



A Hybrid Multiscale and CNN-Based Fusion Framework for Enhanced SAR–MS Image Integration

Priyanka S. Tondewad^{*ID}, Manisha P. Dale^{ID}

Department of Electronics and Telecommunication Engineering, Savitribai Phule Pune University (SPPU), Pune 411007, India

Corresponding Author Email: priyanka.tondewad@mescoepune.org

Copyright: ©2025 The authors. This article is published by IIETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/mmep.121024>

ABSTRACT

Received: 14 August 2025

Revised: 8 October 2025

Accepted: 14 October 2025

Available online: 31 October 2025

Keywords:

satellite image fusion, SAR image, multispectral image, multiscale decomposition, multilayer fusion, augmented dataset, RGF

A successful image fusion framework should effectively integrate the most of information from both Synthetic Aperture Radar (SAR) images and Multispectral (MS) images into the fused image, while reducing the presence of artifacts. SAR and MS image fusion is a fundamental key technology for image quality improvement. Proposed algorithm performs multiscale decomposition on SAR image and intensity (I) component of multispectral image, using optimized Rolling Guidance Filter (RGF). This approach separates the important scale space features into three categories: approximation layer, contour layer and detail layer. Suitable fusion rules to preserve spatial information are designed to fuse approximation layers, contour layers and detail layers based on information contained in that layer. This obtained image is integrated with CNN-fused image to preserve textural information. In the CNN-based approach, the designed network is trained to determine weights from an augmented dataset that includes both types of images. Through comprehensive experimentation, the proposed approach demonstrated robust spectral information preservation, achieving a spectral information metric Erreur Relative Globale Adimensionnelle de Synthese (ERGAS) less than 3 and Spectral Angle Mapper (SAM) less than 1 which represents an optimal balance between contrast and correlation. Moreover, the spatial detail measure Peak Signal to Noise Ratio (PSNR) more than 30 and Average Gradient (AG) maintained in the ideal range, validates the method's capability to retain spatial features. Also, visual interpretation clearly highlights the perceptible difference among the images.

1. INTRODUCTION

In general, the high-resolution satellite images get captured with a long revisit time, while low-resolution satellites revisit frequently. Various applications need satellite images with high spatial as well as spectral resolution [1]. Satellite images can be used for fishing prediction, crop species classification, soil analysis, Land Use Land Cover (LULC) changes, biomass and mineralogy mapping, aerospace and shipbuilding, to name a few. This rich information dataset can be obtained by the combination of two or more dissimilar images of same scene, to form new image with an improved quality and reliability [2, 3]. To obtain such usable dataset satellite image fusion of multitemporal and multiresolution images have been researched in recent decades.

1.1 SAR image

Synthetic Aperture Radar (SAR) is an active microwave sensor that utilizes long-wavelength electromagnetic radiations, enabling it to effectively penetrate atmospheric impairments such as cloud cover, haze, dust, fog, smoke and other adverse environmental factors with the exception of intense penetration. This characteristic allows SAR imagery to be captured continuously, regardless of the time of day or

weather conditions. This mechanism makes SAR imagery richer in spatial information.

1.2 Multispectral image

Multispectral sensors are passive remote sensing instruments that capture solar radiation reflected from the Earth's surface across different parts of the electromagnetic spectrum, such as ultraviolet, visible, and infrared wavelengths. Depending on their spectral resolution, optical sensors are categorized into panchromatic, multispectral and hyperspectral types. Multispectral imaging provides extensive information regarding the spectral signatures of terrestrial objects, thereby facilitating the differentiation of various land cover types. However, its effectiveness is significantly influenced by solar illumination and prevailing weather conditions. As these sensors exhibit high spectral sensitivity, it is more readable than SAR images. Therefore, fusion of SAR and multispectral image can provide more informative image with respect to spatial and spectral information. Broadly, multispectral and SAR image fusion is categorised into three different levels: pixel level, decision level, and feature level as per the different levels of information fusion [4-6]. This article focuses on pixel level fusion, offering a more comprehensive discussion on this technique than other

methods. Pixel level fusion attains more precise and detailed information, facilitating enhanced interpretation and broader application possibilities. Pixel-level fusion methods can be sorted into four main groups:

- 1) Component substitution (CS)
- 2) Multi-scale transform (MST)
- 3) Model optimization (MO)
- 4) Hybrid methods [7]

The working principle of CS methods is to transform the multispectral image into a different domain to separate their spatial and spectral information. Then the spatial components are replaced with SAR images and an inverse transformation is applied to restore the images to their original domain by completing the image fusion process [8]. The CS methods inject more spatial details into fused image however spectral distortion may occur [9]. CS methods are the most widely utilized techniques which encompasses, Generalized Intensity-Hue-Saturation Adaptive Algorithm (GIHSA) [4], Gram-Schmidt Context Adaptive Sharpening (GSA-CA) [10], Principal Component Analysis (PCA) [11], Brovey Transform (BT) [3, 6, 7]. In previous studies, these approaches have been vastly employed for the panchromatic and multispectral image fusion. Further advancements in satellite technology have facilitated the availability of SAR and MS imagery for research purpose, offering enhanced information content. SAR and multispectral images usually have huge illumination difference because of different capturing mechanism. Therefore, applying CS methods for SAR and multispectral image pairs is challenging as these methods are good at retaining global details than the local details [7].

The MST methods typically decompose SAR and multispectral images into low and high frequency components and designs fusion rules specifically tailored to align with their respective spectral characteristics [12]. Multiscale geometric analysis, wavelet transform (WT) [13] and dual-tree complex wavelet transform (DTCWT) [14] are the mostly used traditional methods in MST. Multiscale geometric analysis involves curvelet transform (CVT) [15], non-subsampled contourlet transform (NSCT) [16], Bandlet transform [16], contourlet transform (CT) [17], Shearlet transform (ST) [18], and Gaussian and Laplacian pyramid decomposition-based methods. Pyramidal transformation and wavelet transformation-based methods are prominent categories within this group of methodologies. Pyramidal techniques decompose source images into Gaussian and Laplacian pyramids. While MST methods exhibit superior performance relative to CS methods by offering enhanced edge information and preserving cues within images, they are nevertheless subject to certain limitations. These include restricted directional information, the presence of blocking artifacts, and a moderate signal-to-noise ratio [19].

The MO method has been adapted for multispectral and SAR image fusion, treating the process as image restoration. This approach involves constructing a relationship model between the input images, optimizing the energy function and generating the fused image by refining the model [19]. MO methods are mainly composed of variational optimization models and sparse representation models, both of which demand prior knowledge, intricate model development, and significant computational resources. In contrast, deep learning techniques have gained considerable attention in recent years.

Deep learning-based fusion strategies can be classified according to their underlying architecture into autoencoder (AE), convolutional neural network (CNN), and generative

adversarial network (GAN) approaches [20, 21]. AE methods typically involve pre-training an autoencoder for feature extraction and image reconstruction, with intermediate feature fusion following traditional rules. A notable example is DenseFuse [22], which trains its encoder and decoder on the MS COCO dataset and employs addition and L1-norm fusion strategies. CNN methods incorporate convolutional neural networks into image fusion in two distinct ways. Prabhakar et al. [23] utilize carefully designed loss functions and network structures to perform feature extraction, fusion, and image reconstruction in an end-to-end manner. Zhang et al. [24] proposed a proportional maintenance loss of gradient and intensity to guide direct fused image generation. The trained CNN-based approaches can formulate the fusion rules while relying on traditional methods for image reconstruction [20, 21, 23-25]. Lian et al. [26] utilized CNN to generate fusion weights, with image decomposition and reconstruction handled by Laplacian pyramids. GAN methods leverage the adversarial interplay between the generator unit and discriminator unit to estimate target probability distributions, implicitly accomplishing feature extraction, fusion, and image reconstruction. FusionGAN pioneered GAN-based image fusion, establishing a relation between the fused and visible images to enhance texture details in the resultant fused image [27-29]. Owing to the considerable differences among image fusion tasks, these techniques are applied in distinct ways across various fusion contexts. In summary, the review highlights important challenges associated with state-of-the-art methods:

- (a) Spectral distortion continues to be an issue due to the fusion of heterogeneous images.
- (b) Handcrafted fusion rules lack adaptability.
- (c) The interpretability of deep learning fusion remains limited.

The lacunas in state-of-art fusion models can be minimised by the hybrid method implementation. This research work is synthesized with fusion and reconstruction of detail, approximate and contour layers of input images. The key contributions of the implementation are:

- (a) Computed contour layer using Rolling Guidance Filter (RGF) based image decomposition.
- (b) Effective fusion rule designed for individual layer fusion.
- (c) Dynamic weight computation using augmented dataset training of CNN for SAR and MS image fusion.
- (d) Integration of enhanced spatial and spectral information using multiscale decomposed and CNN based fusion.

The research work is presented as follows: Section 2 introduces the proposed methodology, offering a succinct comparison with state-of-the-art techniques. The following Section 3 elaborates on details of the experimental design, data and the evaluation indexes used. Section 4 discusses the results and provides an analysis. The final section concludes with a synthesis of the principal findings.

