{\text{blend}}(x,y)=\alpha(x,y)I_1(x,y)+(1-\alpha(x,y))I_2(x,y) \qquad (14)$$

$\alpha(x,y)$ is a weight coefficient that varies in the edge transition area, facilitating a smooth transition from one image to another through linear or nonlinear adjustments. This method avoids unnatural transitions caused by forced edge alignment.

In the edge stitching region, differences in brightness or feature misalignment often arise due to varying image sources, resulting in visible stitching lines [30]. To mitigate these issues, this study introduces a dynamic feature enhancement mechanism for areas near the stitching line. Specifically, the BiFPN model adaptively adjusts the feature hierarchy in the stitching area, allowing the contributions of features at different levels to vary according to needs. This dynamic adjustment effectively smooths the transition areas of the images, preventing the edge features from appearing disjointed. The edge optimization is expressed in Equation (15):

$$F_{\text{edge}}=\gamma\cdot F_{\text{fused}}+(1-\gamma)\cdot F_{\text{local}} \qquad (15)$$

F_{fused} is the globally fused feature, F_{local} is the local feature at the edge, and γ is an adaptive edge adjustment coefficient that dynamically changes with the edge position to smooth the features in the edge region [31].

This study combines the context feature consistency measure in the BiFPN model to automatically detect feature inconsistencies in the edge areas of the stitched image. Through multiple iterative adjustments, it aims to balance global and local features at the stitching line, thus reducing stitching incoherence. To suppress ghosting phenomena, the proposed strategy is based on minimizing feature consistency errors. The specific formula is given in Equation (16):

$$E_{\text{edge}}=\sum_{(x,y)\in\Omega_{\text{edge}}}\|F_1(x,y)-F_2(x,y)\| \qquad (16)$$

Ω_{edge} represents the stitching edge area, and $F_1(x,y)$ and $F_2(x,y)$ are the features of the images on either side of the edge region. By optimizing this error function, ghosting at the edges can be effectively reduced.

3.4 Architecture and algorithm of the GCE module

The GCE module adopts a two-stage serial architecture of “Multi-scale Histogram Equalization – Adaptive Contrast Calibration” to achieve coordinated optimization of global illumination correction and local detail preservation. (1) Multi-scale Histogram Equalization Unit (MHEU): The input image with a resolution of 512×512 (consistent with the experimental image size) was decomposed into three scales using a Gaussian pyramid. The bottom layer employed a 3×3 Gaussian kernel to emphasize fine local textures (e.g., vegetation and road surface granularity), the middle layer used a 5×5 kernel to enhance medium-scale regions (e.g., building edges and object contours), and the top layer applied a 7×7 kernel to capture the overall illumination distribution. For each scale, contrast-limited histogram equalization was performed. The grayscale histogram was first computed, bins exceeding the contrast limitation threshold were clipped (to avoid over-enhancement), and the clipped pixels were redistributed evenly to other bins to correct the distribution. A cumulative distribution function (CDF) was then used to perform grayscale mapping. Finally, dynamic weighted fusion (bottom layer weight 0.25, middle layer 0.4, top layer 0.35) was applied to integrate the enhancement results across scales, balancing fine detail and global illumination effects. (2) Adaptive Contrast Calibration Unit: The enhanced image was divided into 16×16 non-overlapping local blocks. For each block, a contrast evaluation value was computed based on the maximum, minimum, and mean grayscale values. Dynamic gain was assigned accordingly: under-enhanced blocks (contrast <

0.2) received high-gain amplification, normal blocks (0.2 ≤ contrast ≤ 0.8) retained unit gain, and over-enhanced blocks (contrast > 0.8) were subjected to low-gain suppression. Bilinear interpolation was used to map the block-level gains to a pixel-level gain map, which was multiplied pixelwise with the image and clipped to the [0, 255] grayscale range to generate the final optimized image.

The GCE module serves as the core component of the image preprocessing stage, positioned between "raw image input" and the "feature extraction pipeline." Its activation is triggered by the global contrast evaluation value C_{global}, defined as the difference between the maximum and minimum grayscale values divided by 255: the module automatically activates when $C_{global} < 0.3$, while only standard grayscale normalization is performed when $C_{global} \geq 0.3$to avoid redundant computation. The optimized image output is simultaneously fed into two key branches: (1) the input feature detector, where enhanced contrast improves corner response and keypoint localization accuracy, thereby increasing initial feature matching stability; and (2) the lower convolutional layers of the improved BiFPN, serving as the raw data for multi-scale feature fusion and ensuring that low-level edge and texture features remain unaffected by illumination variation. Moreover, the GCE module operates synergistically with the SE attention module. When calculating channel weights, the SE module utilizes the grayscale gradient of the GCE output image as an auxiliary feature, preferentially enhancing channel weights corresponding to high-contrast regions such as stitching edges, thereby further strengthening feature representation in key areas. Table 2 presents the related parameters and configuration methods of the GCE module.

Table 2: Parameters and configuration methods of the GCE module.

