MFIFB:Multi-Focus Image Fusion Benchmark

Biying Jian[a], Liuxuan Chen[a] and Jianbo Sheng[a]

[a]Jinan University, Guangzhou 510632, China

ARTICLE INFO

Keywords:
multi-focus image fusion
benchmark
image processing

ABSTRACT

Multi-focus image fusion is an important area in image processing. Although the development of image fusion algorithms has progressed considerably in recent years, there is still a lack of codebases and benchmarks to measure the state-of-the-art. In this paper, we present the Multifocus Image Fusion Benchmark (MFIFB) after briefly reviewing the recent advances in multifocus image fusion. 21 pairs of images, a codebase containing 3 fusion algorithms, and 6 evaluation metrics. We also conducted extensive experiments in this benchmark to understand the performance of these algorithms. By analyzing the qualitative and quantitative results, we identify effective algorithms for robust image fusion and provide some insights into the current state of the art in image fusion, and make some observations about the current state and future prospects of the field. The current state of the field and future prospects.

1. Introduction

The goal of image fusion is to combine information from different images to generate a more information-rich image for subsequent processing. Numerous image fusion algorithms have been proposed, which can generally be categorized into pixel-level, feature-level, and decision-level methods based on the fusion level. Additionally, image fusion can be performed in either the spatial domain or the transform domain. Depending on the application area, image fusion techniques can be divided into several types, such as medical image fusion, multi-focus image fusion, remote sensing image fusion, multi-exposure image fusion, and visible-infrared image fusion. Among these, multi-focus image fusion is one of the most frequently used types. This is because multi-focus image fusion can be applied in various scenarios, such as target tracking, target detection, and biometric recognition. Figure 1 illustrates an example of multi-focus image fusion. Our code can be find in https://github.com/zenobier/MFIFB-Multi-Focus-Image-Fusion-Benchmark/tree/main.

Figure 1: an example of multi-focus image fusion.

The current research on multi-focus image fusion faces several challenges that significantly hinder the development of the field. Firstly, there is no recognized multi-focus image fusion dataset available for comparing performance under the same standards. Consequently, different images are often used in experiments in the literature, making it difficult to compare the performance of various algorithms. Secondly, it is crucial to evaluate the performance of advanced fusion algorithms to highlight their advantages and disadvantages and help determine future research directions in the field. Thirdly, although the source code for some image fusion algorithms, such as ZMFF and LBP [15, 52], has been made public, most algorithms have different interfaces and usage methods, making large-scale performance evaluations inconvenient and time-consuming for researchers.

To address these issues, we have established a multi-focus image fusion benchmark (MFIFB) in this work, which includes 40 pairs of multi-focus images, 3 publicly available fusion algorithms, and 6 evaluation metrics to facilitate performance evaluation tasks.

The main contributions of this paper are reflected in the following aspects:

1. **Dataset**
   We created a test set containing several pairs of multi-focus images. These image pairs were collected from the internet and various tracking datasets, covering a wide range of environments and working conditions such as indoor, outdoor, low-light, and overexposure. Therefore, this dataset can test the generalization capability of image fusion algorithms.
2. **Code Repository**
   We collected three image fusion algorithms developed in the past five years and integrated them into a code repository. This makes it easy to run the algorithms and compare their performance.
3. **Comprehensive Performance Evaluation**
   We implemented six evaluation metrics in MFIFB to comprehensively compare fusion performance. We ran the three collected algorithms on the proposed dataset and conducted a thorough comparison. All results are publicly available for interested readers.

bxj299@student.bham.ac.uk (B. Jian); chenliuxuan2005@163.com (L. Chen); jianbosheng@hotmail.com (J. Sheng)
ORCID(s):

Algorithm	Total Runtime (s)	Average Runtime(s)
ADF [38]	0.082	0.0039
CBF [56]	0.078	0.0037
CNN [45]	0.076	0.0036
DLF [54]	0.082	0.0039
FPDE [4, 26]	0.080	0.0038
GFCE [8]	0.082	0.0039
GFF [61]	0.078	0.0037
GTF [37]	0.085	0.0040
HMSD_GF	0.081	0.0039
Hybrid_MSD [66]	0.082	0.0039
IFEVIP [46]	0.081	0.0039
LatLRR [27]	0.090	0.0043
LP_SR [34]	0.079	0.0038
MGFF	0.088	0.0042
MSVD [22]	0.085	0.0040
NSCT_SR [2]	0.078	0.0037
ResNet [23]	0.078	0.0037
RP_SR	0.078	0.0037
TIF [63]	0.081	0.0039
VSMWLS	0.080	0.0038
IFCNN [60]	0.083	0.0040
SeAFusion [40]	0.077	0.0037
SwinFusion [32]	0.080	0.0038
U2Fusion [49]	0.084	0.0040
YDTR [44]	0.083	0.0040

Table 1
Runtime Performance of Image Fusion Algorithms

2. Related Work

2.1. Image Fusion

Many multi-focus image fusion methods have been proposed. Before the introduction of deep learning into the field of image fusion, the main methods were typically categorized into several types: multi-scale transformation, sparse representation, subspace and saliency methods, hybrid models, and other approaches, classified based on their respective theories.

In recent years, deep learning-based image fusion methods have continuously emerged. Deep learning can help address several key issues in image fusion. Methods such as Convolutional Neural Networks (CNNs) [45], Generative Adversarial Networks (GANs) [65], Siamese Networks [10, 55], and autoencoders have been used for image fusion. Beyond the fusion methods themselves, image quality assessment in image fusion performance evaluation has also benefited from deep learning. These technological advancements are steering image fusion techniques towards machine learning, and more research outcomes are expected in the coming years.

The Table 1 presents the runtime performance of various image fusion algorithms. It includes both the total runtime and the average runtime (in seconds) for each algorithm. The data provides a comprehensive overview of the efficiency of each method in terms of computational speed, aiding in the selection of suitable algorithms for time-sensitive applications.

2.2. Multi-Focus Image Fusion

Multi-focus image fusion is an essential process in computational imaging, designed to synthesize a single image from multiple inputs captured with varying focus levels [58, 19]. This technique is pivotal in scenarios where a single shot cannot capture all relevant features in focus due to depth of field limitations. Below, we delve into various sophisticated methodologies and their applications.

- **Spatial Domain Methods:** These techniques manipulate pixel intensities directly. The simplest approach is pixel averaging, which often leads to blurring. More advanced methods, such as *focus measure operators* like the Laplacian and Sobel filters, identify focused regions by detecting high-frequency components. However, these methods can struggle with artifacts at focus boundaries.

- **Transform Domain Methods:** These methods transform images into frequency or multi-resolution domains. *Wavelet transform* is prevalent due to its multi-scale capability, allowing for effective edge and texture representation. *Curvelet and contourlet transforms* provide enhanced directionality and anisotropy, beneficial for capturing geometric features. For instance, *nonsubsampled contourlet transform (NSCT)* is noted for its shift-invariance and improved performance in preserving directional details.

- **Deep Learning Approaches:** With the advent of deep learning, convolutional neural networks (CNNs) have revolutionized image fusion. Techniques such as *DenseFuse* and *FusionGAN* leverage deep architectures to learn complex fusion rules, achieving superior results in terms of detail preservation and artifact reduction. These models are trained end-to-end, often requiring large datasets for optimal performance. Table 2 summarizes various deep learning-based methods, highlighting their frameworks and supervisory paradigms. Notably, methods like *MFF-GAN* utilize unsupervised learning, offering flexibility in scenarios with limited labeled data.

- **Evaluation Metrics:** Fusion quality is assessed using metrics like *structural similarity index (SSIM)*, *peak signal-to-noise ratio (PSNR)*, and *mutual information (MI)*. Advanced metrics such as *edge preservation index (EPI)* and *visual information fidelity (VIF)* provide deeper insights into the perceptual and informational quality of the fused images.

- **Applications:** Multi-focus image fusion is crucial in medical imaging for enhancing diagnostic accuracy, as in the fusion of CT and MRI images. In remote sensing, it aids in the integration of multi-spectral and panchromatic images to enhance spatial resolution. In consumer electronics, it is employed in smartphone cameras to produce images with extended depth of field.

Methods	Title	Foundational Framework	Supervisory Paradigm
CNN [28]	Multi-focus image fusion with a deep convolutional neural network	CNN	supervised
ECNN [3]	Ensemble of CNN for multi-focus image fusion	CNN	supervised
MLFCNN [53]	Multilevel features convolutional neural network for multi-focus image fusion	CNN	supervised
DRPL [24]	DRPL: Deep Regression Pair Learning for Multi-Focus Image Fusion	CNN	supervised
MMF-Net [30]	An α-Matte Boundary Defocus Model-Based Cascaded Network for Multi-Focus Image Fusion	CNN	supervised
MFFSSIM [50]	Towards Reducing Severe Defocus Spread Effects for Multi-Focus Image Fusion via an Optimization Based Strategy	CNN	Without supervision
MFNet [51]	Structural Similarity Loss for Learning to Fuse Multi-Focus Images	CNN	supervised
GEU-Net [48]	Global-Feature Encoding U-Net (GEU-Net) for Multi-Focus Image Fusion	CNN	Self-supervised
DTMNet [47]	DTMNet: A Discrete Tchebichef Moments-Based Deep Neural Network for Multi-Focus Image Fusion	CNN	Without supervision
MFF-GAN [57]	MFF-GAN: An unsupervised generative adversarial network with adaptive and gradient joint constraints for multi-focus image fusion	GAN	Without supervision
SMFuse [31]	SMFuse: Multi-Focus Image Fusion Via Self-Supervised Mask-Optimization	CNN	Without supervision
ACGAN [16]	A generative adversarial network with adaptive constraints for multi-focus image fusion	GAN	supervised
FuseGAN [12]	Learning to fuse multi-focus image via conditional generative adversarial network	GAN	supervised
D2FMIF [62]	Depth-Distilled Multi-focus Image Fusion	CNN	supervised

Table 2
Comparison of classical common image fusion methods, Frameworks, and Paradigms for Multi-Focus Image Fusion.

- **Challenges and Future Directions:** Current challenges include handling dynamic scenes with motion artifacts and developing real-time fusion systems. Future research is likely to focus on integrating fusion techniques with emerging technologies like computational photography and augmented reality, as well as exploring unsupervised learning frameworks for fusion.

2.3. Image Fusion Benchmark

The evaluation of image fusion techniques is critical for determining their effectiveness and applicability in real-world scenarios. Benchmarks provide standardized datasets and metrics to facilitate objective comparison among different methods.