2. PROPOSED METHODOLOGY

The proposed methodology is integration of CS, MST methods and CNN based method to effectively fuse the SAR and MS image. The experiment mainly includes the six steps:

- (1). IHS transform of MS image.

- (2). Multiscale decomposition of I component using RGF.
- (3). Multilayer fusion.
- (4). CNN based image fusion of original images.
- (5). Combining multilayer fused image and CNN fused image.
- (6). IHS inverse transform to find enhanced color image.

The detailed flow of working is as shown in Figure 1. Stage 1 is to transform the MS image to Ims by using IHS transformation to acquire intensity (I), Hue (H), and saturation (S) components. Then the multiscale decomposition of I component and grayscale SAR image is carried out. In stage 2, rolling guidance filter based multiscale decomposition is accomplished. Stage 3 executes multilayer fusion. CNN based image fusion is carried out in stage 4. The new intensity I' is obtained by combining stage 3 and 4 output. The final enhanced color image is achieved by computing inverse IHS of I' , H, and S components.

Stage 1: IHS transform

The Intensity-Hue-Saturation (HIS) method development was based on the premise that spectral details are primarily found in the hue and saturation elements, while the intensity component retains the spatial details [30]. This research work focuses to improve the quality of spatial information of multispectral image by the fusion of I component of multispectral image and SAR image and IHS inverse transform to reconstruct the color fused image. The IHS transformation is carried out using Eq. (1).

$$\begin{bmatrix} I \\ H \\ S \end{bmatrix} = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ -\sqrt{2}/6 & -\sqrt{2}/6 & 2\sqrt{2}/6 \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1)$$

where, R , G , and B denotes red, green, and blue bands of original input image, respectively. The inverse transform of IHS is expressed in Eq. (2).

$$\begin{bmatrix} R' \\ G' \\ B' \end{bmatrix} = \begin{bmatrix} 1 & -1/\sqrt{2} & 1/\sqrt{2} \\ 1 & -1/\sqrt{2} & -1/\sqrt{2} \\ 1 & \sqrt{2} & 0 \end{bmatrix} \begin{bmatrix} I' \\ H \\ S \end{bmatrix} \quad (2)$$

where, R' , G' , and B' represent values of R , G , and B band values obtained after inverse IHS transform. I' represents modified value of I after fusion.

The traditional approach like Laplace pyramid (LP) multiscale transform fusion [31-33], contrast pyramid transform method [34] were proposed, these methods are less prone to produce artifacts or any kind of distortions due to less consideration for the spatial consistency [6, 35]. Afterwards, transform domain methods incorporated, which have some limitations like loss of detail, when implemented using wavelet transforms (WT), dual-tree complex WT (DTCWT) [36, 37]. In the distant past, Sparse Representation (SR) was built from compressed image sensing [38, 39]. Zhang et al. [40] implemented image fusion using SR as transform domain method for the first time. Afterwards SR method attracted wide attention to fuse the remote sensing (RS) images, owing to their enhanced capability to accurately capture and represent key features and structural details. Nonetheless, the sparse coding SR-based image fusion often entails a substantial computational load, with processing times

escalating significantly as source image dimensions increase. Furthermore, the application of sliding window technique in sparse representation can result in a smoothing effect and a loss of detail, particularly when the overlap between adjacent patches is considerable. To minimize the artifacts that may occur due to traditional methods, this research modifies the I component as explained in subsequent sections.

Stage 2: Multiscale decomposition

In this stage multiscale decomposition is carried out using Rolling Guidance Filter (RGF), which effectively overcomes information redundancy and distortions [6]. Yang and Li [39] demonstrated that RGF, is an edge-preserving smoothing filter. The Joint Bilateral Filter (JBF) is incorporated to achieve iterative operation in RGF. JBF is edge-preserving filter (EPF) [6, 7, 36, 41-46].

The pixel spatial domain kernel determines weights by evaluating both spatial proximity and color similarity, unlike the Gaussian blur which considers only spatial distance and fails to preserve edge details. The RGF encompasses two major procedures: 1) small structure removal and 2) edge recovery. The initial step involves using a Gaussian filter to smooth out minor image details and repeatedly restore the image edges, as depicted in Figure 2. Assume that F_{in} and F_{out} is input and output image pixels respectively. i and j are coordinates of an image. σ_r implies the parameter controlling the scale of Gaussian structure. Following Eq. (3) expresses the Gaussian filtering of image F_{in} at central pixel i .

$$F_{out} = \frac{\sum_{j \in N(i)} \exp\left(-\frac{\|i-j\|^2}{2\sigma_r^2}\right) \cdot F_{in}}{\sum_{j \in N(i)} \exp\left(-\frac{\|i-j\|^2}{2\sigma_r^2}\right)} \quad (3)$$

The edge restoration of the Gaussian-filtered (GF) blurred image is achieved through the application of JBF. The image G , obtained via Gaussian filtering, serves as the initial guide graph J^1 for the JBF process. An iterative procedure is employed, denoted as n iterations, to recover edge information across various scales. In each iteration, the guide graph is derived from the output image J' of the preceding iteration. The expression for this procedure is where J^{n+1} represents the outcome of the n^{th} iteration. The parameter σ_s governs the Gaussian distribution. Eq. (4) gives the relation of F_{in} and J^{n+1} .

$$J^{n+1} = \frac{\sum_{j \in N(i)} e^{\left(-\frac{\|i-j\|^2}{2\sigma_s^2} - \frac{\|J_i^2 - J_j^2\|}{2\sigma_r^2}\right)} \cdot F_{in}}{\sum_{j \in N(i)} e^{\left(-\frac{\|i-j\|^2}{2\sigma_s^2} - \frac{\|J_i^2 - J_j^2\|}{2\sigma_r^2}\right)}} \quad (4)$$

The edge recovered image is updated iteratively as shown in Figure 2. I_m is considered as input image as shown in Eqs. (5)-(7).

$$J^1 = I_m \quad (5)$$

$$J^{n+1} = JBF(I, J, \sigma_s, \sigma_r) \quad (6)$$

$$\begin{cases} S_{d,i} = I_m - J^{i-1} \\ S_{c,i} = J^m - J^{(m+1)} \text{ for } m = (N/2) \\ S_a = J^n \end{cases} \quad (7)$$

where, N is number of iterations, $S_{d,i}$ is detail layer information, $S_{c,i}$ is contour layer information and S_a is approximation layer information at n^{th} level of

decomposition. These obtained layers are further fused by using different fusion rules as explained in next point.