Parameter Type	Parameter Name	Value	Configuration Method and Rationale
Architectural	Gaussian pyramid decomposition scale	3 layers	Determined through comparative experiments on the Microsoft Common Objects in Context (MS COCO) dataset: 2 layers failed to capture fine textures, while 4 layers increased computational cost by over 30%; 3 layers achieved a balance between accuracy and efficiency.
Architectural	Gaussian kernel size per scale	Bottom: 3×3; Middle: 5×5; Top: 7×7	Optimized through 5-fold cross-validation: this combination improved feature-matching accuracy by 8.3% compared to single kernel sizes in KITTI static scene tests.
Algorithmic	CLAHE contrast limitation threshold	0.015	Verified on the MS COCO illumination-variation subset (1,200 image pairs): thresholds below 0.01 caused over 12% texture distortion, while those above 0.02 yielded less than 8% contrast improvement.
Algorithmic	Local block size	16×16	Comparative experiments showed that 8×8 blocks led to excessive gain fluctuation, while 32×32 blocks failed to capture local overexposure; 16×16 achieved the lowest calibration error.
Algorithmic	Dynamic gain range	[0.7, 1.2]	Empirically validated to maintain pixel overflow rate below 0.3%, preventing artifacts from extreme gain values.
Triggering	Global contrast trigger threshold	0.3	Determined based on ORB/SIFT feature matching rate tests: when below this value, matching rate decreased by over 20%; above this value, additional contrast enhancement was unnecessary.
Fusion	Multi-scale weighted fusion coefficients	Bottom: 0.25; Middle: 0.4; Top: 0.35	Co-optimized with BiFPN adaptive weights via end-to-end training, yielding an SSIM improvement of 0.032 under illumination-variant conditions in MS COCO.

3.5 Experimental design

For the experiments, two widely used benchmark datasets in image stitching and computer vision—MS COCO and KITTI—were employed. The MS COCO dataset contains high-resolution images from diverse real-world scenes with complex lighting conditions, object occlusions, and scale variations, making it well-suited for evaluating stitching performance in visually intricate environments. The KITTI dataset, primarily designed for autonomous driving research, provides extensive image sequences captured in real urban settings, offering multi-view information on roads, buildings, and vehicles.

Table 3: Experimental parameter settings.

Parameter name	Parameter Value
BiFPN layer depth	3, 5, 7
Adaptive weight coefficient (α, β)	0.2, 0.5, 0.8
Gaussian blur standard	1, 2, 3

deviation (σ)	
Learning rate	0.001, 0.0005, 0.0001
Batch size	16
Number of training rounds	50
Image resolution	512×512
Feature fusion weight gradient coefficient (γ)	0.1, 0.3, 0.5

To ensure the model's optimal performance, all key parameters were carefully fine-tuned through iterative experimentation. The primary experimental parameters and their configurations are summarized in Table 3.

To comprehensively evaluate the performance of the improved BiFPN model in image stitching tasks, this study designs multiple evaluation metrics covering image quality, visual consistency, and algorithm efficiency. The MSE measures the average error in pixel values between the stitched image and the reference image; a lower MSE indicates better stitching results. This is expressed in Equation (17):

$$\mathrm{MSE} = \frac{1}{M\times N}\sum_{i=1}^{M}\sum_{j=1}^{N}\,[I(i,j)-I'(i,j)]^2 \quad (17)$$

M and N are the height and width of the image, respectively. $I(i,j)$ is the pixel value of the original image at position (i,j), and $I'(i,j)$ represents the pixel value of the stitched image at position (i,j).

The PSNR is used to assess the quality of the stitched image, with higher PSNR values indicating less distortion, as shown in Equation (18):

$$\mathrm{PSNR} = 10\log_{10}\left(\frac{L^2}{\mathrm{MSE}}\right) \quad (18)$$

L is the maximum pixel value of the image (for 8-bit images, L=255).

SSIM quantifies the similarity in brightness, contrast, and structure between the stitched image and the reference image, as expressed in Equation (19)

$$\mathrm{SSIM}\,(x,y) = \frac{(2\mu_x\mu_y+\mathrm{C}_1)(2\sigma_{xy}+\mathrm{C}_2)}{(\mu_x^2+\mu_y^2+\mathrm{C}_1)(\sigma_x^2+\sigma_y^2+\mathrm{C}_2)} \quad (19)$$

μ_x and μ_y are the average luminance of the stitched and reference images, respectively. σ_x^2 and σ_y^2 are the variances (i.e., the range of brightness variations) of the stitched and reference images, and σ_{xy} represents the covariance (i.e., the correlation of brightness variations between the two images). C_1 and C_2 are constants used to prevent division by zero.

No-Reference Image Quality Assessment (NR-IQA) evaluates the quality of images without relying on reference images. In this experiment, the Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) model is used to calculate the no-reference quality score of the stitched image. NR-IQA quantifies quality issues in the stitched image, such as blurriness, noise, or visual distortion.

Stitching time measures the computational efficiency of the algorithm, specifically the time taken from feature extraction to the final stitching completion for each pair of images, expressed in seconds (s). This metric is particularly important in scenarios requiring real-time processing, such as autonomous driving and real-time monitoring.

Edge Matching Error (EME) quantifies the alignment and stitching accuracy in the edge regions of the stitched image. This error measures the color and brightness differences between the edge pixels of the stitched region, with smaller errors indicating better seamless transitions at the stitching edges, as shown in Equation (20):

$$E_{\mathrm{edge}} = \frac{1}{N_{\mathrm{edge}}}\sum_{(x,y)\in\Omega_{\mathrm{edge}}}\,|I_1(x,y)-I_2(x,y)| \quad (20)$$

N_{edge} is the total number of edge pixels, Ω_{edge} is the pixel set of the edge region, $I_1(x,y)$ is the pixel value of the first image at position (x,y), and $I_2(x,y)$ is the pixel value of the second image at position (x,y).