- **Datasets:** Commonly used datasets for image fusion benchmarking include the *Lytro Multi-focus Dataset*, which offers images with varying focus levels, and the *TNO Image Fusion Dataset*, designed for testing multi-modal image fusion methods. Additionally, the *Visible and Infrared Image Fusion Benchmark (VIFB)* [59] consists of 40 image pairs, a code library of 20 fusion algorithms, and 13 evaluation metrics, providing a comprehensive resource for evaluating fusion techniques.
- **Evaluation Metrics:**
  - *Structural Similarity Index (SSIM) [5]:* Measures the perceived quality of the fused image by comparing it to a reference image.
  - *Peak Signal-to-Noise Ratio (PSNR) [36]:* Evaluates the fidelity of the fusion process by quantifying the error between the fused and reference images.
  - *Mutual Information (MI) [21]:* Assesses the amount of information preserved in the fused image from the source images.
  - *Visual Information Fidelity (VIF) [13]:* Provides an advanced metric that evaluates the perceptual quality of the fused image based on human visual system models.
- **Challenges in Benchmarking:**
  - *Lack of Ground Truth:* Many real-world fusion tasks lack a definitive ground truth image, complicating objective assessment.
  - *Diverse Application Needs:* Different applications may prioritize different aspects of the fused image (e.g., detail enhancement vs. noise reduction), making a single metric insufficient.
- **Future Directions:**

– *Development of Comprehensive Benchmarks:* Efforts are underway to create more comprehensive benchmarks that include a wider variety of image types and fusion scenarios.
– *Integration of Subjective Evaluation:* Incorporating human judgment into benchmarking processes to better capture perceptual quality.

3. MFF Overall Concept

Multi-focus image fusion (MFF) is a pivotal technique in image processing, designed to merge multiple images with varying focus depths into a single, coherent image. This process enhances the depth of field by integrating sharp regions from each input image, resulting in a composite image with superior clarity and detail. MFF is crucial in domains such as medical imaging, surveillance, and photography, where capturing all pertinent details in one shot is challenging. The technique employs sophisticated algorithms to evaluate and select the most in-focus regions from each image, optimizing overall visual information. By utilizing advanced computational methods, MFF improves both the qualitative and quantitative aspects of image analysis, providing a robust foundation for subsequent processing tasks. The basic workflow of MFF is illustrated in Figure 2.

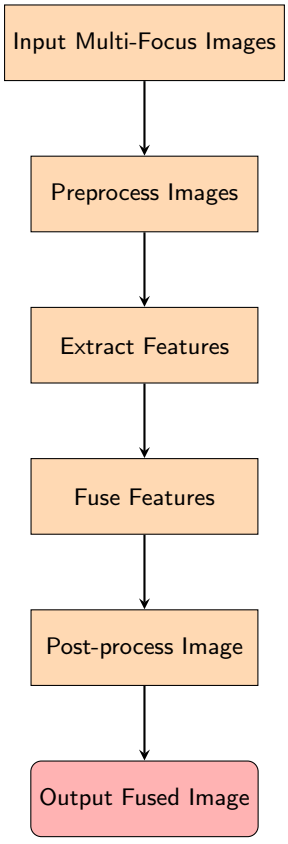


Figure 2: Flowchart of Multi-Focus Image Fusion Process

4. Multi-Focus Image Fusion Benchmark

4.1. Dataset

The dataset in MFIFB is a test set comprising 40 pairs of multi-focus fused images. These images were collected by the authors from the internet and the FusionTrack dataset, as well as transformed using the Gaussian blur method. They cover a wide range of environments and working conditions, such as indoor, outdoor, low light, and overexposure. Each pair of multi-focus fused images is registered to ensure successful image fusion. Some examples from the dataset are shown in Figure 3.

Figure 3: The multi-focus image fusion test set in MFIFB dataset.

4.1.1. Image Collection

The initial step in constructing our dataset involved gathering a diverse set of images from the internet. To ensure diversity, images were carefully selected to represent a wide range of environments and conditions, thereby providing a robust foundation for multi-focus image fusion research.

4.1.2. Applying Gaussian Blur

To simulate different focus areas within each image, we applied Gaussian blur to specific regions. For each image, a half region was selected to remain sharp, while Gaussian blur was applied to the remaining area. This process was repeated for a second image, focusing on a different region, thus creating a pair of images with complementary focus areas.

4.1.3. Gaussian Blur Technique

The Gaussian blur [11] is an image blur filter that calculates the transform for each pixel using a normal distribution,

defined by the equation:

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (1)$$

- **Kernel Size:** Determines the extent of blurring applied to the image.
- **Sigma/StDev:** Controls the spread of the blur; higher values result in more significant blurring.

This systematic approach to dataset construction ensures that the resulting images are well-suited for evaluating and developing multi-focus image fusion algorithms.

4.2. Baseline algorithms

In our study, we carefully selected six multi-focus image fusion methods, taking into consideration factors such as time efficiency, technical complexity, and ease of implementation in MATLAB. These methods are summarized in Table 4. The table provides a detailed comparison of each method's type, advantages, and disadvantages, offering a comprehensive insight into their applicability and performance in various scenarios.

Table 3 provides a comparative analysis of various image fusion methods based on key criteria such as computational complexity, robustness, implementation difficulty, and their respective application scenarios. The Curvelet method, known for its high computational complexity and robustness, is well-suited for medical image processing. In contrast, the Gradient method, with low computational demands and robustness, is preferred for processing natural scene images. The IDWT and Laplacian Pyramid methods offer moderate levels across all criteria, making them versatile for both natural, medical, and multi-scale image processing. The Max method, characterized by low complexity and robustness, is ideal for real-time processing. Lastly, the NSCT method, with its high complexity and robustness, is best applied to complex structured images.

4.2.1. Curvelet

The Curvelet transform is a powerful tool for image fusion, particularly effective in capturing edges and other singularities along curves. It is an extension of the wavelet transform, designed to handle the anisotropic features of images more efficiently. The mathematical formulation of the Curvelet transform involves decomposing an image into different scales and orientations, capturing the intrinsic geometrical structure.

$$C = \text{fdct\_wrapping}(I, \text{is\_real}, \text{finest}, \text{nbscales}, \text{nbangles\_coarse}) \quad (2)$$

In the above equation, I represents the input image, and the parameters control various aspects of the transform. The **fdct_wrapping** function performs the Fast Discrete Curvelet Transform, where:

- **is_real**: Specifies whether the input is a real image.
- **finest**: The number of scales at the finest level.
- **nbscales**: The total number of scales.
- **nbangles_coarse**: The number of angles at the coarsest level.

The fusion process involves several key steps:

1. **Compute Curvelet Transforms**: For each input image, compute the Curvelet coefficients using the **fdct_wrapping** function.
2. **Coefficient Fusion**: For each scale and orientation, apply a maximum selection rule. This is done by comparing the absolute values of the coefficients from the two images and selecting the larger one:

$$\text{fusedC}_{s,w} = \begin{cases} C1_{s,w}, & \text{if } |C1_{s,w}| > |C2_{s,w}| \\ C2_{s,w}, & \text{otherwise} \end{cases} \quad (3)$$

   Here, $C1$ and $C2$ are the Curvelet coefficients of the two images, and fusedC are the fused coefficients.
3. **Inverse Curvelet Transform**: Reconstruct the fused image from the fused Curvelet coefficients using the inverse transform function **ifdct_wrapping**.

The provided MATLAB implementation efficiently executes these steps. The function **curveletFusion** accepts two images, computes their Curvelet transforms, applies the fusion rule, and reconstructs the final fused image. The non-algorithmic parts, such as setting paths and reading images, facilitate the processing of multiple image pairs.

4.2.2. Gradient

The gradient-based fusion algorithm is a technique that utilizes the gradient magnitude to merge two images. The primary goal is to retain the most detailed regions from each image by evaluating the gradient magnitude. The mathematical formulation and implementation details are outlined below.

$$\text{Gradient Magnitude} = \sqrt{G_x^2 + G_y^2} \quad (4)$$

where G_x and G_y are the gradients in the x and y directions, respectively.

1. **Compute Image Gradients:** For each input image, calculate the gradients G_x and G_y using a function such as `imgradientxy`.
2. **Calculate Gradient Magnitude:** Determine the gradient magnitude for both images using the formula above.
3. **Pixel Selection:** Create a mask by comparing the gradient magnitudes of the two images. Select pixels from the image with the higher gradient magnitude, as it indicates a more detailed region.
4. **Fuse Images:** Use the mask to combine the two images, selecting pixels from the image with the higher gradient magnitude.

Method	Computational Complexity	Robustness	Implementation Difficulty	Application Scenarios
Curvelet [7]	High	Moderate	High	Medical image processing
Gradient [9]	Low	Low	Moderate	Natural scene images
IDWT(Inverse Discrete Wavelet Transform) [35]	Moderate	Moderate	Moderate	Natural and medical images
Laplacian Pyramid [6]	Moderate	Moderate	Moderate	Multi-scale image processing
Max	Low	Low	Low	Real-time processing
NSCT(Non-Subsampled Contourlet Transform) [64]	High	High	High	Complex structured images

Table 3
Evaluation of Image Processing Methods: Computational Complexity, Robustness, Implementation Difficulty, and Application Scenarios.

Method	Type	Advantages	Disadvantages
Curvelet	Transform Domain	Excels at handling curves and edges	High computational complexity
Gradient	Spatial Domain	Emphasizes edges and details	Sensitive to noise
IDWT	Transform Domain	Preserves multi-resolution details	May cause artifacts
Laplacian Pyramid	Multi-scale Decomposition	Handles features at various scales	Possible detail loss
Max	Simple Pixel-level	Simple and efficient	Possible detail and contrast loss
NSCT	Transform Domain	Good directional selectivity and multi-scale analysis	High computational complexity

Table 4
Comparison of Image Processing Methods: Advantages and Disadvantages.

4.2.3. IDWT

The Inverse Discrete Wavelet Transform (IDWT) is a fundamental technique in image processing, particularly in the context of image fusion. It allows for the reconstruction of an image from its wavelet coefficients, effectively reversing the Discrete Wavelet Transform (DWT). The process involves several key steps:

$$\text{IDWT:} \quad f(x,y) = \sum_{i,j} c_{i,j}\psi_{i,j}(x,y) \tag{5}$$

In this context, $c_{i,j}$ are the wavelet coefficients, and $\psi_{i,j}(x,y)$ are the wavelet basis functions. The IDWT reconstructs the image by summing the contributions from all wavelet coefficients and their corresponding basis functions.

The implementation of IDWT in the image fusion process involves the following steps:

1. **Wavelet Decomposition:** Each image channel is decomposed into approximation and detail coefficients using the DWT. This includes horizontal, vertical, and diagonal details.
2. **Coefficient Fusion:** For each sub-band, the fusion rule selects the coefficients with the maximum absolute value from the corresponding sub-bands of the input images. This ensures that the most significant features are preserved.
3. **Image Reconstruction:** The fused coefficients are used to reconstruct the image using the IDWT. This step combines the selected coefficients to form the final fused image.
4. **Channel Processing:** The process is repeated for each color channel to ensure a complete and accurate fusion of the entire image.

This method leverages the strengths of the wavelet transform, capturing both spatial and frequency information, which is crucial for effective image fusion.

4.2.4. Laplacian Pyramid

The Laplacian Pyramid is a multi-scale image representation technique that decomposes an image into a set of band-pass filtered images. This method is particularly useful for image fusion, allowing the combination of multiple images by preserving important features at various scales.

$$L_i = G_i - \text{expand}(G_{i+1}) \tag{6}$$

where L_i is the Laplacian level at scale i, G_i is the Gaussian pyramid level at scale i, and $\text{expand}(G_{i+1})$ is the expanded version of the next Gaussian level. The final level of the Laplacian pyramid is simply the last level of the Gaussian pyramid.