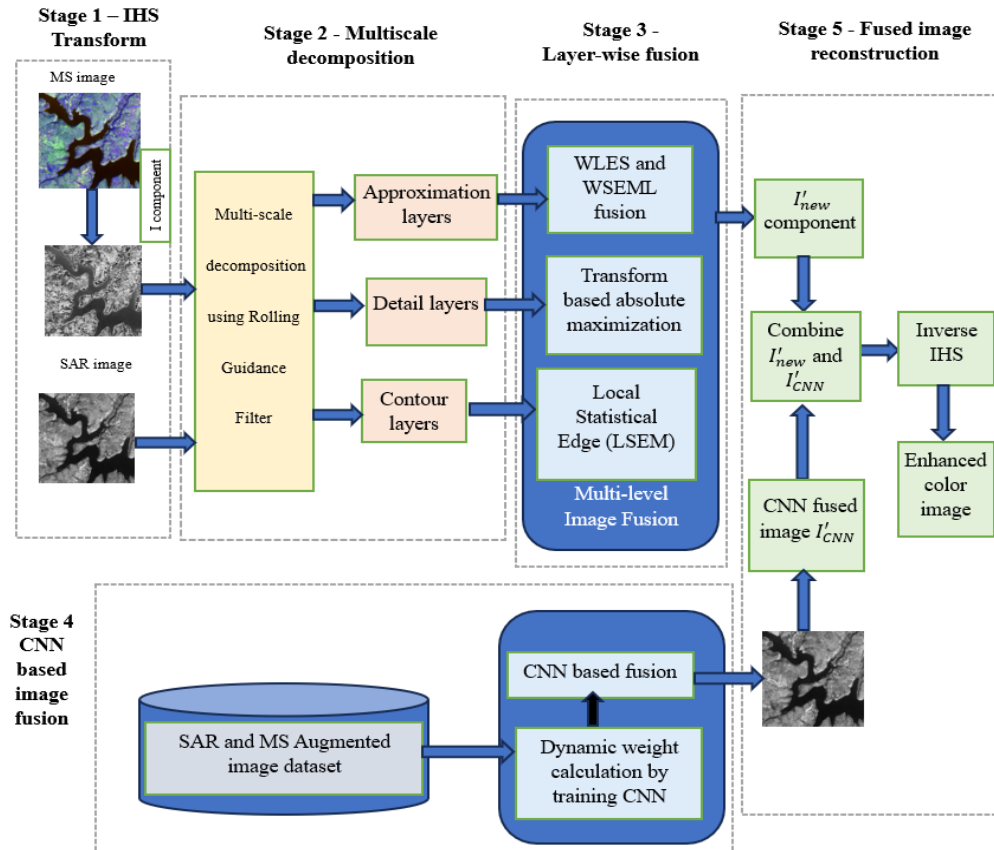


Figure 1. Proposed work flow diagram

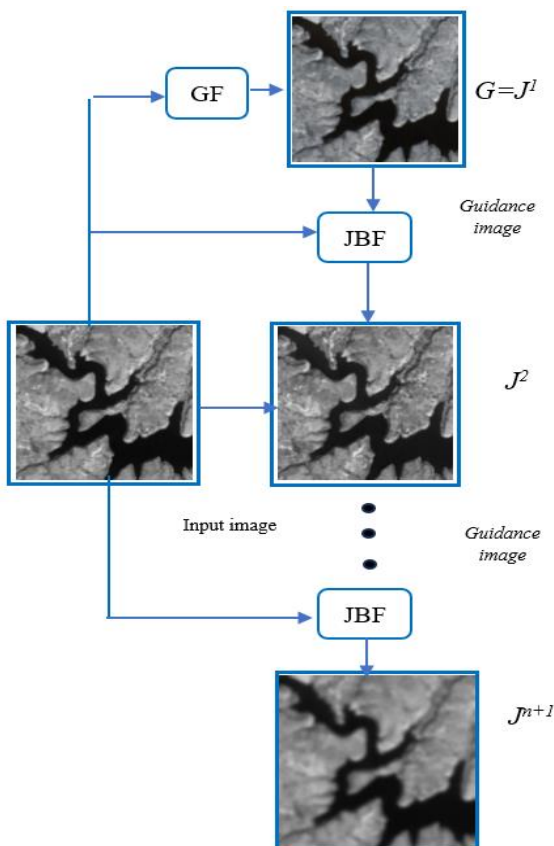


Figure 2. Joint bilateral filter

Stage 3: Layer-wise fusion

Wavelet based traditional methods usually decompose the image into low and high frequency components and design the fusion rule for different frequency components. Fusion rules are designed according to the information contained in that layer and the fused image is reconstructed by inversion process. With this process it is observed that there is loss of information. To avoid the loss, this research work, fuses the scale spaced image layers with different fusion rules, is explained in this section.

a. Approximation layer fusion

The approximation layer contains coarse scale structure information and overall appearance of the image. Duan et al. [42] employed a multiscale decomposition of images utilizing a weighted least squares framework. For the base image, the average value of N base images is considered as the fused image. Lewis et al. [36], Liu et al. [46], and Gong et al. [47] proposed a hybrid approach combining MST and SR, where low-pass bands are fused using a conventional averaging method [48, 49]. Jian et al. [6] utilized a weighted average and global variance. The averaging-based method has limitations, such as contrast loss when input images have different intensity distributions, and noisy images may significantly affect the image reconstruction process.

The most frequently used fusion methods for approximation layer are like averaging rule which may lose residual details and structural information after fusion because they are created for ideal approximation layer which is difficult to

achieve in satellite images. Therefore, Weighted Local energy Sum (WLES) method and Weighted Sum of Entropy and Mean of Laplacian (WSEML) method are employed to extract details and edge information. The WLE is expressed as:

$$WLES_f(p, q) = \sum_{m=-r}^r \sum_{n=-r}^r w \times ((m+r+1), (n+r+1)) \times S_a((m+p), (n+q))^2 \quad (8)$$

where, $f \in \{SAR, MS\}$ and $S_a(p, q)$ symbolise the sub-band of low frequency at position (p, q) . Weight matrix w is of $(2r+1) \times (2r+1)$, where each element value of w is set to 2^{2r-d} , r is the radius of w matrix, and four-neighborhood distance from the corresponding element to the center of matrix is represented with d , when r is set to 1, the normalization matrix W can be expressed as:

$$\frac{1}{16} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix} \quad (9)$$

The Eight-Neighborhood Modified Laplacian (EML) considers the impact of diagonal coefficients and effectively leverages the information from neighborhood [47, 48]. WSEML is a weighted representation of EML. The mathematical expression is:

$$WSEML_f(p, q) = \sum_{m=-r}^r \sum_{n=-r}^r w \times (m+r+1, n+r+1) \times EML_f(m+p, n+q) \quad (10)$$

$$\begin{aligned} EML_f(p, q) = & |2S_a(p, q) - S_a(p-1, q) \\ & - S_a(p+1, q)| \\ & + |2S_a(p, q) \\ & - S_a(p, q-1) - S_a(p, q+1)| \\ & + \frac{1}{\sqrt{2}} |2S_a(p, q) - S_a(p-1, q-1) \\ & - S_a(p+1, q+1)| \\ & + \frac{1}{\sqrt{2}} |2S_a(p, q) - S_a(p-1, q+1) \\ & - S_a(p+1, q-1)| \end{aligned} \quad (11)$$

Then, according to WLE and WSEML, the fused low-frequency sub-band coefficient $A_{fused}(p, q)$ is obtained, and the mathematical expression is

$$A_{fused}(p, q) = \begin{cases} A_{SAR}(p, q) & \text{if condition (1) is true} \\ A_{MS}(p, q) & \text{otherwise} \end{cases} \quad (12)$$

The condition (1) is $WLE_{SAR}(p, q) \cdot WSEML_{SAR}(p, q) \geq WLE_{SAR}(p, q) WSEML_{SAR}(p, q)$.

While the WLE method emphasizes sharpness by focusing on frequency domain features, the WSEML approach primarily considers a weight matrix based on horizontal, vertical, and diagonal distances, making it effective for observing spatial details. Consequently, the combination of WLE and WSEML can identify rich information block in the image.

b. Contour layer

Many of the researchers have decomposed the input images into low and high frequency components whereas those have decomposed it to contour layer have mostly used absolute maxima rule. Gong et al. [7] have utilized Coupled Neural P (CNP) for contour layer fusion, which is quite parameter sensitive, complex implementation with high computational complexity. Gong et al. [47] have fused the contour layer using absolute maximum principle, with the possibility of contrast loss. Hence in order to preserve textural and edge information with a good blend from both images, the strategy of Local Statistical Edge Model (LSEM) is employed, it focuses on preserving edge structures by analysing local statistical properties of the decomposed images. The local statistical properties taken into consideration are, gradient magnitude and local statistics like standard deviation. The edge strength is computed using following mathematical relation in Eq. (13).

$$E(t) = \alpha \cdot G(t) + \beta \cdot Std(t) \quad (13)$$

Here, $E(t)$ is edge strength, $G(t)$ is gradient, $Std(t)$ is standard deviation, α and β are tuneable weights.

The fusion of gradient and variance provides a more adaptive representation of edge and texture information. Utilizing the edge strength to calculate the adaptive weights for fusion, using Eqs. (14) and (15).

$$\omega_1 = \frac{E_1}{E_1 + E_2 + \forall} \quad (14)$$

and

$$\omega_1 = 1 - \omega_2 \quad (15)$$

Here, ω_1 and ω_2 are the weight calculations using E_1 and E_2 edge strengths of both input detail layers. $\forall$ is constant to make the denominator non zero value. The design of local statistical edge model is explained in Eq. (16).

$$D_F^{c,i}(u, v) = \omega_1(u, v) \cdot D_{SAR}^{c,i}(u, v) + \omega_2(u, v) \cdot D_{MS}^{c,i}(u, v) \quad (16)$$

where, $D_{SAR}^{c,i}$, $D_{MS}^{c,i}$ are contour layers of SAR and MS image respectively.

This approach combines image gradient and standard deviation as complementary focus metrics, allowing for the concurrent assessment of local edge strength and overall intensity variation. The gradient aspect captures intricate structural and directional details, while the standard deviation measures the overall contrast and texture richness throughout the area. By merging these metrics, the method successfully maintains both sharp local features and broad intensity variations, leading to better information retention compared to methods that overlook either local or global pixel variation.

c. Detail layer fusion

These are considered as high frequency components of image. These contains abundant textural and edge information. Jian et al. [6] have employed JBF based detail layer fusion to preserve edge information. As detail layer consists of high frequency information, absolute maximization is suitable method [47, 50, 51]. Therefore, to fuse this layer absolute maximization using Gaussian filter is employed. The mathematical representation is as follows:

$$D_F^{d,i} = \text{Gaussian}(W_{d,\sigma_0}) \cdot D_{MS}^{d,n} + (1 - \text{Gaussian}(W_{d,\sigma_0})) \cdot D_{SAR}^{d,n} \quad (17)$$

Here, $W_j = \begin{cases} 1 & \text{if } |D_{MS}^{d,n}| \cdot |D_{MS}^{d,n}| > |D_{SAR}^{d,n}| \cdot |D_{SAR}^{d,n}|, n = 2, \\ 0 & \text{otherwise} \end{cases}$
3, 4, ..., N.