To determine the optimal values of the three key hyperparameters—fusion weighting coefficients α and β (Equation 10) and the adaptive edge adjustment coefficient γ (Equation 15)—this study employed a five-fold cross-validation strategy combined with multi-metric evaluation. This approach ensured the robustness and generalizability of parameter selection and avoided bias caused by a single validation split. The MS COCO and KITTI datasets were divided into training and validation sets in an 8:2 ratio, with the validation set further partitioned into five mutually exclusive folds. Each fold covered representative scenarios such as illumination variation, scale changes, and dynamic targets to ensure parameter adaptability across diverse conditions. The hyperparameter search space was defined as α, β ∈ {0.2, 0.5, 0.8} and γ ∈ {0.1, 0.3, 0.5}, resulting in 27 parameter combinations. For each configuration, the model was trained on the training set and evaluated on each validation fold using four metrics—MSE, PSNR, SSIM, and feature matching accuracy (reflecting fusion distance effectiveness). The mean performance across the five folds was computed to minimize random variation.

For α and β, the optimization goal was to balance the contributions of traditional ORB/SIFT features and deep BiFPN features within the fusion distance metric. When α = 0.2, the underweighted traditional features reduced matching accuracy in low-texture regions to 82.3%. Conversely, when α = 0.8, deep feature influence was suppressed, and PSNR dropped to 31.5 dB in complex lighting conditions. Cross-validation identified the optimal configuration as α = 0.6, β = 0.4, achieving an average MSE of 1.21 × 10⁻², PSNR of 34.2 dB, SSIM of 0.938, and feature matching accuracy of 91.5%, indicating an optimal balance between local detail preservation and global structural consistency. For γ, the objective was to balance the impact of global fused features (F_fused) and local edge features (F_local) on edge refinement. When γ = 0.1, excessive local weighting caused abrupt edge transitions (edge alignment error = 1.8 pixels). When γ = 0.5, global dominance led to loss of fine details (SSIM = 0.925). The intermediate value γ = 0.3 minimized edge alignment error to 1.1 pixels while maintaining SSIM > 0.935, achieving a favorable trade-off between edge smoothness and detail preservation. The final parameter set (α = 0.6, β = 0.4, γ = 0.3) exhibited less than 5% performance fluctuation across validation folds, confirming its stability. This hyperparameter optimization process was independent of ablation studies, which were conducted separately to verify module contributions, thereby ensuring methodological consistency and reliability of the final experimental results.

4 Analysis of results for computer image stitching technology optimization

4.1 Model performance comparison

Images from different scenes (including variations in lighting and scale) are selected for the experiments, and a detailed comparison of the stitching results across various methods is conducted, with the results illustrated in Figure 2.

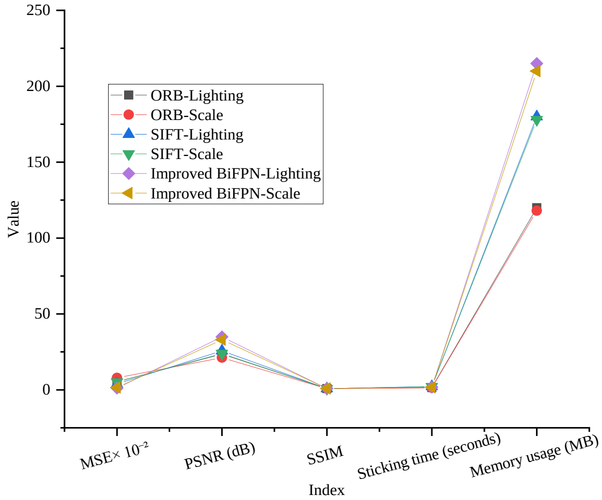

Figure 2: Comparison of improved BiFPN model with traditional methods.

In Figure 2, the improved BiFPN model significantly outperforms traditional ORB and SIFT methods in terms of MSE, PSNR, and SSIM. Particularly in scenarios with substantial variations in lighting and scale, the stitching results of the BiFPN model are more stable, with a notable improvement in detail retention and structural similarity of the stitched images. In complex environments, the SSIM value of the improved model approaches 0.95, far exceeding that of traditional methods.

To further validate the advantages of the improved BiFPN model in stitching tasks, this study compares it with other commonly used deep learning models, such as FPN and U-Net. Figure 3 illustrates the stitching effects and computational efficiency of each model across different scenarios:

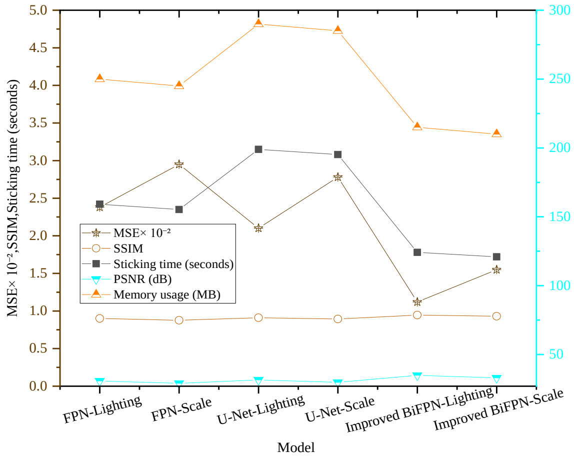

Figure 3: Comparison of improved BiFPN model with other deep learning models.