The implementation involves the following steps:

1. **Gaussian Pyramid Construction:** Start with the original image and iteratively apply a Gaussian filter followed by downsampling to create multiple levels.

2. **Laplacian Pyramid Construction:** For each level, subtract the expanded version of the next Gaussian level from the current Gaussian level to obtain the Laplacian level.
3. **Fusion Rule:** For each level of the pyramids of two images, select the coefficients with the maximum absolute value to form the fused pyramid.
4. **Reconstruction:** Reconstruct the fused image by expanding and summing the levels of the fused pyramid.

In the code implementation, each color channel of the input images is processed independently. The function `laplacianPyramid` constructs the pyramid for an image, while `laplacianPyramidFusion` performs the fusion by comparing the absolute values of the Laplacian coefficients from two images at each level. The fusion process ensures that the most prominent features from each input image are preserved in the output.

4.2.5. Max

The Max algorithm is a straightforward method used in image fusion. It operates by selecting the maximum intensity value for each pixel position across the input images. This approach ensures that the most prominent features from each image are preserved in the fused result.

$$F(i,j,c) = \max(I_A(i,j,c), I_B(i,j,c)) \tag{7}$$

where $F(i,j,c)$ is the fused image, $I_A(i,j,c)$ and $I_B(i,j,c)$ are the input images, i and j denote the pixel position, and c represents the color channel.

Implementation:

The implementation involves iterating over each pair of input images and applying the max operation across all color channels. For each pixel, the algorithm compares the intensity values of the two images and selects the higher value for the fused image. This process is repeated for all pixels and channels, resulting in a final image that combines the most intense features from both inputs.

4.2.6. NSCT

The Non-Subsampled Contourlet Transform (NSCT) is a powerful multi-scale and multi-directional image decomposition technique. It is particularly effective for image fusion tasks due to its ability to capture geometrical structures in images. The NSCT consists of a shift-invariant filter bank that decomposes an image into low-frequency and high-frequency components. The process is defined by the following key steps:

1. **Decomposition:** The input images are decomposed using NSCT into a set of coefficients. This involves applying a pyramidal filter and a directional filter to obtain multi-scale and multi-directional representations. The decomposition levels are specified by the vector $nlevels = [0, 1, 1, 1]$.

$$\text{coeffs} = \text{nsctdec}(\text{image}, nlevels, \text{dfilt}, \text{pfilt}) \tag{8}$$

where dfilt is the directional filter ('dmaxflat7') and pfilt is the pyramidal filter ('maxflat').

2. **Fusion of Low-Frequency Subbands:** The low-frequency components of the two images are averaged to retain the approximate information from both images.

$$\text{fusedCoeffs}_1 = \frac{\text{coeffs1}_1 + \text{coeffs2}_1}{2} \tag{9}$$

3. **Fusion of High-Frequency Subbands:** For the high-frequency components, a selection rule based on the maximum absolute value is applied. This ensures that the most significant features from both images are preserved.

$$\text{fusedCoeffs}_i^j = \begin{cases} \text{coeffs1}_i^j & \text{if } |\text{coeffs1}_i^j| > |\text{coeffs2}_i^j| \\ \text{coeffs2}_i^j & \text{otherwise} \end{cases} \tag{10}$$

4. **Reconstruction:** The fused coefficients are then used to reconstruct the final fused image using the inverse NSCT.

$$\text{fusedImage} = \text{nsctrec}(\text{fusedCoeffs}, \text{dfilt}, \text{pfilt}) \tag{11}$$

In the implementation, the NSCT toolbox is utilized to perform the decomposition and reconstruction. The process iterates over each color channel of the input images, applying the NSCT-based fusion method to produce a final, coherent image.

4.3. Evaluation metrics

The evaluation of multi-focus image fusion has introduced numerous metrics[43]. As described in the literature [41, 42, 33], these image fusion metrics can be categorized into four types: information theory-based, image feature-based, image structural similarity-based, and human perception-based metrics. However, no single metric is superior to all others. To enable a comprehensive and objective performance comparison, we have implemented 10 evaluation metrics in MFIFB. All the metrics implemented in MFIFB and their corresponding categories are listed in Table 5. As shown in the table, the metrics implemented in MFIFB cover all four categories. Calculating these metrics for each method is convenient, allowing for easy comparison of different methods' performance. More information on the evaluation metrics can be found in the literature [39].

4.3.1. Average Gradient

The Average Gradient (AG) is a metric used to assess the sharpness and clarity of an image. It quantifies the amount of detail by calculating the gradients of the image intensity values. A higher average gradient indicates a sharper image with more defined edges.

Metric	Description	Category
AG [25]	Measures the sharpness of an image by calculating the average gradient.	Sharpness
CC [17, 20]	Evaluates the degree of correlation between two images.	Similarity
Hab [18]	Assesses the suitability of an image for certain applications.	Application Suitability
MSE	Quantifies the difference between values implied by an estimator and the true values.	Error Measurement
PSNR [1]	Measures the peak error between two images, indicating quality.	Quality Assessment
QCB [29]	Evaluates the quality components of an image.	Quality Assessment
QG	Measures the quality of an image based on gradient analysis.	Quality Assessment
QP [14]	Assesses perceived quality by human observers.	Perception
SD	Indicates the amount of variation or dispersion in an image.	Statistical Measurement
VIF [13]	Quantifies the fidelity of visual information in the pixel domain.	Fidelity

Table 5
We selected 10 evaluation criteria for assessing the quality of fused images, along with their descriptions and categories.

$$AG = \frac{1}{(r-1)(c-1)} \sum_{k=1}^{b} \sum_{i=1}^{r-1} \sum_{j=1}^{c-1} \sqrt{\frac{\left(\frac{\partial I}{\partial x}\right)^2 + \left(\frac{\partial I}{\partial y}\right)^2}{2}} \quad (12)$$

Where r, c, and b represent the dimensions of the image, specifically the number of rows, columns, and bands (or channels), respectively. The partial derivatives $\frac{\partial I}{\partial x}$ and $\frac{\partial I}{\partial y}$ are computed using the gradient function, which approximates the rate of change in pixel intensity in both the horizontal and vertical directions.

1. **Convert the image to double precision:** This ensures that calculations are performed with sufficient accuracy.
2. **Compute gradients for each band:** For each color band, calculate the horizontal and vertical gradients using finite differences.
3. **Calculate gradient magnitude:** For each pixel, compute the gradient magnitude as the square root of the average of squared horizontal and vertical gradients.
4. **Average over the entire image:** Sum the gradient magnitudes and normalize by the image size to obtain the average gradient for each band.
5. **Final output:** The overall average gradient is the mean of the values obtained for all bands.

4.3.2. Correlation Coefficient

The Correlation Coefficient (CC) is a statistical measure used to evaluate the similarity between two images. It quantifies how well the fused image retains the information from the source images. The CC is calculated using the following formula:

$$r_{XY} = \frac{\sum_{i=1}^{N} \sum_{j=1}^{M} (X_{ij} - \overline{X})(Y_{ij} - \overline{Y})}{\sqrt{\sum_{i=1}^{N} \sum_{j=1}^{M} (X_{ij} - \overline{X})^2 \sum_{i=1}^{N} \sum_{j=1}^{M} (Y_{ij} - \overline{Y})^2}} \quad (13)$$

where X and Y represent the image matrices, $\overline{X}$ and $\overline{Y}$ are the mean values of the respective images, and N and M are the dimensions of the images.

In the implementation, the correlation coefficient is computed for both pairs of source and fused images:

1. Calculate the correlation between the first source image A and the fused image F.
2. Calculate the correlation between the second source image B and the fused image F.
3. The final correlation coefficient CC is the average of these two correlation values.

This process ensures that the fused image F maintains a high degree of similarity with both source images A and B, indicating effective fusion.

4.3.3. Hab Metric Explanation

The Hab metric is used to compute the mutual information between two images, which is a measure of the amount of information shared between them. This metric is particularly useful in evaluating the quality of image fusion, as it quantifies how much information from the original images is retained in the fused image.

$$\mathrm{HabR} = -\sum_{i=1}^{N} \sum_{j=1}^{M} p(i,j) \log_2(p(i,j)) \quad (14)$$

where $p(i,j)$ is the joint probability distribution of the grey values of the two images, calculated from their joint histogram.

1. **Initialization:** The function initializes a counter matrix to store the joint histogram of the two images.
2. **Histogram Calculation:** Each pixel in the images is iterated over, and the counter matrix is updated to reflect the frequency of each pair of grey values.
3. **Probability Distribution:** The joint probability distribution is computed by normalizing the counter matrix with the total number of pixels.
4. **Entropy Calculation:** The mutual information, HabR, is calculated by summing over the joint probability distribution, weighted by the logarithm of the distribution itself.

This process provides a quantitative measure of the similarity between the two images in terms of the information they share, which is crucial for assessing the effectiveness of image fusion techniques.

4.3.4. Mean Squared Error

The Mean Squared Error (MSE) is a widely used metric for evaluating the quality of image fusion. It measures the average of the squares of the errors between the fused image and the source images. The MSE is defined mathematically as follows:

$$\text{MSE} = \frac{1}{m \times n} \sum_{i=1}^{m} \sum_{j=1}^{n} (F(i,j) - A(i,j))^2 + (F(i,j) - B(i,j))^2 \tag{15}$$

where F is the fused image, A and B are the source images, and m and n are the dimensions of the images.

To compute the MSE, the following steps are performed:

1. Normalize the pixel values of images A, B, and F by dividing by 255.0 to ensure they are in the range [0, 1].
2. Calculate the MSE between the fused image F and each source image A and B.
3. Compute the average of these two MSE values to obtain the final MSE.

This metric provides a quantitative measure of the fidelity of the fused image with respect to the original source images.

4.3.5. Peak Signal to Noise Ratio

The Peak Signal-to-Noise Ratio (PSNR) is a widely used metric for evaluating the quality of image reconstruction. It measures the ratio between the maximum possible power of a signal and the power of corrupting noise that affects the fidelity of its representation. The PSNR is expressed in decibels (dB) and is calculated using the following formula:

$$\text{PSNR} = 20 \cdot \log_{10} \left(\frac{255}{\sqrt{\text{MSE}}} \right) \tag{16}$$

where **MSE** (Mean Squared Error) is computed as the average of the squared differences between the pixel values of the original and the fused images. In the given implementation, the MSE is calculated by averaging the MSE between the fused image F and each of the input images A and B:

$$\text{MSE} = 0.5 \cdot \text{MSE}_{AF} + 0.5 \cdot \text{MSE}_{BF} \tag{17}$$

1. First, the images A, B, and F are normalized by dividing by 255.0 to scale the pixel values between 0 and 1.
2. The dimensions of the fused image F are obtained to calculate the MSE.
3. **MSE_AF** is computed as the sum of squared differences between F and A, divided by the number of pixels.
4. **MSE_BF** is similarly computed between F and B.
5. The final **MSE** is the average of **MSE_AF** and **MSE_BF**.
6. Finally, the PSNR is calculated using the formula, providing a measure of the quality of the fused image compared to the original images.