The computation is carried out using kernel size = 5. The last fusion image is reconstructed by using Eq. (18).

$$I'_{new} = \sum_{i=1}^n (D_F^{c,i} + D_F^{d,i}) + A_{fused} \quad (18)$$

Using Eq. (18), the new I component is calculated. For the further processing the recovered image is combined with CNN fused image, as explained in next point.

Stage 4: CNN based image fusion

As depicted in the workflow block diagram Figure 1, involves the preparation of an augmented dataset. The augmented dataset is constructed by generating 100 samples from each original input image. Each input image is processed through three convolutional layers, with a progressive increase in the number of feature maps to extract modality-specific features. The kernel size for the convolution operation is set at 3×3 . The ReLU activation is used to introduce non-linearity. Padding set to same to ensure preservation of spatial resolution. These feature maps are concatenated along the channel axis to enable joint feature encoding. The concatenated features are passed through additional convolutional layers, enhancing network's ability to learn spatial correlations and rich hierarchical representations. Further the tensor is flattened and a fully connected layer is used to refine the feature representation. Subsequently, the sigmoid activation is used to convert it into weighting map. An epoch count of 100 is employed to achieve effective results, SSIM loss function used, and learning rate = 0.0001. The detailed layer specifications of proposed CNN based fusion network is explained in Table 1.

Table 1. Detailed layer specifications

Layers (Type)	Output Shape	Parameters
input_layer (InputLayer)	(None, 224, 224, 1)	0
input_layer_1 (InputLayer)	(None, 224, 224, 1)	0
conv2d (Conv2D)	(None, 224, 224, 16)	160
conv2d_1 (Conv2D)	(None, 224, 224, 16)	160
conv3 (Conv2D)	(None, 224, 224, 32)	4,640
conv3 (Conv2D)	(None, 224, 224, 32)	4,640
conv2d (Conv2D)	(None, 224, 224, 64)	18,496
conv2d (Conv2D)	(None, 224, 224, 64)	18,496
conv2d (Conv2D)	(None, 1)	2,049
Concatenate (Concatenate)	(None, 224, 224, 128)	0
conv2d_2 (Conv2D)	(None, 224, 224, 32)	36,896
conv2d_3 (Conv2D)	(None, 224, 224, 64)	18,496
flatten_1 (Flatten)	(None, 3211264)	0
dense (Dense)	(None, 128)	411,041,920
dense_1 (Dense)	(None, 50176)	6,472,704
reshape (Reshape)	(None, 224, 224, 1)	0
subtract (Subtract)	(None, 224, 224, 1)	0
multiply (Multiply)	(None, 224, 224, 1)	0
Add (Add)	(None, 224, 224, 1)	0

Stage 5: Fused image reconstruction

The modified I'_{new} component from layer-wise fusion is combined with CNN fused output image I'_{CNN} . As CNN-based methods primarily emphasize spatial features, they are capable of producing images with enhanced spatial detail. The multiscale decomposition is responsible for preserving the spectral information. Thus, to acquire information that is both spectrally and spatially rich, the integration of the two fused outputs is carried out. Finally, a color image is reconstructed using the H & S components along with the updated I' component, as described in Eq. (2).

3. EXPERIMENTAL DESIGN

A. Experimental data

In the course of rigorous experimentation, the MS image dataset employed was procured from the National Remote Sensing Centre (NRSC) in Hyderabad, India. Conversely, the SAR image dataset was sourced from the Earth Resources Observation and Science (EROS) Centre of the U.S. Geological Survey via the freely accessible QGIS tool 3.18 version. This dataset encompasses the regions within the state of Maharashtra, India. The SAR image possesses a resolution of 15 meter, whereas the LISS III Multi Spectral image exhibits a resolution of 8 meter. The dataset used consists of more than 100 MS and SAR image pairs, in this presentation results of pairs is discussed. Prior to initiating the fusion process, an initial pairwise registration is conducted to achieve optimal results. The experimental setup used is CPU Intel Core i7-12500 H_z 12th gen 2.50 GHZ, 2GB GPU, 64 bit operating system Windows 11, Programming environment PyCharm 2024.3.

B. Evaluation indexes

Qualitative evaluation entails a subjective analysis of outcomes by comparing texture details, color information, spatial structure, visual effects, and other features of the combined images. On the other hand, quantitative evaluation offers an objective analysis based on specific evaluation metrics.

In this section, we report the experimental process for diverse satellite image dataset using proposed algorithm. The experiments on four SAR and MS image datasets to validate the proposed fusion algorithm. For the purpose of benchmarking, we chosen the listed baseline methods in the experimentations as, NSCT, PCA, BT, DWT [50], FusionGAN [29], and Siamese Network [51]. The experiments compare fused image with the reference image to analyse the performance of proposed method. Here, we utilize Peak Signal to Noise Ratio (PSNR), Spectral Angle Mapper (SAM), Erreur Relative Globale Adimensionnelle de Synthèse (ERGAS), Average Gradient (AG), and Information Entropy five measures for proposed method evaluation. PSNR is a measure of accuracy of an algorithm. SAM signifies the spectral distortion after fusion. ERGAS indicates radiometric and spatial quality of an image.

(a) PSNR

This is crucial metric for measuring the quality of an image after processing. It evaluates the ratio of the maximum achievable power of an input signal to power of the noise, defined using Eq. (19).

$$PSNR = 10 \times \log_{10} \frac{255}{MSE} \quad (19)$$

Here, MSE denotes mean square error of fused image as explained in Eq. (20). Let's consider $I_f(x'_k, y'_k)$ denote the transformed coordinates of $I_a(x_k, y_k)$ from original image.

$$MSE = \frac{1}{N} \sum_{k=1}^N (I_a(x_k, y_k) - I_f(x'_k, y'_k))^2 \quad (20)$$

Here, N represents the number of difference pairs. I_a, I_f indicates input and fused image respectively.

(b) SAM

The spectral quality is evaluated using SAM, it typically compares pixel-wise spectral similarity of fused image with reference image by computing angle between two vectors. Eq. (21) evaluates spectral similarity between K and T , fused and reference image spectral pixel vectors respectively.

$$SAM(K, T) = \cos^{-1} \frac{(K \times T)}{\|K\| \|T\|} \quad (21)$$

(c) ERGAS

The ERGAS is used for global relative error computation, as shown in Eqs. (22) and (23) [7].

$$ERGAS = 100 \frac{a}{b} \sqrt{\frac{1}{P} \sum_{p=1}^P \left(\frac{RMSE(I_a, I_f)}{\mu I_a} \right)^2} \quad (22)$$

$$\frac{1}{NM} \sqrt{\sum_{i=1}^N \sum_{j=1}^M (I_a(x_k, y_k) - I_f(x'_k, y'_k))^2} \quad (23)$$

where, a/b is ratio of resolution of SAR and MS image, P denotes the number of bands and μ is mean value of the image.

(d) AG

This signifies the average magnitude of the image gradient, computed using discrete derivatives as follows

$$AG = \frac{1}{(M-1)(N-1)} \sum_{i=1}^{M-1} \sum_{j=1}^{N-1} \sqrt{\left(\frac{\partial f}{\partial x} \right)^2 + \left(\frac{\partial f}{\partial y} \right)^2} \quad (24)$$

Here, M, N is image dimensions, $f(i, j)$ is pixel intensity.

$$\begin{aligned} \frac{\partial f}{\partial x} &= f(i+1, j) - f(i, j) \\ \frac{\partial f}{\partial y} &= f(i, j+1) - f(i, j) \end{aligned}$$

This measures the strength of edges and textures, averaged throughout the image [49].

(e) Entropy

This parameter gives idea about the richness of information in the data under consideration, calculated using Eq. (25). The higher the entropy value, better the quality of fused image [4].

$$e = - \sum_{n=1}^{M-1} p(n) \log_2 p(n) \quad (25)$$

where, $p(n)$ is probability of occurrence of n^{th} gray level, M is the dynamic range of the image under analysis.

(f) Standard Deviation of difference image

This calculation determines the Standard Deviation (SD) of residual information, indicating the extent to which the fusion process has altered the input image. The absolute difference between the source and fused image is initially calculated using Eq. (26), and this result is then used to determine the SD_{diff} using Eq. (27).

$$I_{diff} = |I_{source} - I_{fused}| \quad (26)$$

$$SD_{diff} = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (I_{diff}(i) - \mu_{diff})^2} \quad (27)$$

where, I_{diff} is absolute difference image, μ_{diff} is mean of difference pixel image and N is total number of pixels.

4. RESULTS AND DISCUSSION

To highlight the effectiveness of the proposed method, comparison with several traditional image fusion techniques is done. These comparative methods include the nonsubsampling contourlet transform (NSCT), PCA, BT, Discrete Wavelet Transform (DWT), as well as more advanced techniques like FusionGAN and Siamese Network. Tables 2-5 present the results of all comparative methods across four distinct SAR and MS image datasets. Figures 3-9 illustrate the MS input, SAR input, NSCT output, PCA output, BT output, DWT output, FusionGAN output, Siamese network output, our method's grayscale output image, and our method's color output image in subgraphs from Figures 3 to 6 respectively.