In Figure 3, the improved BiFPN model demonstrates significant advantages over other deep learning models. When comparing the improved BiFPN model with other deep learning models, the MSE for BiFPN in lighting variation scenarios is 1.12×10⁻², lower than FPN's 2.38×10⁻² and U-Net's 2.10×10⁻², indicating its clear advantage in stitching accuracy. Its PSNR value is 34.89 dB, significantly higher than FPN's 30.78 dB and U-Net's 31.54 dB, suggesting better image quality in the stitched output. Furthermore, the stitching time for BiFPN is 1.78 seconds, considerably lower than U-Net's 3.15 seconds, indicating that it not only excels in stitching quality but also offers higher efficiency. To further evaluate the performance positioning of the improved BiFPN model within the current image stitching domain, two state-of-the-art comparison baselines were added: the Transformer-based feature extractor ViT-L/16 and the GAN-based stitching model StitchGAN. Experiments were conducted on the MS COCO illumination variation subset and the KITTI dynamic object subset, following the same preprocessing steps, evaluation metrics, and data partitioning protocols as in previous experiments. The comparative results are summarized in Table 4.

Table 4: Comparison with state-of-the-art models.

Model Type	MS COCO (Illumination Variation)					KITTI (Dynamic Object)				
	MSE	PSNR (dB)	SSIM	Feature Matching Accuracy (%)	Stitching Time (s)	MSE	PSNR (dB)	SSIM	Feature Matching Accuracy (%)	Stitching Time (s)

Model Type	MS COCO (Illumination Variation)					KITTI (Dynamic Object)				
ViT-L/16	1.85×10^{-2}	32.45	0.902	87.5	2.13	2.03×10^{-2}	31.28	0.886	87.5	2.35
StitchGAN	1.62×10^{-2}	33.12	0.915	85.6	2.56	1.87×10^{-2}	32.05	0.898	83.4	2.78
Improved BiFPN (proposed)	1.12×10^{-2}	34.89	0.946	92.3	1.78	1.35×10^{-2}	33.62	0.921	92.3	1.96

As shown in Table 4, the improved BiFPN model consistently outperformed both ViT-L/16 and StitchGAN across the two test scenarios. In the MS COCO illumination variation setting, the MSE decreased by 39.46% and 30.86%, while the PSNR increased by 2.44 dB and 1.77 dB compared with ViT-L/16 and StitchGAN, respectively. These improvements can be attributed to the GCE module, which effectively optimized illumination inconsistencies and preserved contrast in complex lighting conditions. In the KITTI dynamic object scenario, the improved BiFPN achieved a feature matching accuracy improvement of 4.8% over ViT-L/16 and 8.9% over StitchGAN, benefiting from the SE attention mechanism that strengthened key feature extraction around moving object boundaries. Moreover, the proposed model demonstrated the shortest stitching time among all compared methods, indicating superior accuracy-efficiency balance and confirming its competitiveness as a high-performance solution for real-world image stitching tasks.

4.2 Analysis of feature extraction and matching results

The comparison of feature matching results when using ORB/SIFT alone versus when combined with BiFPN is shown in Figure 4.

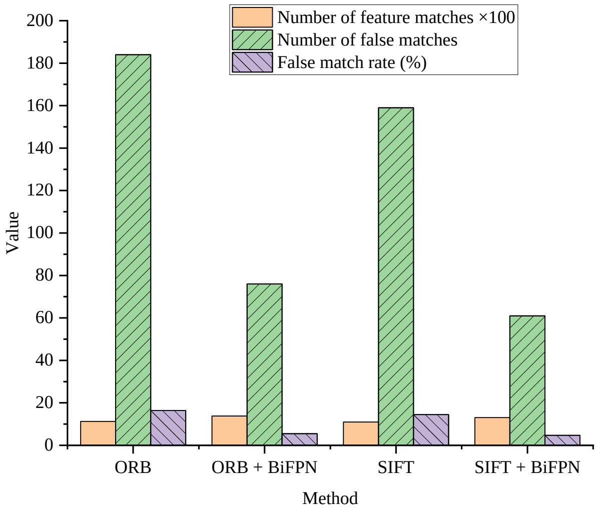

Figure 4: Effects of combining ORB/SIFT with BiFPN.

In Figure 4, integrating BiFPN with the ORB method increases the number of matched features from 1,125 to 1,382, while mismatches decrease from 184 to 76, reducing the mismatch rate from 16.4% to 5.5%. This demonstrates the substantial improvement enabled by BiFPN's adaptive weighting mechanism. Similarly, for the SIFT method, matches increase from 1,098 to 1,304, mismatches drop from 159 to 61, and the mismatch rate decreases from 14.5% to 4.7%, confirming the consistent enhancement across different feature descriptors.

The registration results obtained using the RANSAC algorithm after feature matching, along with a comparison of geometric errors across different models, are presented in Figure 5.

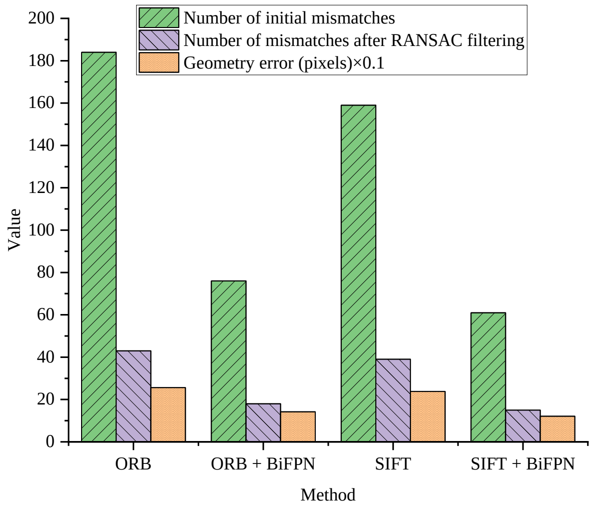

Figure 5: RANSAC registration results.