4.3.6. Quality Composite Evaluation

The Quality Composite Evaluation (QCB) is a metric designed to assess the quality of a fused image by comparing it to its source images. This metric is particularly useful in evaluating multi-focus image fusion results. The QCB is computed by analyzing the local contrast of the images and measuring the structural similarity between the fused image and each source image.

$$QCB = \frac{1}{N} \sum_{i=1}^{N} \text{SSIM}(C_f, C_{s_i}) \tag{18}$$

where N is the number of source images, C_f is the local contrast of the fused image, and C_{s_i} is the local contrast of the i-th source image. The function SSIM denotes the Structural Similarity Index, which is used to measure the similarity between two images.

1. **Local Contrast Computation:** The local contrast of both the fused image and the source images is computed using a Gaussian-weighted neighborhood. This involves calculating the local mean and variance, and then deriving the standard deviation as the measure of contrast.
2. **Normalization:** Before computing the SSIM, the local contrast maps are normalized to ensure they are within the range [0, 1].
3. **Similarity Calculation:** The SSIM is calculated between the normalized contrast map of the fused image and each source image. The average of these values provides the QCB score, indicating the overall quality of the fusion.

This method provides a comprehensive evaluation by focusing on local contrast and structural similarity, offering a robust assessment of the fusion quality.

4.3.7. Gradient Quality Evaluation

The Gradient Quality Evaluation (QG) is a crucial metric for assessing the quality of fused images by analyzing the preservation of edge information. The process involves several key steps:

1. **Compute Gradient Maps:** For each image (input images and the fused image), compute the gradient maps using filters. The gradients in the x and y directions are calculated, followed by the magnitude and angle of the gradients. To avoid division by zero, a small constant is added where necessary.

2. **Edge Preservation Estimation:** The edge preservation is estimated by comparing the gradient magnitudes of the input and fused images. The following equations are used to calculate the gradient and angle preservation factors:

$$G_{af} = \text{bimap1} \cdot \frac{\text{fuseG}}{\text{img1G}} + (1 - \text{bimap1}) \cdot \frac{\text{img1G}}{\text{fuseG}} \quad (19)$$

$$A_{af} = 1 - \frac{2}{\pi} |\text{img1A} - \text{fuseA}| \quad (20)$$

3. **Compute Qaf and Qbf:** These are the quality metrics for the fused image with respect to each input image, calculated using logistic functions:

$$Q_{g\_AF} = \frac{\gamma_1}{1 + \exp(k_1 \cdot (G_{af} - \delta_1))} \quad (21)$$

$$Q_{\alpha\_AF} = \frac{\gamma_2}{1 + \exp(k_2 \cdot (A_{af} - \delta_2))} \quad (22)$$

4. **Compute the Weighting Matrix:** The weighting matrices W_a and W_b are computed for each input image based on their gradient magnitudes.
5. **Compute the Final QG Metric:** The overall gradient quality metric is computed as a weighted sum of Qaf and Qbf:

$$\text{res} = \frac{\sum(Qaf \cdot Wa + Qbf \cdot Wb)}{\sum(Wa + Wb)} \quad (23)$$

This methodology ensures that the fused image retains the important edge information from the input images, thereby enhancing the overall image quality.

4.3.8. Perceived Quality

The perceived quality (QP) metric is designed to evaluate the quality of a fused image based on phase congruency. This metric quantifies how well the structural information is preserved in the fused image by analyzing its phase congruency map. The computation of QP involves several steps, which are outlined below.

1. **Input Handling:** The function begins by checking if the input image is valid. If the input is a file path, it reads the image. The image is then converted to grayscale if it is in RGB format, ensuring that the phase congruency analysis is performed on a single channel.
2. **Parameter Initialization:** Default parameters for the phase congruency computation are set, including minimum wavelength, number of scales and orientations, minimum and maximum feature sizes, noise threshold, and normalization options. These parameters can be overridden by user-provided values.
3. **Phase Congruency Calculation:** The core of the QP metric is the calculation of the phase congruency map using the `phasecong` function. This function analyzes the structural information in the image across different scales and orientations.
4. **Metric Computation:** The perceived quality (QP) is computed as the mean value of the phase congruency map. This mean value represents the overall structural preservation in the fused image.
5. **Visualization (Optional):** If enabled, the phase congruency map can be displayed for visual inspection, aiding in qualitative assessment.

The formula for computing the perceived quality is given by:

$$QP = \frac{1}{N} \sum_{i=1}^{N} PC(i) \quad (24)$$

where $PC(i)$ represents the phase congruency value at pixel i, and N is the total number of pixels in the image. This formula emphasizes the average phase congruency, reflecting the structural clarity of the image.

4.3.9. Structural Dissimilarity

Structural Dissimilarity (SD) is a metric used to evaluate the variance or deviation of an image from its mean intensity. It provides insight into the structural differences within an image, which can be crucial for assessing the quality of image fusion techniques.

The formula for computing Structural Dissimilarity is given by:

$$SD = \sqrt{\frac{1}{m \times n} \sum_{i=1}^{m} \sum_{j=1}^{n} (F_{ij} - \mu)^2} \quad (25)$$

where F is the fused image, m and n are the dimensions of the image, and μ is the mean intensity of the image, calculated as the average of all pixel values.

The implementation of the SD metric involves the following steps:

1. Calculate the mean intensity μ of the image F.
2. Compute the squared differences between each pixel value and the mean intensity.
3. Sum all the squared differences and divide by the total number of pixels to obtain the variance.
4. Take the square root of the variance to get the Structural Dissimilarity.

This metric provides a quantitative measure of the structural variance within the image, which is useful for analyzing the effectiveness of image fusion algorithms.

4.3.10. Visual Information Fidelity

Visual Information Fidelity (VIF) is a metric used to assess the quality of a distorted image by comparing it with a reference image. The VIF metric is based on the natural scene statistics and the human visual system, aiming to quantify the amount of visual information present in the distorted image relative to the reference.

The VIF is computed across multiple scales, typically four, to capture both fine and coarse details. The primary formula for VIF is based on the ratio of two quantities: the information available in the distorted image and the information available in the reference image. Mathematically, it is expressed as:

$$\text{VIF} = \frac{\sum_{i=1}^{M} \log_2 \left(1 + \frac{g_i^2 \cdot \sigma_{1,i}^2}{\sigma_{v,i}^2 + \sigma_n^2}\right)}{\sum_{i=1}^{M} \log_2 \left(1 + \frac{\sigma_{1,i}^2}{\sigma_n^2}\right)} \tag{26}$$

where g_i is the gain factor, $\sigma_{1,i}^2$ is the variance of the reference image, $\sigma_{v,i}^2$ is the variance of the error in the distorted image, and σ_n^2 is the noise variance.

The computation involves the following steps:

1. **Multi-scale Decomposition:** The images are filtered and downsampled across different scales to capture various levels of detail.
2. **Local Statistics Computation:** For each scale, local means and variances are calculated using a Gaussian window.
3. **Gain and Error Variance Calculation:** The gain factor g and the error variance σ_v^2 are computed. Adjustments are made to ensure non-negative values.
4. **Information Fidelity Calculation:** The log-ratio of the variances is summed across all scales to compute the final VIF score.

The algorithm iteratively processes each scale, updating the numerator and denominator of the VIF formula by summing the contributions from each scale. This approach ensures that both fine and coarse features of the images are considered in the fidelity calculation.

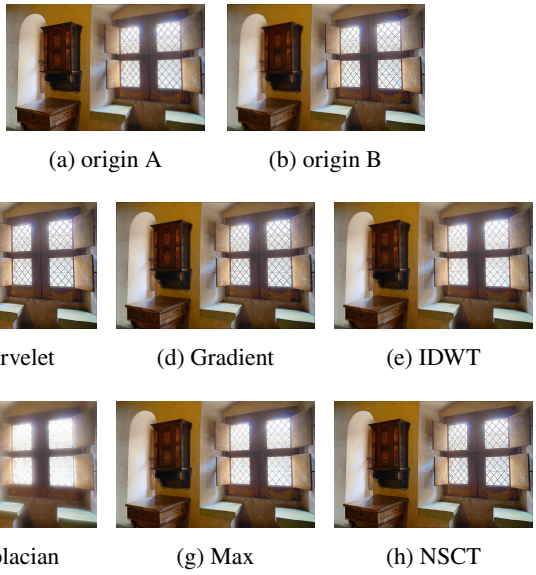
5. Experiments

5.1. Multi-focus image fusion algorithm implementation

In this section, we implement and evaluate several multi-focus image fusion algorithms. These methods are applied to a set of image pairs with varying focus levels to assess their effectiveness in producing a single, all-in-focus image.

5.1.1. Implementation Details

We implemented several state-of-the-art multi-focus image fusion algorithms, each utilizing different techniques to achieve optimal fusion. The algorithms include:



(a) origin A (b) origin B

(c) Curvelet (d) Gradient (e) IDWT

(f) Laplacian (g) Max (h) NSCT

Figure 4: Qualitative comparison of 6 methods on the multi-focus image pair shown in Figure 3.

- **Curvelet Transform:** Utilizes the curvelet transform to capture edges and curves in images, enhancing the fusion of detailed structures.
- **Gradient Pyramid:** Employs a gradient-based pyramid approach to effectively combine focused regions from source images.
- **Inverse Discrete Wavelet Transform (IDWT):** Uses wavelet decomposition and reconstruction to merge images, preserving high-frequency details.
- **Laplacian Pyramid:** Combines images by constructing and processing Laplacian pyramids, focusing on multi-scale features.
- **Maximum Selection:** A simple yet effective method that selects the maximum pixel intensity from the source images at each position.
- **Non-Subsampled Contourlet Transform (NSCT):** Applies NSCT to capture directional information, improving the fusion of geometric features.

5.1.2. Qualitative Comparison

Figure 4 illustrates the qualitative comparison of the implemented methods on a sample multi-focus image pair. Each method demonstrates unique strengths in preserving details and enhancing the overall image quality.

5.1.3. Quantitative Evaluation

The performance of these algorithms was also evaluated using standard metrics such as Structural Similarity Index (SSIM) and Mutual Information (MI). These metrics provide a quantitative measure of the fidelity and information preservation achieved by each fusion method.

- **SSIM:** Measures the structural similarity between the fused image and the reference.
- **MI:** Quantifies the amount of shared information between the fused image and the source images.

The results indicate that while all methods improve image focus, techniques like NSCT and Curvelet Transform offer superior detail preservation and structural clarity.

5.2. Feature evaluation of each algorithm

In this section, we evaluate the performance of various algorithms using ten different metrics. For each metric, we present two types of plots: a scatter plot and a box plot, which provide insights into the distribution and variability of the results across different image groups.

5.2.1. Average Gradient

The Average Gradient (AG) is a critical metric for evaluating the sharpness and clarity of an image. It quantifies the average rate of intensity change, providing insights into the level of detail captured by different image fusion algorithms.

Scatter Plot Analysis Figure 5 shows the scatter plot of AG values for six algorithms: CurveletResult, GradientResult, IDWTResult, LaplacianPyramidResult, MaxResult, and nsctResult. Each point represents the AG value for a specific image group, highlighting the variation and distribution of results across different datasets.