Visual analysis: NSCT decomposes the input images and performs energy based weighted fusion, is implemented. NSCT is poor at capturing direction edges which produces visual effect and blurring at the edges and reduced spectral fidelity. Therefore, poorly retained spectral resolution can be seen through Figure 3 for dataset pair 1. PCA considers most important components based on variance, which causes potential information loss and brightness distortion, leads to reduced spatial resolution as well as spectral fidelity. BT does the uniform enhancement without considering local texture or edges which eventually suppress the high frequency information, causes blurriness. DWT based methods decomposes the image and handles the individual frequency band. As SAR and MS image have basic different characteristics which demands variation in basis function and level of decomposition. This causes improper feature extraction from both input images, loss of information during fusion. Figure 3 clearly shows the spectral fidelity and overly smoothed image using FusionGAN method as it fails to acquire important details from dissimilar images makes image not interpretable. Though Siamese network is able to obtain relatively good results, but it performs weakly in integrating dissimilar features from SAR and MS image. Siamese network method does not perform well in terms spectral fidelity. These

observations are graphically represented in Figures 10 to 12. The affected area is zoomed in Figure 13 for the NSCT, PCA, BT, DWT and Siamese network output respectively. Whereas, proposed method is able to recover both spectral and spatial features.

Quantitative analysis: Many a times image may look good but may not have the enough important information for further processing therefore, quantitative evaluation is required. Quantitative analysis mainly relies on the retention of information, the restoration of spatial features, the fidelity of spectral data, and the degree of image interpretability. Tables 2-5 portray performance parameters for all datasets, where most efficient results are highlighted. Figures 7-12 depict the range of performance parameters for all datasets. In terms of spatial information, quality of fused image, PSNR and AG are the indicators. Figures 7 and 8 show the range of PSNR and AG respectively. This implies PCA and Siamese network method performs poorly whereas NSCT and BT method relatively performs well but not up-to the mark for PSNR metric. Proposed method consistently excels in PSNR evaluation. Regarding AG indicator, PCA, BT, DWT and FusionGAN methods show less spatial information retention.

On the contrary traditional NSCT with multidirectional decomposition is able to perform well. Siamese network result indicates slightly improved response because of improved training mechanism than the FusionGAN method. Whereas proposed method maintains the highest spatial retention with largest value of AG as 9.77. From the perspective of spectral information measure, ERGAS and SAM are crucial indicators, signifies the level of spectral distortion in the fused image. From respective result tables of different dataset and Figures 9 and 10, it illustrates that PCA output image shows highest level of distortion which can be verified by visual inspection also. In contrast to the Siamese network method, which holds the second position, the proposed approach achieves superior outcomes, with improvements of 2.22 in ERGAS and 0.54 in SAM, respectively. Successively, entropy signifies the overall enhancement of input image. The proposed method performs best as presented in Figure 11. The standard deviation of the difference image for the proposed method, falls within the healthy range of 12 to 18, indicating minimal deviation in pixel values, depicted in Figure 12. This suggests that the designed method exhibits a clear advantage in preserving the original image information.

Table 2. Quantitative evaluation results for dataset pair P1

Method	PSNR	Entropy	ERGAS	SAM	AG	SD _{diff}
NSCT	25.33	7	5.23	5.54	8.41	115.87
PCA	15.33	6.67	16.56	17.27	8.06	86.32
BT	22.69	7.21	7.09	6	5.71	81.64
DWT	23.89	7.11	6.17	6.19	5.07	46.21
FusionGAN	20.07	6.58	5.98	5.89	3.54	119.65
Siamese	5.14	6.74	53.5	1.5	8.63	25.65
Our	32.15	7.93	2.48	0.54	8.89	17.56

Table 3. Quantitative evaluation results for dataset pair P2

Method	PSNR	Entropy	ERGAS	SAM	AG	SD _{diff}
NSCT	25.04	7	6.88	5.77	8.41	105.27
PCA	25.08	7.04	6.845	5.82	5.06	113.56
BT	20.76	7.08	11.26	4.74	6.18	51.37
DWT	20.5	7.05	11.6	6.48	6.44	95.96
FusionGAN	10.12	6.74	10.14	2.56	5.12	120.45
Siamese	6.84	6.43	55.92	0.35	8.67	28.21
Our	32.84	7.24	2.98	0.78	8.89	15.36

Table 4. Quantitative evaluation results for dataset pair P3

Method	PSNR	Entropy	ERGAS	SAM	AG	SD _{diff}
NSCT	25.34	7.09	5.88	4.93	8.27	72.76
PCA	16.08	6.39	17.11	17.28	7.23	100.08
BT	25.65	7.02	5.68	4.61	6.34	88.71
DWT	22.69	7.06	7.99	6.86	6.38	92.06
FusionGAN	19.07	5.98	8.45	4.98	4.95	94.78
Siamese	6.31	6.8	52.6	1.45	8.68	21.64
Our	34.26	7.27	2.22	0.68	8.76	14.47

Table 5. Quantitative evaluation results for dataset pair P4

Method	PSNR	Entropy	ERGAS	SAM	AG	SD _{diff}
NSCT	24.88	7.25	6.19	5.16	9.87	66.89
PCA	8.53	7.3	40.69	35.54	5.63	77.09
BT	23.85	7.14	6.97	4.43	7.1	53.13
DWT	16.7	6.96	15.89	13.88	5.04	73.61
FusionGAN	17.62	6.98	5.24	3.94	2.38	84.51
Siamese	6.31	6.8	52.6	8.48	0.89	24.14
Our	34	7.55	2.26	0.742	9.77	12.45

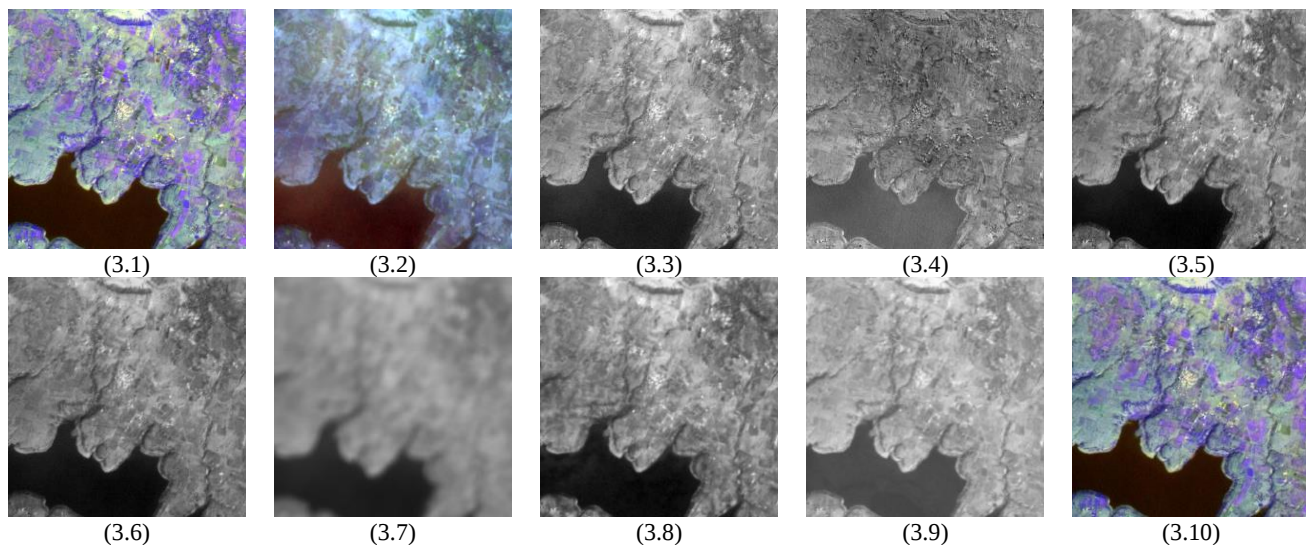


Figure 3. Dataset pair P1 results: (3.1) MS input, (3.2) SAR input, (3.3) NSCT, (3.4) PCA, (3.5) BT, (3.6) DWT, (3.7) FusionGAN, (3.8) Siamese network, (3.9) Our method output, (3.10) Our method color image