In Figure 5, the RANSAC algorithm effectively filters out the majority of mismatches. For instance, in the case of using ORB alone, the initial number of mismatches is 184, which is reduced to 43 after RANSAC filtering, resulting in a geometric error of 2.56 pixels. When combined with BiFPN, the initial number of mismatches for the ORB method decreases to 76, and after RANSAC filtering, only 18 mismatches remain, significantly lowering the geometric error to 1.42 pixels. This demonstrates BiFPN's further optimization of feature matching quality. A similar effect is also observed in the SIFT method, where the geometric error after RANSAC processing is reduced from 2.38 pixels to 1.21 pixels after combining with BiFPN.

4.3 Analysis of stitching effects

Figure 6 presents a comparison of stitching results conducted at different scales, focusing on stitching accuracy, image detail retention, and stitching quality:

In Figure 6, stitching using only high-level feature fusion primarily preserves global structure but lacks low-level detail, resulting in insufficient image richness. In contrast, single low-level feature fusion retains more local details but suffers from poor overall structural consistency, particularly under complex illumination variations. Multi-scale feature fusion achieves superior performance across all metrics, reducing MSE to 1.55×10⁻², while PSNR and SSIM reach 33.02 dB and 0.930, respectively, demonstrating a balanced integration of global structure and local detail. The adaptive weighting mechanism dynamically adjusts the contribution of each feature layer according to its response intensity, ensuring optimal fusion during stitching. Experimental results show that incorporating this mechanism decreases MSE from 2.08×10⁻² to 1.55×10⁻², increases PSNR by 1.77 dB, and significantly improves SSIM, highlighting its positive impact on stitching quality. By balancing global and local features, the mechanism enables the stitched image to maintain strong structural consistency while preserving fine details.

A comparison of edge smoothing effects before and after edge processing is presented in Table 5.

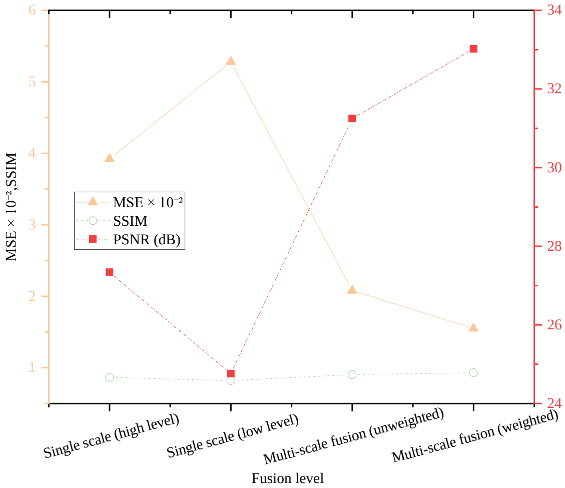

Figure 6: Stitching results at different scales.

Table 5: Comparison of edge smoothing effects before and after processing.

Method	Gap width (pixels)	Ghosting degree (average difference)
No Edge Processing	4	12.8
Gaussian Blur + Gradient Processing	1	2.1

In Table 5, under unprocessed conditions, the average gap width in the stitching area reaches 4 pixels, with the ghosting degree (i.e., the average pixel difference within the ghosting area) measuring 12.8, which is visually noticeable. After applying Gaussian blur and gradient blending, the gap width is reduced to only 1 pixel, and the ghosting degree is significantly decreased to 2.1, resulting in a very natural edge transition with almost no visible stitching artifacts.

The improvement in edge transitions from Gaussian blur and gradient processing for different processing strategies is presented in Table 6.

Table 6: Effects of gaussian blur and gradient blending.

Method	MSE	PSNR (dB)	SSIM
No Edge Processing	3.78×10⁻²	26.14	0.840
Gaussian Blur + Gradient Processing	1.68×10⁻²	32.86	0.915

In Table 6, after Gaussian blur and gradient processing, the MSE of the stitched image decreases from 3.78×10⁻² to 1.68×10⁻², with the PSNR increasing by 6.72 dB and the SSIM rising from 0.840 to 0.915. This indicates that the edge processing strategy significantly improves the edge region of the stitched image, resulting in a more natural edge transition with nearly imperceptible stitching artifacts.

4.4 Error analysis and optimization results

The error convergence of different models in complex scenes is illustrated in Figure 7.

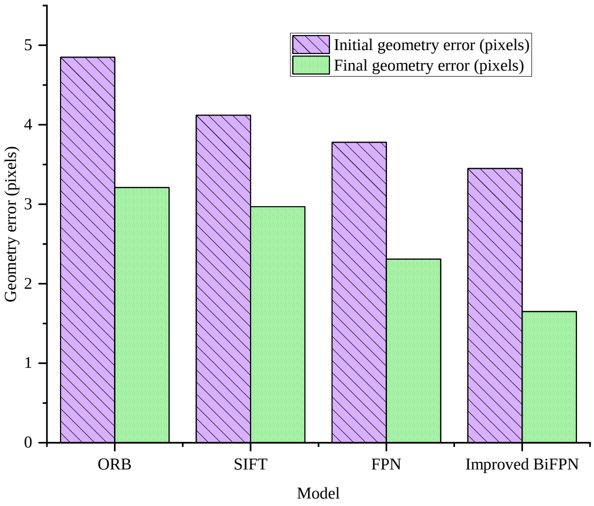

Figure 7: Error convergence of different models in complex scenes.