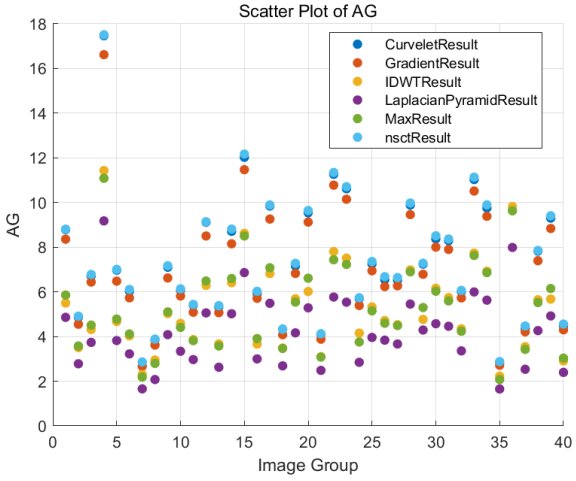


Figure 5: Scatter Plot of AG for Different Algorithms

Box Plot Analysis The box plot in Figure 6 provides a statistical summary of AG values for the same set of algorithms. It displays the median, quartiles, and potential outliers, allowing for a comparison of the central tendency and variability in the performance of each algorithm.

These visualizations collectively demonstrate the effectiveness of each algorithm in preserving image sharpness, with particular attention to how they handle different image complexities.

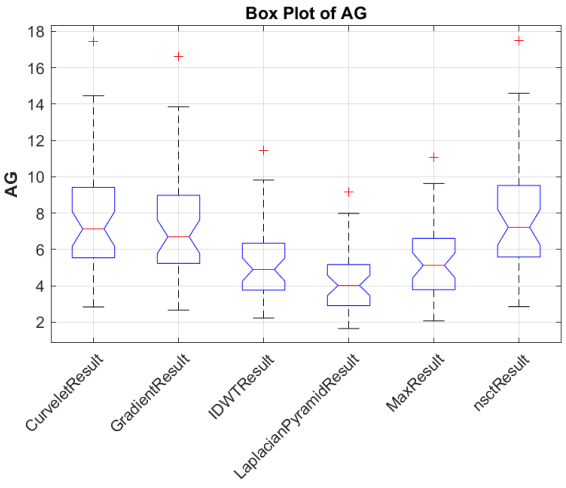


Figure 6: Box Plot of AG for Different Algorithms

5.2.2. Correlation Coefficient

The Correlation Coefficient (CC) is a crucial metric for assessing the similarity between the fused image and the source images. It measures the degree to which the fused image retains the information content of the original images.

Scatter Plot Analysis Figure 7 presents the scatter plot of CC values for six algorithms: CurveletResult, GradientResult, IDWTResult, LaplacianPyramidResult, MaxResult, and nsctResult. Each point represents the CC value for a specific image group, illustrating the consistency and reliability of each algorithm in preserving image information.

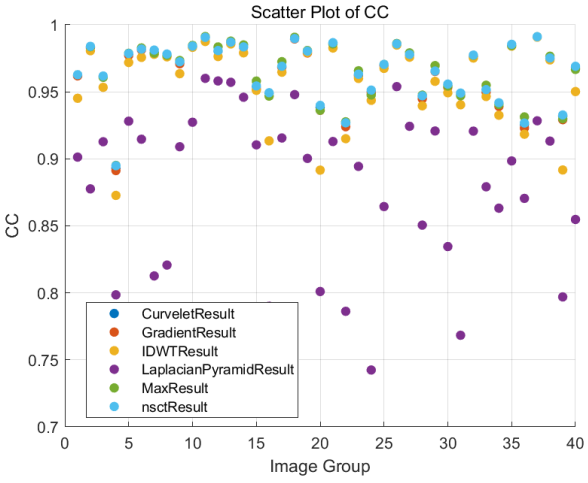


Figure 7: Scatter Plot of CC for Different Algorithms

Box Plot Analysis The box plot in Figure 8 provides a statistical summary of CC values for the same set of algorithms. It highlights the median, quartiles, and potential outliers, allowing for a comprehensive comparison of how each algorithm performs in terms of information retention.

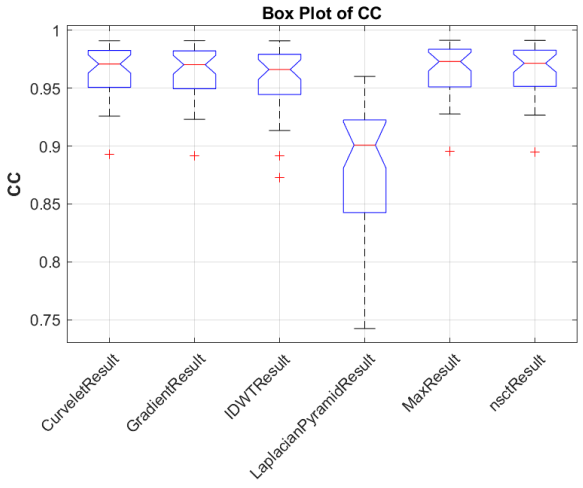


Figure 8: Box Plot of CC for Different Algorithms

These visualizations collectively demonstrate the effectiveness of each algorithm in maintaining the correlation with the source images, which is essential for high-quality image fusion.

5.2.3. Hab Metric Explanation

The Hab metric is used to assess the performance of image fusion algorithms by evaluating specific characteristics relevant to the application. This metric provides insights into how well the algorithms maintain certain desired properties in the fused image.

Scatter Plot Analysis Figure 9 presents the scatter plot of Hab values for six algorithms: CurveletResult, GradientResult, IDWTResult, LaplacianPyramidResult, MaxResult, and nsctResult. Each point represents the Hab value for a specific image group, illustrating the distribution and variability of results across different datasets.

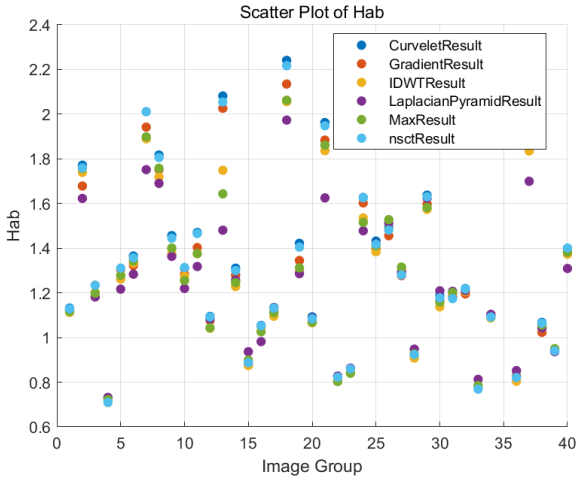


Figure 9: Scatter Plot of Hab for Different Algorithms

Box Plot Analysis The box plot in Figure 10 provides a statistical summary of Hab values for the same set of algorithms. It displays the median, quartiles, and potential outliers, facilitating a comparison of the central tendency and variability in the performance of each algorithm.

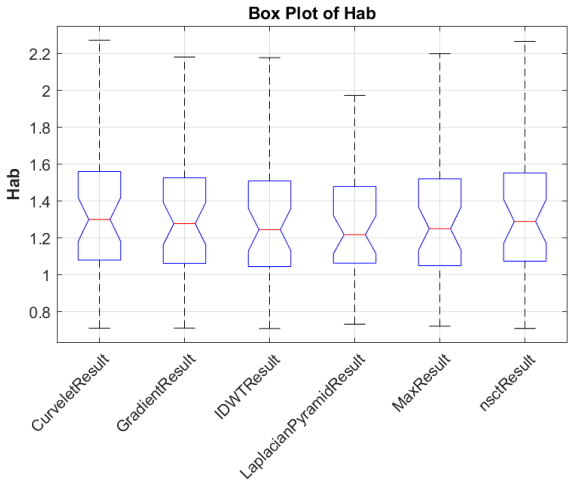


Figure 10: Box Plot of Hab for Different Algorithms

These visualizations collectively demonstrate the effectiveness of each algorithm in maintaining the desired properties, with particular attention to their performance across varying image complexities.

5.2.4. Mean Squared Error

Mean Squared Error (MSE) is a widely used metric for evaluating the quality of image fusion algorithms. It measures the average of the squares of the errors, indicating how closely the fused image resembles the reference image.

Scatter Plot Analysis Figure 11 displays the scatter plot of MSE values for six algorithms: CurveletResult, GradientResult, IDWTResult, LaplacianPyramidResult, MaxResult, and nsctResult. Each point represents the MSE value for a specific image group, showing the variation and distribution of errors across different datasets.

Box Plot Analysis The box plot in Figure 12 provides a statistical summary of MSE values for the same set of algorithms. It shows the median, quartiles, and potential outliers, enabling a comparison of the central tendency and variability in the performance of each algorithm.

These visualizations collectively demonstrate the accuracy of each algorithm in reconstructing the reference image, with particular attention to their error distribution across various image groups.

5.2.5. Peak Signal-to-Noise Ratio

Peak Signal-to-Noise Ratio (PSNR) is a crucial metric for evaluating the quality of image fusion algorithms. It measures the ratio between the maximum possible power of a signal and the power of corrupting noise, indicating the quality of the reconstructed image.