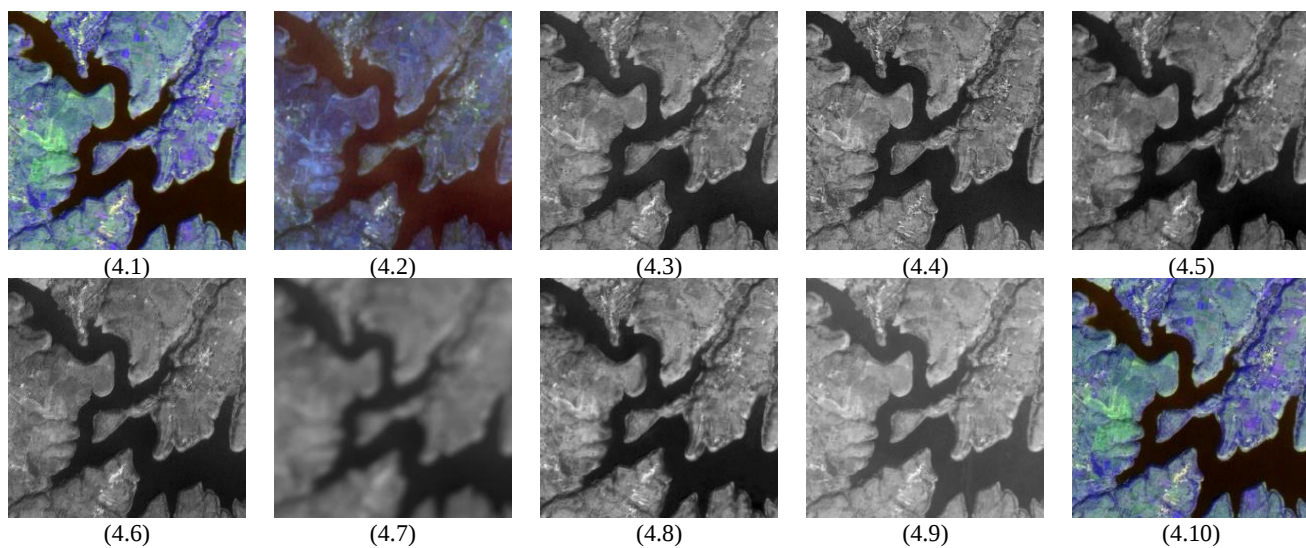


Figure 4. Dataset pair P2 results: (4.1) MS input, (4.2) SAR input, (4.3) NSCT, (4.4) PCA, (4.5) BT, (4.6) DWT, (4.7) FusionGAN, (4.8) Siamese network, (4.9) Our method output, (4.10) Our method color image

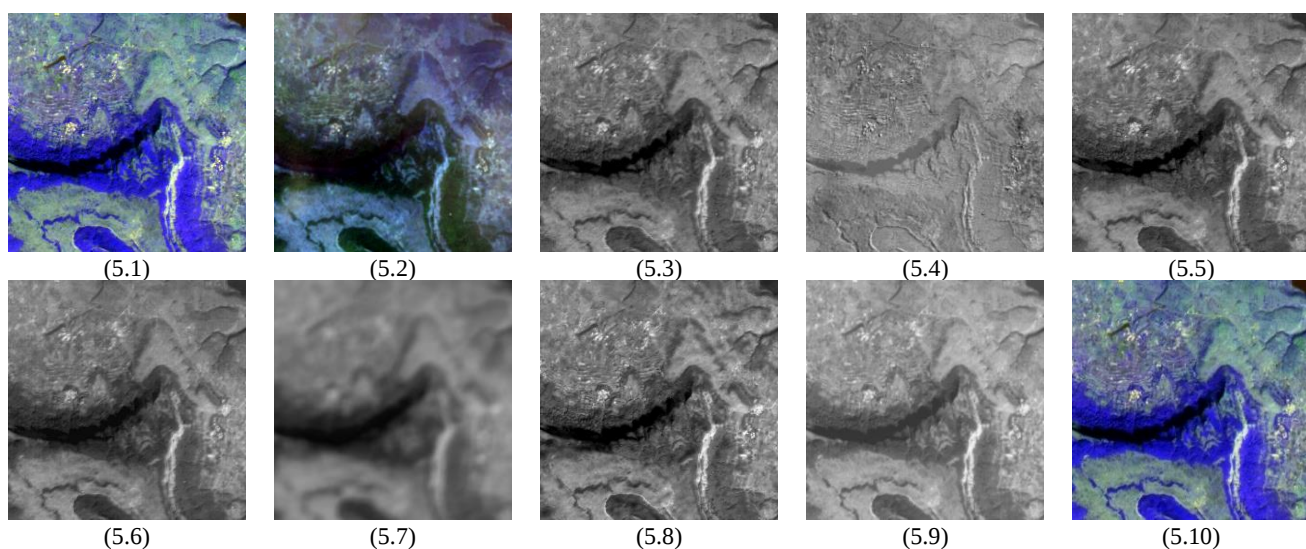


Figure 5. Dataset pair P3 results: (5.1) MS input, (5.2) SAR input, (5.3) NSCT, (5.4) PCA, (5.5) BT, (5.6) DWT, (5.7) FusionGAN, (5.8) Siamese network, (5.9) Our method output, (5.10) Our method color image

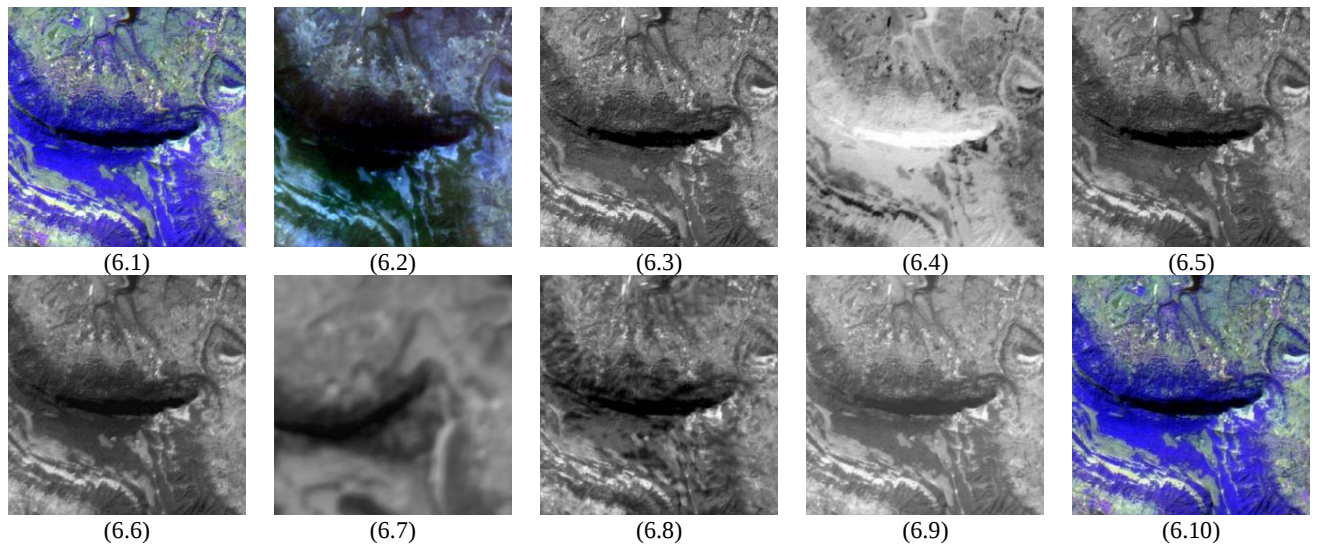


Figure 6. Dataset pair P4 results: (6.1) MS input, (6.2) SAR input, (6.3) NSCT, (6.4) PCA, (6.5) BT, (6.6) DWT, (6.7) FusionGAN, (6.8) Siamese network, (6.9) Our method output, (6.10) Our method color image