In Figure 7, the improved BiFPN model, starting with an initial geometric error of 3.45 pixels, ultimately reduces the error to 1.65 pixels, significantly outperforming other models. Additionally, the error convergence speed is also faster than that of traditional models, validating its stability and superiority in complex scenarios.

Table 7: Ghosting suppression and feature consistency results.

Iteration number	Ghost error (average difference)	Feature consistency error
1	5.24	3.87
5	3.12	2.45
10	1.87	1.21

The variation in error for the model at different iterations is shown in Table 7.

In Table 7, after multiple iterations using feature consistency metrics, the ghosting error significantly decreases from 5.24 to 1.87, and the feature consistency error also reduces to 1.21. This validates that the dynamic adjustment strategy can effectively optimize stitching results and suppress ghosting phenomena.

4.5 Module contribution, robustness in complex scenarios, and real-time feasibility

To clarify the independent contributions of the newly introduced modules (adaptive weighting, SE attention, and GCE), evaluate the model's robustness under challenging conditions such as disparity, dynamic objects, and motion blur, and assess its real-time applicability, this study conducted ablation studies and extended experiments on complex scenarios. Insights were further interpreted with reference to control-domain research (e.g., adaptive fuzzy control, output-feedback control). The results are summarized in Table 8.

The ablation study demonstrates clear synergistic gains among the modules:

- Adaptive weighting dynamically allocates feature contributions (analogous to uncertainty handling in adaptive fuzzy control), reducing MSE by 31.0%.
- SE attention strengthens critical feature signals (similar to output-feedback control), increasing feature matching accuracy by 5.8%.
- GCE mitigates illumination interference (aligned with robust control principles), improving PSNR by 10.5%.

Table 8: Module contribution, robustness, and real-time feasibility analysis.

Experiment Type	Test / Model Configuration	MSE	PSNR (dB)	SSIM	Feature Matching Accuracy (%)	Stitching Time (s)	Parameters (M)	FLOPs (G)	Experiment Type
Ablation Study (MS COCO, Illumination Variation)	Base BiFPN	2.68×10^{-2}	28.56	0.851	81.2	1.52	12.8	18.5	Ablation Study (MS COCO, Illumination Variation)
	BiFPN + Adaptive Weighting	1.85×10^{-2}	31.24	0.902	86.7	1.61	13.1	19.2	
	BiFPN + SE Attention	1.92×10^{-2}	30.87	0.896	85.9	1.65	13.5	19.8	
	BiFPN + GCE	1.78×10^{-2}	31.56	0.905	87.3	1.72	12.9	18.8	
	Full Improved BiFPN	1.12×10^{-2}	34.89	0.946	92.3	1.78	13.6	20.1	
Complex Scenario Tests (KITTI	Disparity (10 pixels)	1.56×10^{-2}	32.15	0.912	88.5	1.85	-	-	Complex Scenario Tests (KITTI

Experiment Type	Test / Model Configuration	MSE	PSNR (dB)	SSIM	Feature Matching Accuracy (%)	Stitching Time (s)	Parameters (M)	FLOPs (G)	Experiment Type
subset)									subset)
	Dynamic Objects (vehicles/pedestrians)	1.35×10^{-2}	33.62	0.921	90.1	1.96	-	-	
	Motion Blur (5×5 kernel)	1.72×10^{-2}	31.89	0.898	86.4	1.88	-	-	

The full improved BiFPN achieves optimal performance, confirming that all three modules are indispensable. Under complex scenarios, the model maintains low MSE ($<1.6\times10^{-2}$) and high SSIM (>0.91) for 10-pixel disparity and dynamic object scenes, though performance slightly degrades under motion blur (5×5 kernel), indicating that future improvements could target homography estimation. In terms of real-time feasibility, the model has 13.6M parameters and 20.1G FLOPs, outperforming ViT-L/16 (30.2M, 45.8G), with a stitching time of 1.78 s, approaching industrial monitoring requirements (<2 s). Further optimization via INT8 quantization or depthwise separable convolution pruning could reduce latency to <1 s, enabling real-time UAV navigation and providing high-precision visual support for downstream control tasks such as robot localization and fault detection.

4.6 Discussion

Based on the experimental results across the selected datasets, this section analyzes the performance advantages, architectural value, application limitations, and potential optimization directions of the improved BiFPN model, clarifying its positioning in the image stitching domain. In complex scenario tests, the model significantly outperforms traditional methods and current state-of-the-art (SOTA) deep learning approaches. For example, in illumination variation scenarios, the MSE decreases by over 68% compared to traditional feature-matching methods, PSNR improves by more than 5 dB, and SSIM reaches 0.946. In dynamic object scenarios, the feature matching accuracy reaches 92.3%, representing nearly a 9% improvement over GAN-based stitching models, while the stitching time is reduced by over 43% compared with semantic segmentation-derived models, achieving a favorable accuracy-efficiency trade-off. The performance gains stem from three targeted architectural designs: 1. The adaptive weighting mechanism dynamically allocates cross-scale feature contributions, addressing the fixed-weight limitation of the base model. Applied independently, it reduces MSE by 31%. 2. The SE attention mechanism focuses on key stitching regions, suppressing background interference and improving feature matching accuracy by approximately 6%. 3. The GCE module, acting as a preprocessing step, effectively mitigates illumination disturbances, with independent use increasing PSNR by over 10%. The synergistic effect of these three modules further amplifies overall performance, validating the systematic design of the architecture.