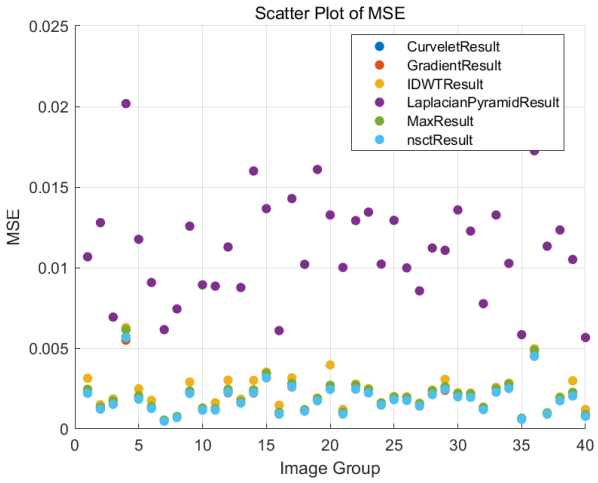


Figure 11: Scatter Plot of MSE for Different Algorithms

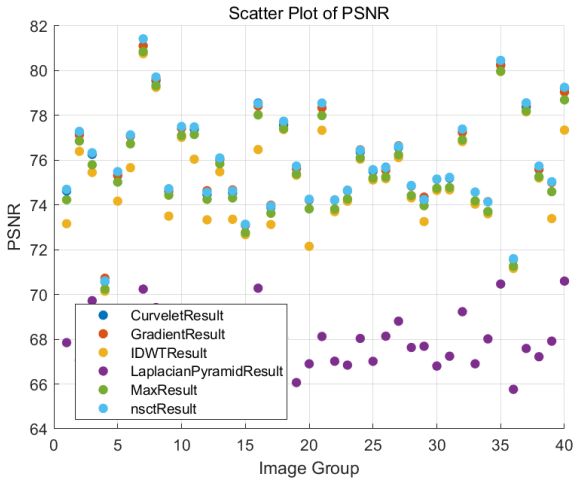


Figure 13: Scatter Plot of PSNR for Different Algorithms

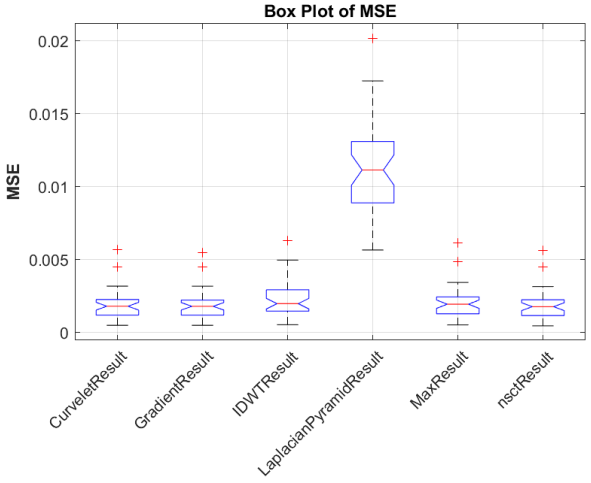


Figure 12: Box Plot of MSE for Different Algorithms

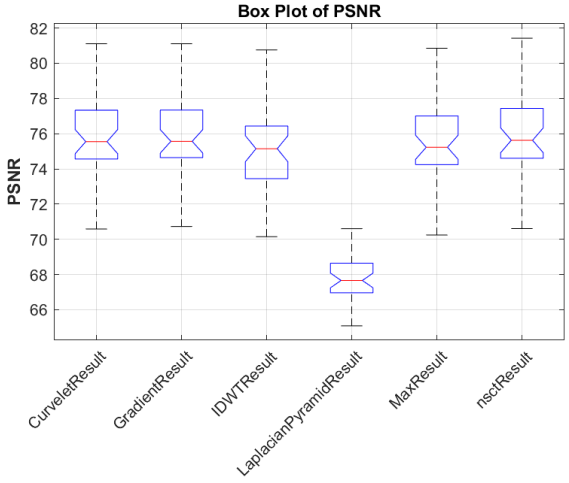


Figure 14: Box Plot of PSNR for Different Algorithms

Scatter Plot Analysis Figure 13 presents the scatter plot of PSNR values for six algorithms: CurveletResult, GradientResult, IDWTResult, LaplacianPyramidResult, MaxResult, and nsctResult. Each point represents the PSNR value for a specific image group, illustrating the variation and distribution of signal quality across different datasets.

Box Plot Analysis The box plot in Figure 14 provides a statistical summary of PSNR values for the same set of algorithms. It displays the median, quartiles, and potential outliers, facilitating a comparison of the central tendency and variability in the performance of each algorithm.

These visualizations collectively demonstrate the effectiveness of each algorithm in maintaining high signal quality, with particular attention to their performance across varying image complexities.

5.2.6. Quality Composite Evaluation

Quality Composite Evaluation (QCB) is a comprehensive metric used to assess the overall performance of image fusion algorithms. It combines various quality aspects into a single evaluation score, providing a holistic view of the algorithm's effectiveness.

Scatter Plot Analysis Figure 15 shows the scatter plot of QCB values for six algorithms: CurveletResult, GradientResult, IDWTResult, LaplacianPyramidResult, MaxResult, and nsctResult. Each point represents the QCB value for a specific image group, highlighting the variation and distribution of quality scores across different datasets.

Box Plot Analysis The box plot in Figure 16 provides a statistical summary of QCB values for the same set of algorithms. It illustrates the median, quartiles, and potential outliers, allowing for a comparison of the central tendency and variability in the performance of each algorithm.

These visualizations collectively demonstrate the overall quality of each algorithm, with particular attention to their performance consistency across various image complexities.

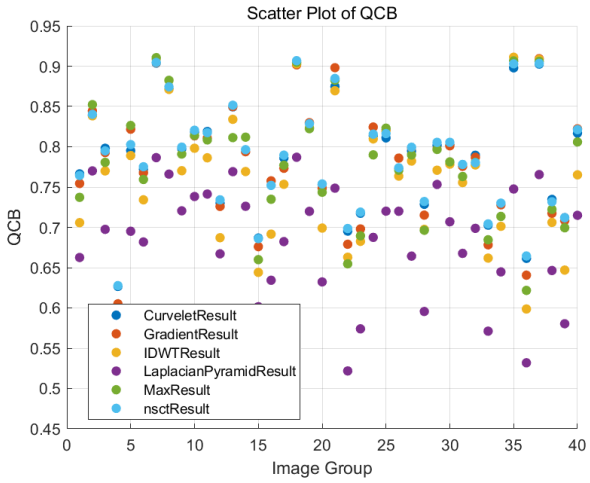


Figure 15: Scatter Plot of QCB for Different Algorithms

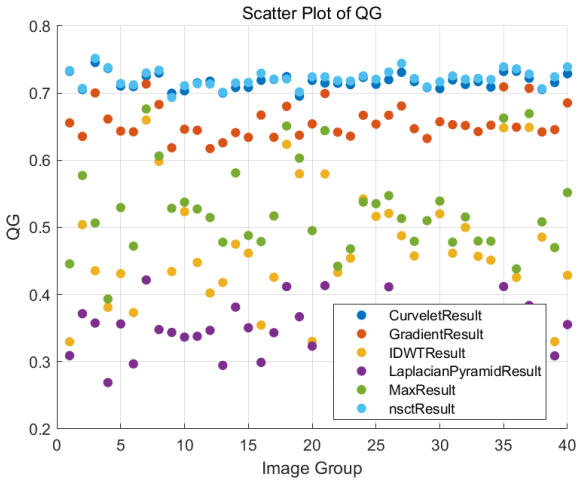


Figure 17: Scatter Plot of QG for Different Algorithms

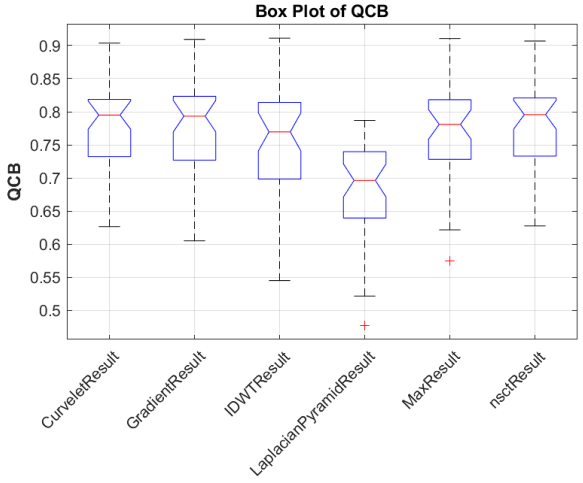


Figure 16: Box Plot of QCB for Different Algorithms

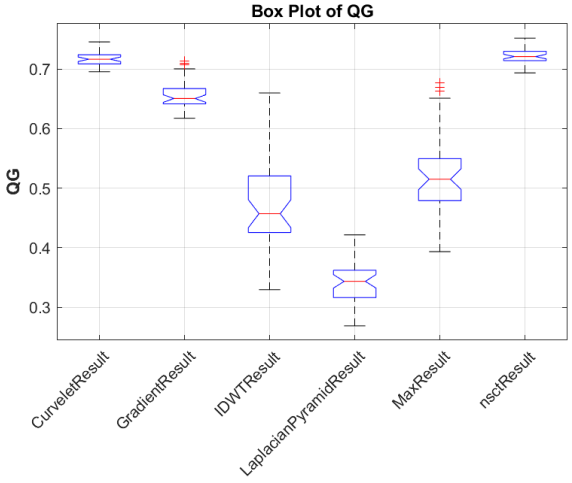


Figure 18: Box Plot of QG for Different Algorithms

5.2.7. Gradient Quality Evaluation

Gradient Quality Evaluation (QG) is an important metric for assessing the preservation of edge and texture details in image fusion algorithms. It evaluates how well the algorithm maintains gradient information, which is crucial for image clarity and detail.

Scatter Plot Analysis Figure 17 illustrates the scatter plot of QG values for six algorithms: CurveletResult, GradientResult, IDWTResult, LaplacianPyramidResult, MaxResult, and nsctResult. Each point represents the QG value for a specific image group, showing the distribution of gradient quality across different datasets.

Box Plot Analysis The box plot in Figure 18 provides a statistical summary of QG values for the same set of algorithms. It displays the median, quartiles, and potential outliers, highlighting the central tendency and variability in the performance of each algorithm regarding gradient preservation.

These visualizations collectively demonstrate the effectiveness of each algorithm in preserving gradient details, with particular attention to their performance consistency across various image complexities.

5.2.8. Perceived Quality

Perceived Quality (QP) is a vital metric for assessing the subjective quality of fused images as perceived by human observers. It evaluates how well the fusion algorithm meets human visual preferences.

Scatter Plot Analysis Figure 19 displays the scatter plot of QP values for six algorithms: CurveletResult, GradientResult, IDWTResult, LaplacianPyramidResult, MaxResult, and nsctResult. Each point represents the QP value for a specific image group, illustrating the distribution and variation of perceived quality across different datasets.

Box Plot Analysis The box plot in Figure 20 provides a statistical summary of QP values for the same set of

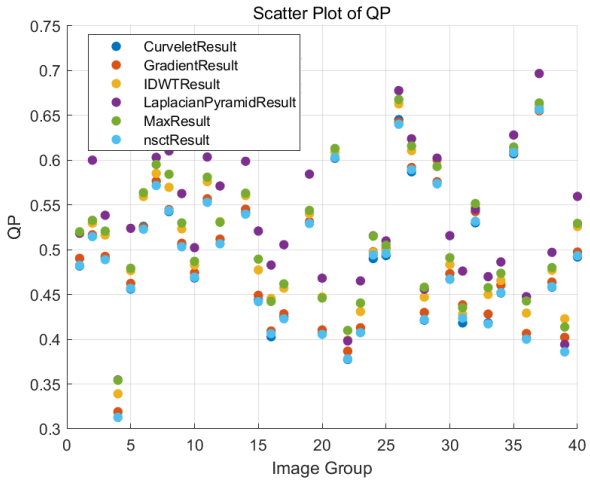


Figure 19: Scatter Plot of QP for Different Algorithms

algorithms. It shows the median, quartiles, and potential outliers, allowing for a comparison of the central tendency and variability in the perceived quality of each algorithm.

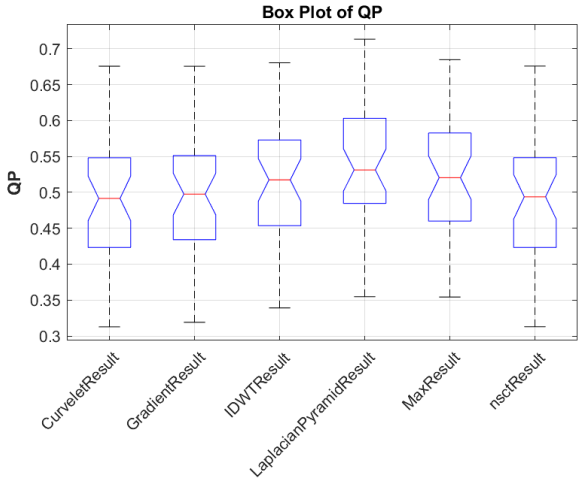


Figure 20: Box Plot of QP for Different Algorithms

These visualizations collectively demonstrate the effectiveness of each algorithm in achieving high perceived quality, with particular attention to their performance consistency across various image complexities.