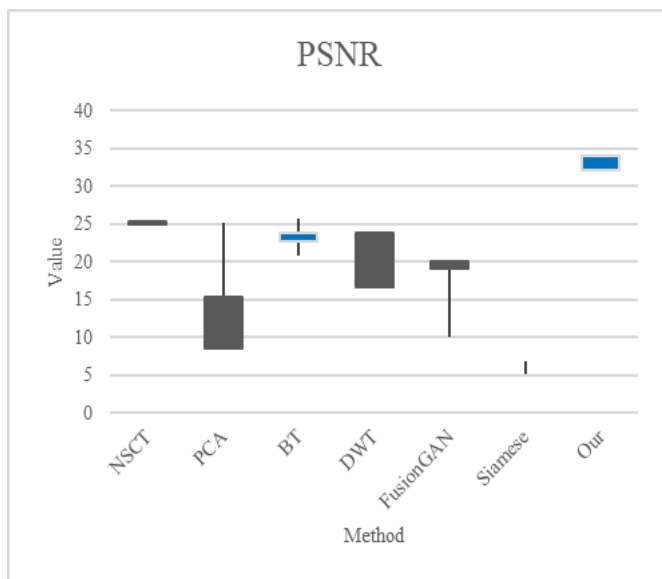


Figure 7. Comparative analysis of PSNR

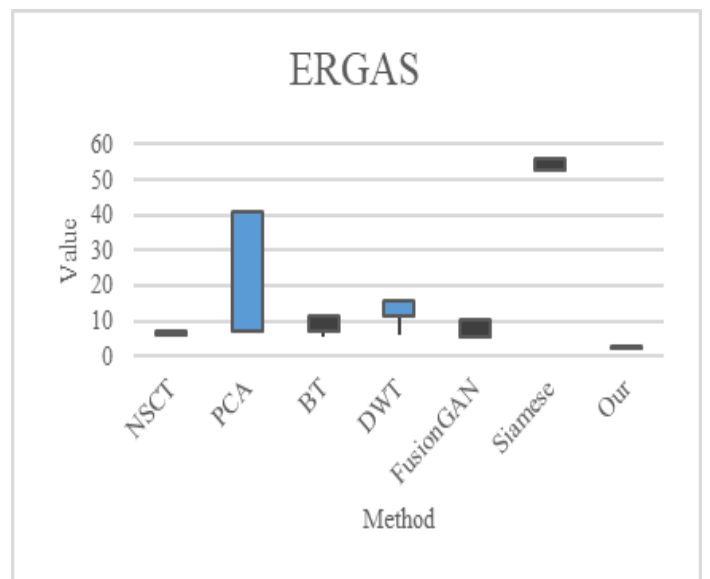


Figure 9. Comparative analysis of ERGAS

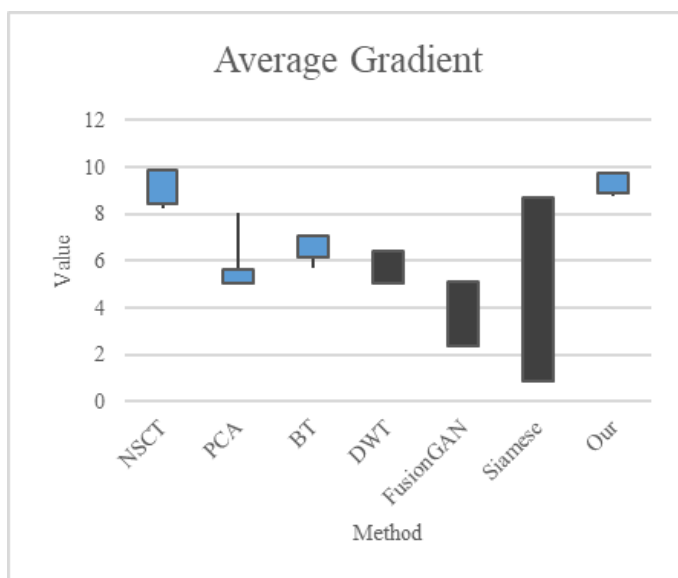


Figure 8. Comparative analysis of AG

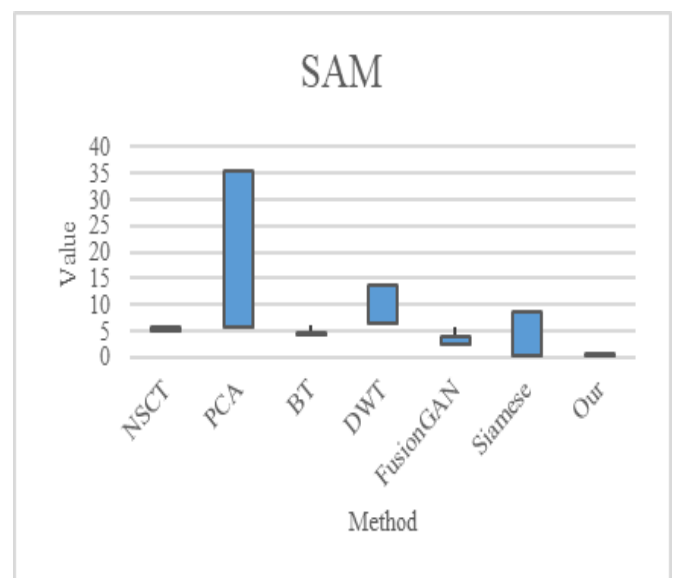


Figure 10. Comparative analysis of SAM

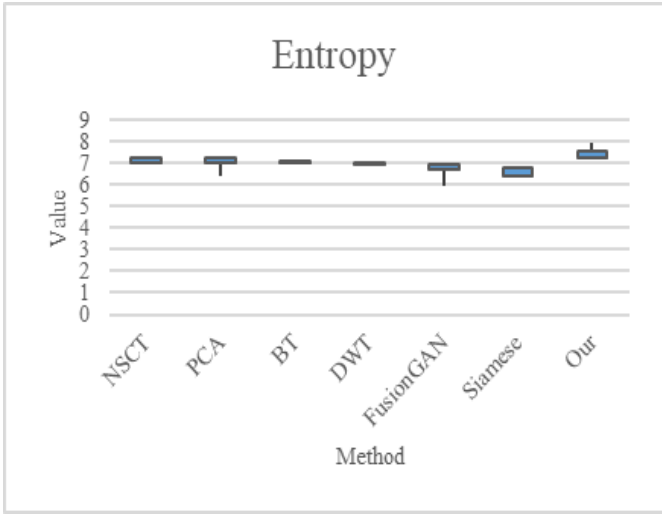


Figure 11. Comparative analysis of entropy

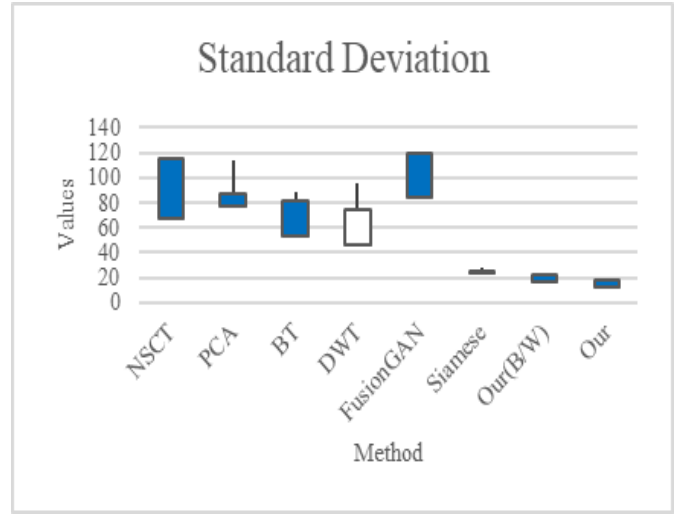


Figure 12. Comparative analysis of standard deviation of difference image

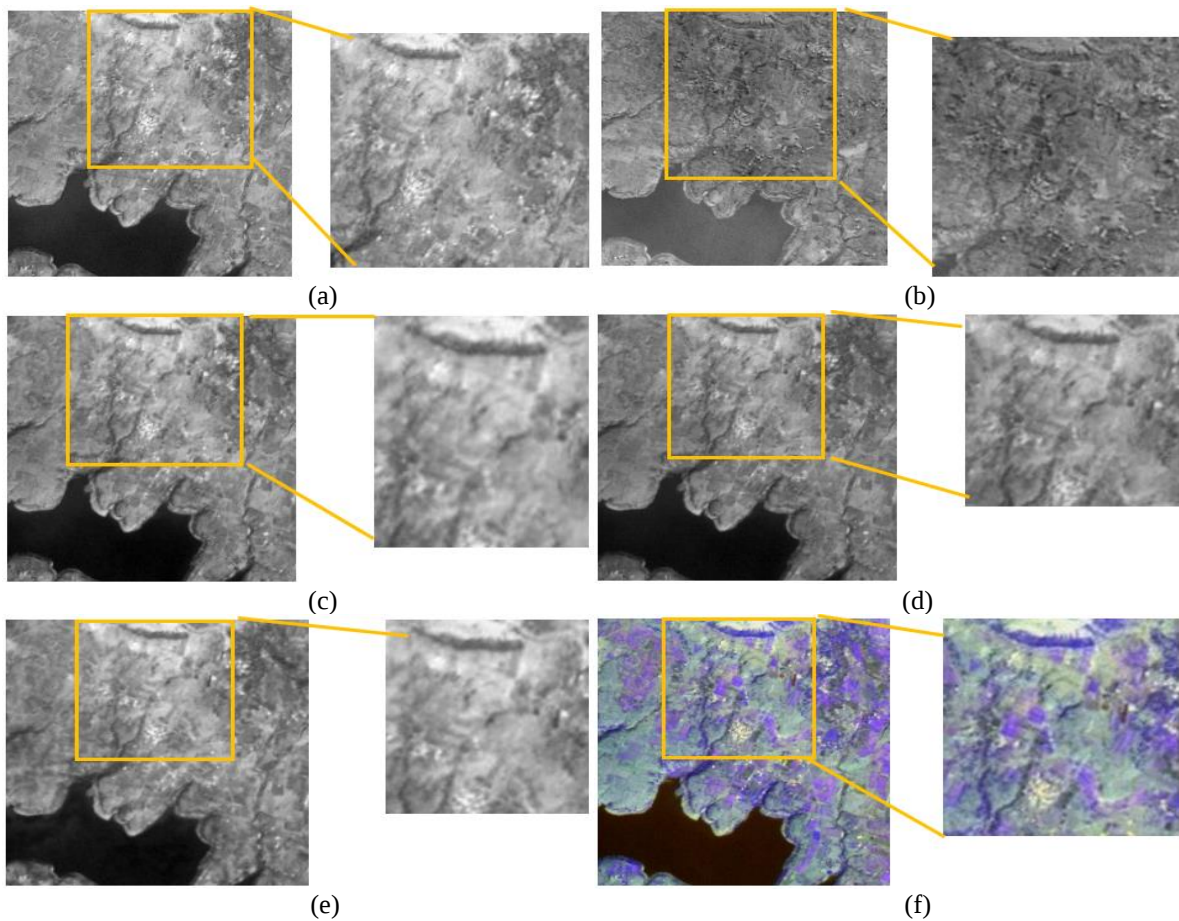


Figure 13. Zoomed sections of (a) NSCT, (b) PCA, (c) BT, (d) DWT, (e) Siamese and (f) our method output images