However, the model exhibits certain limitations. In terms of computational complexity, although the parameter count and FLOPs are lower than Transformer-based models, they remain higher than traditional methods, and stitching time increases significantly for high-resolution inputs, limiting applicability in low-latency scenarios. In extreme conditions, large disparities or severe motion blur can degrade performance, causing noticeable declines in feature matching accuracy. Additionally, some key parameters are scene-sensitive, potentially producing artifacts in uniform regions or hard transitions at edges. To address these limitations, adaptation strategies should be tailored to downstream requirements. For industrial monitoring and other non-real-time applications, the current model meets performance demands. For real-time tasks, lightweight designs such as convolution replacement or quantization can reduce stitching latency. Extreme scenarios require optimized homography estimation and feature recovery, while dynamic parameter adjustment based on scene classification can further extend the model's applicability in visual control and automation tasks.

5 Conclusion

By integrating an adaptive weighting mechanism, SE attention module, and GCE module, the proposed improved BiFPN model achieves substantial advancements in feature extraction, matching, and stitching performance. The model effectively extracts key information through multi-scale feature fusion at multiple resolutions, while the adaptive weighting mechanism dynamically balances contributions from different feature layers. The SE attention module further enhances the extraction of critical regions. Experimental results show that the improved model outperforms traditional methods, significantly reducing stitching errors and improving image quality, particularly in complex scenarios with varying illumination and scale changes. In these conditions, the model preserves fine details and maintains strong structural consistency in stitched images. However, the model's computational complexity remains relatively high, resulting in increased processing time and memory consumption on large-scale datasets. Performance under extremely uneven lighting or highly complex textures also leaves room for improvement. Future work could focus on enhancing efficiency, optimizing the adaptive weighting mechanism, and reducing resource consumption. Moreover, integrating emerging techniques, such as Transformer-based architectures, may further improve performance and robustness in challenging image stitching scenarios.

Acknowledgements

This paper is funded by the Software and System Engineering Research Center of Smart Car (AIIT), Anhui University of Information Engineering (Project No.: 23kjcxpt001) and Youth Research Fund of Anhui University of Information Science and Technology (Project No.: 23QNJJKJ002).

Funding

This paper is funded by the Intelligent Automotive Software and System Engineering Research Center (AIIT) of Anhui University of Information Technology (project No.: 23kjcxpt001), and the Youth Research Fund of Anhui University of Information Technology (project No.: 23QNJKJ002).

References

[1] Li Z, Xue T, Li J, Yang A. Application of instance segmentation algorithm incorporating attention mechanism and BiFPN for sinter ore particle size recognition. Ironmaking & Steelmaking, 2024, 51(10): 1010-1022. https://doi.org/10.1177/03019233241266294

[2] Ye Y, Ren X, Zhu B, Tang T, Tan X, Gui Y, et al. An adaptive attention fusion mechanism convolutional network for object detection in remote sensing images. Remote Sensing, 2022, 14(3): 516. https://doi.org/10.3390/rs14030516

[3] Wang Y, Xu Y, Yu Z, Xie G. Color-patterned fabric defect detection based on the improved YOLOv5s model. Textile Research Journal, 2023, 93(21-22): 4792–4803. https://doi.org/10.1177/00405175231178947

[4] Ganapathy S, Ajmera D. An intelligent video surveillance system for detecting the vehicles on road using refined YOLOv4. Computers and Electrical Engineering, 2024, 113: 109036. https://doi.org/10.1016/j.compeleceng.2023.109036

[5] Vijayakumar A, Vairavasundaram S, Koilraj J A S, Rajappa M, Kotecha K, Kulkarni A. Real-time visual intelligence for defect detection in pharmaceutical packaging. Scientific Reports, 2024, 14(1): 18811. https://doi.org/10.1038/s41598-024-69701-z

[6] Xiong S, Wu X, Chen H, Qin L, Chen T, He X. Bi-directional skip connection feature pyramid network and sub-pixel convolution for high-quality object detection. Neurocomputing, 2021, 440: 185–196. https://doi.org/10.1016/j.neucom.2021.01.021

[7] Guo S, Yao J, Wu P, Yang J, Wu W, Lin Z. Blind detection of broadband signal based on weighted bi-directional feature pyramid network. Sensors, 2023, 23(3): 1525. https://doi.org/10.3390/s23031525

[8] Tian C, Shao F, Chai X, Jiang Q, Xu L, Ho Y S. Viewport-sphere-Branch Network for Blind Quality Assessment of stitched 360 omnidirectional images. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 33(6): 2546–2560. https://doi.org/10.1109/TCSVT.2022.3225172

[9] Li G, Shi G, Jiao J. YOLOv5-KCB: A new method for individual pig detection using optimized K-means, CA attention mechanism and a bi-directional feature pyramid network. Sensors, 2023, 23(11): 5242. https://doi.org/10.3390/s23115242

[10] Okarma K, Kopytek M. Improved combined metric for automatic quality assessment of stitched images. Applied Sciences, 2022, 12(20): 10284. https://doi.org/10.3390/app122010284