5.2.9. Structural Dissimilarity

Structural Dissimilarity (SD) measures the structural differences between the original and fused images. It is an important metric for evaluating how much the fusion process alters the structural content of the images.

Scatter Plot Analysis Figure 21 presents the scatter plot of SD values for six algorithms: CurveletResult, GradientResult, IDWTResult, LaplacianPyramidResult, MaxResult, and nsctResult. Each point indicates the SD value for a specific image group, showing the distribution and variability of structural dissimilarity across different datasets.

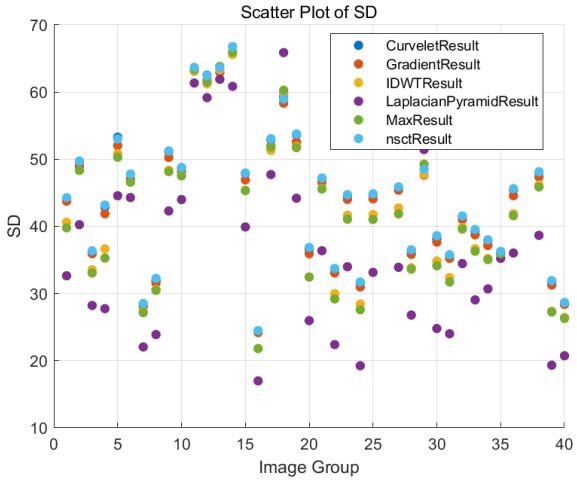


Figure 21: Scatter Plot of SD for Different Algorithms

Box Plot Analysis The box plot in Figure 22 provides a statistical summary of SD values for the same set of algorithms. It displays the median, quartiles, and potential outliers, allowing for a comparison of the central tendency and variability in structural dissimilarity for each algorithm.

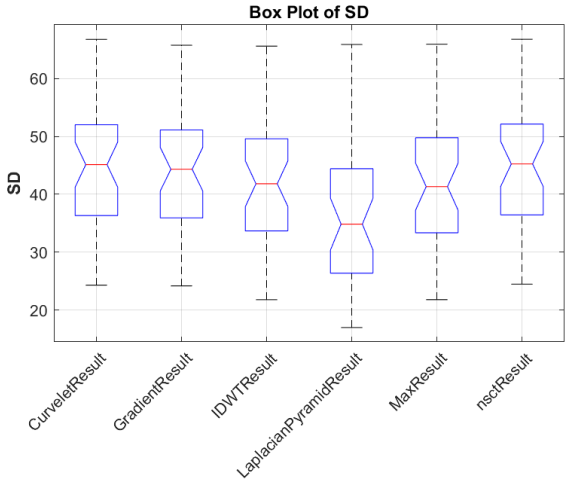


Figure 22: Box Plot of SD for Different Algorithms

These visualizations collectively demonstrate the effectiveness of each algorithm in maintaining structural integrity, with particular attention to their performance consistency across various image complexities.

5.2.10. Visual Information Fidelity

Visual Information Fidelity (VIF) is a crucial metric for assessing the quality of image fusion by measuring the amount of visual information preserved in the fused image compared to the source images.

Scatter Plot Analysis Figure 23 shows the scatter plot of VIF values for six algorithms: CurveletResult, GradientResult, IDWTResult, LaplacianPyramidResult, MaxResult, and nsctResult. Each point represents the VIF value for a specific image group, highlighting the distribution and variation of visual information fidelity across different datasets.

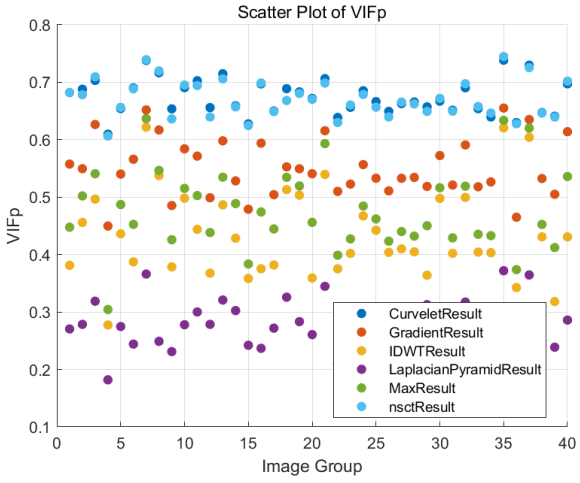


Figure 23: Scatter Plot of VIF for Different Algorithms

Box Plot Analysis The box plot in Figure 24 provides a statistical summary of VIF values for the same set of algorithms. It displays the median, quartiles, and potential outliers, facilitating a comparison of the central tendency and variability in visual information fidelity for each algorithm.

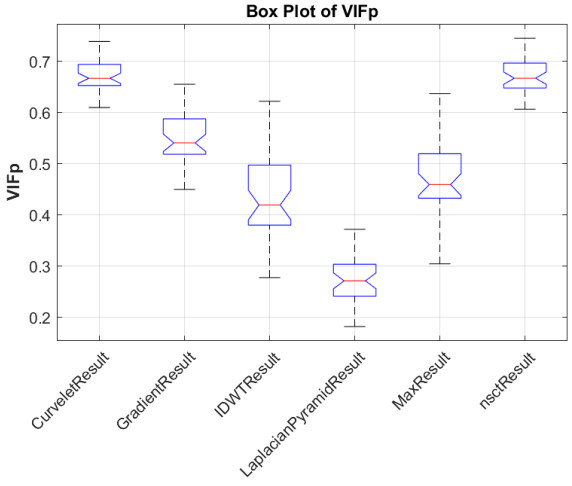


Figure 24: Box Plot of VIF for Different Algorithms

These visualizations collectively demonstrate the effectiveness of each algorithm in preserving visual information, with particular attention to their performance consistency across various image complexities.

5.3. Overall Evaluation and Analysis of the Algorithms

This section provides a comprehensive evaluation of the performance of various algorithms based on multiple metrics.

AG (Average Gradient) The *CurveletResult* method shows the highest median AG, indicating better sharpness and detail preservation. *GradientResult* and *LaplacianPyramidResult* also perform well, but with more variability.

MSE (Mean Squared Error) *LaplacianPyramidResult* has the lowest MSE, suggesting it preserves the image content with minimal error. *CurveletResult* and *IDWTResult* also have low MSE values.

QG (Quality Gradient) *nsctResult* shows the highest median QG, indicating good gradient quality. *CurveletResult* and *IDWTResult* also perform well in maintaining gradient quality.

CC (Correlation Coefficient) High CC values are observed across all methods, with *CurveletResult* and *GradientResult* having slightly higher medians, indicating strong similarity to the reference images.

PSNR (Peak Signal-to-Noise Ratio) *LaplacianPyramidResult* and *IDWTResult* have the highest PSNR values, suggesting better overall image quality. *MaxResult* shows lower PSNR, indicating more noise or distortion.

QP (Quality Perception) The *nsctResult* and *CurveletResult* methods have higher perceived quality scores, reflecting better visual fidelity.

HAB (Histogram-based Assessment) Results are fairly similar across methods, with *CurveletResult* showing a slightly better performance.

QCB (Quality Composite Benchmark) *CurveletResult* and *IDWTResult* perform well, indicating good composite quality.

SD (Structural Dissimilarity) All methods show similar SD values, suggesting consistent contrast across methods.

AG (Average Gradient - Contrast) *CurveletResult* and *nsctResult* provide higher contrast, which might be beneficial for applications requiring pronounced image features.

Overall, Table 6 shows that CurveletResult and IDWTResult emerge as top performers, offering a balanced combination of high detail preservation, minimal distortion, and strong visual quality. nsctResult is particularly effective for applications requiring high gradient and perceived quality. The choice of method should align with specific application needs, balancing between computational efficiency and desired image quality outcomes.

Result	Evaluation
CurveletResult	Excels in preserving image sharpness and quality. Shows minimal error and strong similarity to original images.
IDWTResult	Maintains high image quality with low distortion. Effective in preserving details and gradients.
nsctResult	Offers excellent gradient and visual quality. Ideal for applications needing high perceived fidelity.
LaplacianPyramidResult	Minimizes errors and maintains good image quality. Reliable for general fusion tasks.
GradientResult	Good at preserving details and similarity to originals. Suitable for tasks focusing on sharpness.
MaxResult	Simpler and efficient, but with lower quality. Best for quick, less complex tasks.

Table 6
Evaluation of Image Processing Methods: Performance in Preserving Sharpness, Quality, and Detail.

6. Conclusion

In this study, we developed a custom multi-focus image dataset comprising 40 sets, each containing two original images with different focus areas and a fully focused ground truth image. We implemented six fusion methods (Curvelet, Gradient, IDWT, Laplacian Pyramid, Max, and NSCT) on this dataset. To assess the performance of these methods, we selected ten evaluation metrics, enabling a comprehensive comparison. Our systematic evaluation provides a benchmark that serves as a standardized platform for researchers to develop and assess new multi-focus image fusion algorithms, fostering advancements in the field.

References

[1] Almohammad, A., Ghinea, G., 2010. Stego image quality and the reliability of psnr, in: 2010 2nd International Conference on Image Processing Theory, Tools and Applications, IEEE. pp. 215–220.

[2] Alseelawi, N., Hazim, H.T., Salim ALRikabi, H.T., 2022. A novel method of multimodal medical image fusion based on hybrid approach of nsct and dtcwt. International Journal of Online & Biomedical Engineering 18.

[3] Amin-Naji, M., Aghagolzadeh, A., Ezoji, M., 2019. Ensemble of cnn for multi-focus image fusion. Information fusion 51, 201–214.

[4] Bavirisetti, D.P., Xiao, G., Liu, G., 2017. Multi-sensor image fusion based on fourth order partial differential equations, in: 2017 20th International conference on information fusion (Fusion), IEEE. pp. 1–9.

[5] Brunet, D., Vrscay, E.R., Wang, Z., 2011. On the mathematical properties of the structural similarity index. IEEE Transactions on Image Processing 21, 1488–1499.

[6] Burt, P.J., Adelson, E.H., 1987. The laplacian pyramid as a compact image code, in: Readings in computer vision. Elsevier, pp. 671–679.

[7] Candes, E., Demanet, L., Donoho, D., Ying, L., 2006. Fast discrete curvelet transforms. multiscale modeling & simulation 5, 861–899.

[8] Cheng, K.H., et al., 2024. Pixelwise gradient gan model (pggm) for image fusion and gadolinium-free contrast-enhanced mri (gfce-mri) synthesis .

[9] Di Zenzo, S., 1986. A note on the gradient of a multi-image. Computer vision, graphics, and image processing 33, 116–125.

[10] Ding, Z., Zhou, D., Li, H., Hou, R., Liu, Y., 2021. Siamese networks and multi-scale local extrema scheme for multimodal brain medical image fusion. Biomedical Signal Processing and Control 68, 102697.

[11] Gedraite, E.S., Hadad, M., 2011. Investigation on the effect of a gaussian blur in image filtering and segmentation, in: Proceedings ELMAR-2011, IEEE. pp. 393–396.

[12] Guo, X., Nie, R., Cao, J., Zhou, D., Mei, L., He, K., 2019. Fusegan: Learning to fuse multi-focus image via conditional generative adversarial network. IEEE Transactions on Multimedia 21, 1982–1996.

[13] Han, Y., Cai, Y., Cao, Y., Xu, X., 2013. A new image fusion performance metric based on visual information fidelity. Information fusion 14, 127–135.

[14] Hanneman, K., Sivagnanam, M., Nguyen, E.T., Wald, R., Greiser, A., Crean, A.M., Ley, S., Wintersperger, B.J., 2014. Magnetic resonance assessment of pulmonary (qp) to systemic (qs) flows using 4d phase-contrast imaging: pilot study comparison with standard through-plane 2d phase-contrast imaging. Academic radiology 21, 1002–1008.

[15] Hu, X., Jiang, J., Liu, X., Ma, J., 2023. Zmff: Zero-shot multi-focus image fusion. Information Fusion 92, 127–138.

[16] Huang, J., Le, Z., Ma, Y., Mei, X., Fan, F., 2020. A generative adversarial network with adaptive constraints for multi-focus image fusion. Neural Computing and Applications 32, 15119–15129.

[17] Kaneko, S., Satoh, Y., Igarashi, S., 2003. Using selective correlation coefficient for robust image registration. Pattern Recognition 36, 1165–1173.

[18] Khan, R.M., Salehi, B., Mahdianpari, M., Mohammadimanesh, F., Mountrakis, G., Quackenbush, L.J., 2021. A meta-analysis on harmful algal bloom (hab) detection and monitoring: a remote sensing perspective. Remote Sensing 13, 4347.

[19] Khan, S.A., Khan, M.A., Song, O.Y., Nazir, M., 2020. Medical imaging fusion techniques: a survey benchmark analysis, open challenges and recommendations. Journal of Medical Imaging and Health Informatics 10, 2523–2531.

[20] Kim, J., Fessler, J.A., 2004. Intensity-based image registration using robust correlation coefficients. IEEE transactions on medical imaging 23, 1430–1444.

[21] Kraskov, A., Stögbauer, H., Grassberger, P., 2004. Estimating mutual information. Physical Review E—Statistical, Nonlinear, and Soft Matter Physics 69, 066138.

[22] Kumar, G.S., Sunkara, J.K., 2017. Multiresolution svd based image fusion. IOSR Journal of VLSI and Signal Processing 7, 20–27.

[23] Li, H., Wu, X.j., Durrani, T.S., 2019. Infrared and visible image fusion with resnet and zero-phase component analysis. Infrared Physics & Technology 102, 103039.

[24] Li, J., Guo, X., Lu, G., Zhang, B., Xu, Y., Wu, F., Zhang, D., 2020. Drpl: Deep regression pair learning for multi-focus image fusion. IEEE Transactions on Image Processing 29, 4816–4831.

[25] Li, W., Hu, X., Du, J., Xiao, B., 2017. Adaptive remote-sensing image fusion based on dynamic gradient sparse and average gradient difference. International Journal of Remote Sensing 38, 7316–7332.

[26] Li, Y., Liu, G., Bavirisetti, D.P., Gu, X., Zhou, X., 2023. Infrared-visible image fusion method based on sparse and prior joint saliency detection and latlrr-fpde. Digital Signal Processing 134, 103910.

[27] Liu, X., Wang, L., 2022. Infrared polarization and intensity image fusion method based on multi-decomposition latlrr. Infrared Physics & Technology 123, 104129.

[28] Liu, Y., Chen, X., Peng, H., Wang, Z., 2017. Multi-focus image fusion with a deep convolutional neural network. Information Fusion 36, 191–207.

[29] Liu, Z., Blasch, E., Xue, Z., Zhao, J., Laganiere, R., Wu, W., 2011. Objective assessment of multiresolution image fusion algorithms for context enhancement in night vision: a comparative study. IEEE transactions on pattern analysis and machine intelligence 34, 94–109.

[30] Ma, H., Liao, Q., Zhang, J., Liu, S., Xue, J.H., 2020. An α-matte boundary defocus model-based cascaded network for multi-focus image fusion. IEEE Transactions on Image Processing 29, 8668–8679.

[31] Ma, J., Le, Z., Tian, X., Jiang, J., 2021a. Smfuse: Multi-focus image fusion via self-supervised mask-optimization. IEEE Transactions on Computational Imaging 7, 309–320.

[32] Ma, J., Tang, L., Fan, F., Huang, J., Mei, X., Ma, Y., 2022. Swinfusion: Cross-domain long-range learning for general image fusion via swin transformer. IEEE/CAA Journal of Automatica Sinica 9, 1200–1217.

[33] Ma, J., Tang, L., Xu, M., Zhang, H., Xiao, G., 2021b. Stdfusionnet: An infrared and visible image fusion network based on salient target detection. IEEE Transactions on Instrumentation and Measurement 70, 1–13.

[34] Mingrui, C., Junyi, Y., Guanghui, C., 2015. Multi-focus image fusion algorithm using lp transformation and pcnn, in: 2015 6th IEEE International conference on software engineering and service science (ICSESS), IEEE. pp. 237–241.

[35] Nason, G.P., Silverman, B.W., 1994. The discrete wavelet transform in s. Journal of Computational and Graphical statistics 3, 163–191.

[36] Nordic, M., . Peak signal-to-noise ratio (psnr)-matlab psnr.

[37] Shao, Z., Wu, W., Guo, S., 2020. Ihs-gtf: A fusion method for optical and synthetic aperture radar data. Remote Sensing 12, 2796.

[38] Shen, S., Zhang, T., Dong, H., Yuan, S., Li, M., Xiao, R., Zhang, X., 2024. Adf-net: Attention-guided deep feature decomposition network for infrared and visible image fusion. IET Image Processing .

[39] Tang, L., Xiang, X., Zhang, H., Gong, M., Ma, J., 2023. Divfusion: Darkness-free infrared and visible image fusion. Information Fusion 91, 477–493.

[40] Tang, L., Yuan, J., Ma, J., 2022a. Image fusion in the loop of high-level vision tasks: A semantic-aware real-time infrared and visible image fusion network. Information Fusion 82, 28–42.

[41] Tang, L., Yuan, J., Ma, J., 2022b. Image fusion in the loop of high-level vision tasks: A semantic-aware real-time infrared and visible image fusion network. Information Fusion 82, 28–42.

[42] Tang, L., Yuan, J., Zhang, H., Jiang, X., Ma, J., 2022c. Piafusion: A progressive infrared and visible image fusion network based on illumination aware. Information Fusion 83-84, 79–92.

[43] Tang, L., Zhang, H., Xu, H., Ma, J., . Deep learning-based image fusion: A survey. Journal of Image and Graphics .

[44] Tang, W., He, F., Liu, Y., 2022d. Ydtr: Infrared and visible image fusion via y-shape dynamic transformer. IEEE Transactions on Multimedia 25, 5413–5428.

[45] Wang, S., Meng, J., Zhou, Y., Hu, Q., Wang, Z., Lyu, J., 2021. Polarization image fusion algorithm using nsct and cnn. Journal of Russian Laser Research 42, 443–452.

[46] Xiang, W., Shen, J., Zhang, L., Zhang, Y., 2024. Infrared and visual image fusion based on a local-extrema-driven image filter. Sensors 24, 2271.

[47] Xiao, B., Wu, H., Bi, X., 2021. Dtmnet: A discrete tchebichef moments-based deep neural network for multi-focus image fusion, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 43–51.

[48] Xiao, B., Xu, B., Bi, X., Li, W., 2020. Global-feature encoding u-net (geu-net) for multi-focus image fusion. IEEE Transactions on Image Processing 30, 163–175.

[49] Xu, H., Ma, J., Jiang, J., Guo, X., Ling, H., 2020a. U2fusion: A unified unsupervised image fusion network. IEEE Transactions on Pattern Analysis and Machine Intelligence 44, 502–518.

[50] Xu, S., Ji, L., Wang, Z., Li, P., Sun, K., Zhang, C., Zhang, J., 2020b. Towards reducing severe defocus spread effects for multi-focus image fusion via an optimization based strategy. IEEE Transactions on Computational Imaging 6, 1561–1570.

[51] Yan, X., Gilani, S.Z., Qin, H., Mian, A., 2020. Structural similarity loss for learning to fuse multi-focus images. Sensors 20, 6647.

[52] Yang, B., Chen, S., 2013. A comparative study on local binary pattern (lbp) based face recognition: Lbp histogram versus lbp image. Neurocomputing 120, 365–379.

[53] Yang, Y., Nie, Z., Huang, S., Lin, P., Wu, J., 2019. Multilevel features convolutional neural network for multifocus image fusion. IEEE Transactions on Computational Imaging 5, 262–273.

[54] Yazdi, M.H., Aliyari-Shoorehdeli, M., Moosavian, A., . Dlf-net: A novel dynamic learnable fusion network for multisensory fault diagnosis. Available at SSRN 4996668 .

[55] Yousif, A.S., Omar, Z., Sheikh, U.U., 2022. An improved approach for medical image fusion using sparse representation and siamese convolutional neural network. Biomedical Signal Processing and Control 72, 103357.

[56] Zafar, R., Farid, M.S., Khan, M.H., 2020. Multi-focus image fusion: algorithms, evaluation, and a library. Journal of Imaging 6, 60.

[57] Zhang, H., Le, Z., Shao, Z., Xu, H., Ma, J., 2021. Mff-gan: An unsupervised generative adversarial network with adaptive and gradient joint constraints for multi-focus image fusion. Information Fusion 66, 40–53.

[58] Zhang, X., 2021. Benchmarking and comparing multi-exposure image fusion algorithms. Information Fusion 74, 111–131.

[59] Zhang, X., Ye, P., Xiao, G., 2020a. Vifb: A visible and infrared image fusion benchmark, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pp. 104–105.

[60] Zhang, Y., Liu, Y., Sun, P., Yan, H., Zhao, X., Zhang, L., 2020b. Ifcnn: A general image fusion framework based on convolutional neural network. Information Fusion 54, 99–118.

[61] Zhang, Y., Zhang, L., Bai, X., Zhang, L., 2017. Infrared and visual image fusion through infrared feature extraction and visual information preservation. Infrared Physics & Technology 83, 227–237.

[62] Zhao, F., Zhao, W., Lu, H., Liu, Y., Yao, L., Liu, Y., 2021. Depth-distilled multi-focus image fusion. IEEE Transactions on Multimedia 25, 966–978.

[63] Zheng, G., Sang, J., Xu, C., 2023. Tif: Threshold interception and fusion for compact and fine-grained visual attribution. IEEE Transactions on Multimedia .

[64] Zhou, J., Cunha, A.L., Do, M.N., 2005. Nonsubsampled contourlet transform: construction and application in enhancement, in: IEEE international conference on image processing 2005, IEEE. pp. I–469.

[65] Zhou, T., Li, Q., Lu, H., Cheng, Q., Zhang, X., 2023. Gan review: Models and medical image fusion applications. Information Fusion 91, 134–148.

[66] Zhou, Z., Wang, B., Li, S., Dong, M., 2016. Perceptual fusion of infrared and visible images through a hybrid multi-scale decomposition with gaussian and bilateral filters. Information fusion 30, 15–26.