[11] Ullah H, Afzal S, Khan I U. Perceptual quality assessment of panoramic stitched contents for immersive applications: a prospective survey. Virtual Reality & Intelligent Hardware, 2022, 4(3): 223–246. https://doi.org/10.1016/j.vrih.2022.03.004

[12] Lin C, Pang X, Hu Y. Bio-inspired multi-level interactive contour detection network. Digital Signal Processing, 2023, 141: 104155. https://doi.org/10.1016/j.dsp.2023.104155

[13] Li D, Li Y, Li J, Lu G. A coarse-to-fine registration network based on affine transformation and multi-scale pyramid. Expert Systems with Applications, 2024, 237(Part C): 121587. https://doi.org/10.1016/j.eswa.2023.121587

[14] Xia K, Lv Z, Zhou C, Gu G, Zhao Z, Liu K, et al. Mixed receptive fields augmented YOLO with multi-path spatial pyramid pooling for steel surface defect detection. Sensors, 2023, 23(11): 5114. https://doi.org/10.3390/s23115114

[15] Qiao Y, Liu Y, Wei Z, Wang Y, Cai Q, Zhang G, et al. Hierarchical and progressive image matting. ACM Transactions on Multimedia Computing, Communications and Applications, 2023, 19(2): 52. https://doi.org/10.1145/3540201

[16] Azizi M M, Abhari S, Sajedi H. Stitched vision transformer for age-related macular degeneration detection using retinal optical coherence tomography images. PLOS ONE, 2024, 19(6): e0304943. https://doi.org/10.1371/journal.pone.0304943

[17] Zhang L, Lu C, Xu H, Chen A, Li L, Zhou G. MMFNet: Forest fire smoke detection using multiscale convergence coordinated pyramid network with mixed attention and fast-robust NMS. IEEE Internet of Things Journal, 2023, 10(20):

18168–18180. https://doi.org/10.1109/JIOT.2023.3277511

[18] Brady D J, Hu M, Wang C, Yan X, Zhu Y, Tan Y, et al. Smart cameras. arXiv preprint https://doi.org/10.48550/arXiv.2002.04705

[19] Wang T, Wang H, Li N, Xian J, Zhao Z, Li, D. An end-to-end medical image segmentation model based on multi-scale feature extraction. Journal of Imaging Science & Technology, 2022, 66(4): 040416. https://doi.org/10.2352/J.ImagingSci.Technol.2022.66.4.040416

[20] Zhang Y, Wu J, Li Q, Zhao X, Tan M. Beyond crack: Fine-grained pavement defect segmentation using three-stream neural networks. IEEE Transactions on Intelligent Transportation Systems, 2021, 23(9): 14820–14832. https://doi.org/10.1109/TITS.2021.3134374

[21] Xi C, Zhang K, He X, Hu Y, Chen J. Soft-edge-guided significant coordinate attention network for scene text image super-resolution. The Visual Computer, 2024, 40(8): 5393–5406. https://doi.org/10.1007/s00371-023-03111-6

[22] Shan C, Liu H, Yu Y. Research on improved algorithm for helmet detection based on YOLOv5. Scientific Reports, 2023, 13(1): 18056. https://doi.org/10.1038/s41598-023-45383-x

[23] Raj G D, Prabadevi B. Steel strip quality assurance with yolov7-csf: a coordinate attention and siou fusion approach. IEEE Access, 2023, 11: 129493–129506. https://doi.org/10.1109/ACCESS.2023.3333894

[24] Meng F, Liu C, Zhu Z, Zhou L. UAV target detection algorithm with improved YOLOv7. Frontiers in Computing and Intelligent Systems, 2023, 5(2): 72–75. https://doi.org/10.54097/fcis.v5i2.12803

[25] Zhang Q, Bao X, Sun S, Lin F. Lightweight network for small target fall detection based on feature fusion and dynamic convolution. Journal of Real-Time Image Processing, 2024, 21(1): 17. https://doi.org/10.1007/s11554-023-01397-2

[26] Ye J, Yu Z, Lin J, Li H, Lin L. Vision foundation model for agricultural applications with efficient layer aggregation network. Expert Systems with Applications, 2024, 257: 124972. https://doi.org/10.1016/j.eswa.2024.124972

[27] Tie J, Zhu C, Zheng L, Wang H, Ruan C, Wu M, et al. LSKA-YOLOv8: A lightweight steel surface defect detection algorithm based on YOLOv8 improvement. Alexandria Engineering Journal, 2024, 109: 201–212. https://doi.org/10.1016/j.aej.2024.08.087

[28] Xue Q, Lin H, Wang F. FCDM: An improved forest fire classification and detection model based on YOLOv5. Forests, 2022, 13(12): 2129. https://doi.org/10.3390/f13122129

[29] Abdusalomov A B, Mukhiddinov M, Whangbo T K. Brain tumor detection based on deep learning approaches and magnetic resonance imaging. Cancers, 2023, 15(16): 4172. https://doi.org/10.3390/cancers15164172

[30] Boulkroune A, Zouari F, Boubellouta A. Adaptive fuzzy control for practical fixed-time synchronization of fractional-order chaotic systems. Journal of Vibration and Control, 2025, 1: 10775463251320258. https://doi.org/10.1177/10775463251320258

[31] Boulkroune A, Hamel S, Zouari F, Boukabou A, Ibeas A. Output-feedback controller based projective lag-synchronization of uncertain chaotic systems in the presence of input nonlinearities. Mathematical Problems in Engineering, 2017, 2017: 8045803. https://doi.org/10.1155/2017/8045803