5. CONCLUSIONS

Despite the higher computational demands of pixel-level fusion techniques compared to feature-level and decision-level fusion methods, they continue to be widely used in remote sensing image fusion due to their superior accuracy. In this article pixel level fusion of SAR and MS image is carried out using RGF multiscale decomposition in combination with CNN based fusion. In the initial phase of fusion, the Intensity (I) component of the multispectral (MS) image is

predominantly considered, as it encapsulates the most essential perceptual information. Modifying the I component facilitates the effective integration of spatial details, particularly from the SAR image, thereby enhancing the spatial richness of the fused image. Processing of the I component allow for more controlled enhancement while preserving spectral integrity, and any distortions introduced are relatively easier to rectify during the inverse transformation process. RGF based multiscale decomposition can reserve edge information better than other decomposition

methods such as wavelet-based decomposition. After decomposition approximation layer, detail layer and contour layers are obtained and fused individually. Approximation layer is fused using WLE and WSEML methods to obtain most of the information from both input images. WLE method gives advantage of analysing frequency band-wise better to keep spectral information intact and WSEML method primarily focuses on improving the entropy and mean of Laplacian is responsible to ensure spectral consistency. The fusion of the contour layer is executed using a Local Statistical Edge Model, which is designed by taking into account both standard deviation and gradient. This approach aids in preserving crucial edge information, particularly in relation to directional data. Whereas, the detail layer fusion utilises absolute maximization to preserve high frequency information. Subsequently, the average of all layers is integrated with the fusion output from the CNN approach. In CNN-based fusion, the network is trained using the SSIM loss function and augmented dataset pairs of SAR and MS images, which enhances the network's robustness to illumination variations and maintains a balance in smoothness that promotes the preservation of fine details. Through this the weights are calculated and image fusion is accomplished which has ability to find and preserve complex features from dissimilar datasets. By integrating multiscale decomposition with CNN-based fusion methods, the resulting image is enriched with both spatial and spectral information. Following this, the modified *I* component is processed through an inverse IHS transformation. This innovative method demonstrates superior performance compared to both traditional and network-based techniques. It achieves a peak PSNR of 34 and a minimum SAM of 0.54, indicating significant improvements in the spatial and spectral resolution of the final image. Also ensures high fidelity preservation by maintaining the standard deviation of different images at a low value, ranging from 12 to 19.

REFERENCES

- [1] Ying, J.C., Shen, H.L., Cao, S.Y. (2022). Unaligned hyperspectral image fusion via registration and interpolation modeling. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 1-14. <https://doi.org/10.1109/TGRS.2021.3081136>
- [2] Wang, P., Huang, M.X., Shi, S.P., Huang, B., Zhou, B.L., Xu, G. (2024). Landsat-8 and Sentinel-2 image fusion based on multiscale smoothing-sharpening filter. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 17: 17957-17970. <https://doi.org/10.1109/JSTARS.2024.3469974>
- [3] Aburaed, N., Alkhatib, M.Q., Marshall, S., Zabalza, J., Al Ahmad, H. (2023). A review of spatial enhancement of hyperspectral remote sensing imaging techniques. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 16: 2275-2300. <https://doi.org/10.1109/JSTARS.2023.3242048>
- [4] Kulkarni, S.C., Rege, P.P. (2020). Pixel level fusion techniques for SAR and optical images: A review. *Information Fusion*, 59: 13-29. <https://doi.org/10.1016/j.inffus.2020.01.003>
- [5] Wang, X., Dong, S.W., Song, H.J., Sun, B.Q., Wu, W.J., Wang, W.X., Guo, D., Gao, Z. (2024). Time Transfer Link fusion algorithm based on wavelet multi-resolution analysis. *Measurement*, 232: 114599. <https://doi.org/10.1016/j.measurement.2024.114599>
- [6] Jian, L.H., Yang, X.M., Zhou, Z.L., Zhou, K., Liu, K. (2018). Multi-scale image fusion through rolling guidance filter. *Future Generation Computer Systems*, 83: 310-325. <https://doi.org/10.1016/j.future.2018.01.039>
- [7] Gong, X.J., Hou, Z.Y., Wan, Y.T., Zhong, Y.F., Zhang, M., Lv, K.Y. (2024). Multispectral and SAR image fusion for multiscale decomposition based on least squares optimization rolling guidance filtering. *IEEE Transactions on Geoscience and Remote Sensing*, 62: 1-20. <https://doi.org/10.1109/TGRS.2024.3353868>
- [8] Huang, D.S., Tang, Y.L., Wang, Q.S. (2022). An image fusion method of SAR and multispectral images based on non-subsampled Shearlet transform and activity measure. *Sensors*, 22(18): 7055. <https://doi.org/10.3390/s22187055>
- [9] Wang, H.X., Jiang, W.S., Lei, C.Q., Qin, S.L., Wang, J.L. (2014). A robust image fusion method based on local spectral and spatial correlation. *IEEE Geoscience and Remote Sensing Letters*, 11(2): 454-458. <https://doi.org/10.1109/LGRS.2013.2265915>
- [10] Gao, G., Wang, M.X., Zhang, X., Li, G.S. (2025). DEN: A new method for SAR and optical image fusion and intelligent classification. *IEEE Transactions on Geoscience and Remote Sensing*, 63: 1-18. <https://doi.org/10.1109/TGRS.2024.3500036>
- [11] Shahdoosti, H.R., Ghassemian, H. (2016). Combining the spectral PCA and spatial PCA fusion methods by an optimal filter. *Information Fusion*, 27: 150-160. <https://doi.org/10.1016/j.inffus.2015.06.006>
- [12] Hill, P., Al-Mualla, M.E., Bull, D. (2017). Perceptual image fusion using wavelets. *IEEE Transactions on Image Processing*, 26(3): 1076-1088. <https://doi.org/10.1109/TIP.2016.2633863>
- [13] Arif, M., Wang, G.J. (2020). Fast curvelet transform through genetic algorithm for multimodal medical image fusion. *Soft Computing*, 24: 1815-1836. <https://doi.org/10.1007/s00500-019-04011-5>
- [14] Iqbal, M.Z., Ghafoor, A., Siddiqui, A.M. (2013). Satellite image resolution enhancement using dual-tree complex wavelet transform and nonlocal means. *IEEE Geoscience and Remote Sensing Letters*, 10(3): 451-455. <https://doi.org/10.1109/LGRS.2012.2208616>
- [15] Amolins, K., Zhang, Y., Dare, P. (2007). Wavelet based image fusion techniques — An introduction, review and comparison. *ISPRS Journal of Photogrammetry and Remote Sensing*, 62(4): 249-263. <https://doi.org/10.1016/j.isprsjprs.2007.05.009>
- [16] Miao, Q.G., Lou, J.J., Xu, P.F. (2012). Image fusion based on NSCT and bandelet transform. In 2012 Eighth International Conference on Computational Intelligence and Security, Guangzhou, China, pp. 314-317. <https://doi.org/10.1109/CIS.2012.77>
- [17] Nencini, F., Garzelli, A., Baronti, S., Alparone, L. (2007). Remote sensing image fusion using the curvelet transform. *Information Fusion*, 8(2): 143-156. <https://doi.org/10.1016/j.inffus.2006.02.001>
- [18] Khare, A., Khare, M., Srivastava, R. (2021). Shearlet transform based technique for image fusion using median fusion rule. *Multimedia Tools and Applications*, 80: 11491-11522. <https://doi.org/10.1007/s11042-020-10184-1>

- [19] Pandit, V.R., Bhiwani, R.J. (2015). Image fusion in remote sensing applications: A review. *International Journal of Computer Applications*, 120(10): 22-32. <https://doi.org/10.5120/21263-3846>
- [20] Liu, Y., Chen, X., Wang, Z.F., Wang, Z.J., Ward, R.K., Wang, X.S. (2018). Deep learning for pixel-level image fusion: Recent advances and future prospects. *Information Fusion*, 42: 158-173. <https://doi.org/10.1016/j.inffus.2017.10.007>
- [21] Zhang, W.F., Zhao, R.P., Yao, Y.X., Wan, Y., Wu, P.H., Li, J.Y., Li, Y.S., Zhang, Y.J. (2025). Multi-resolution SAR and optical remote sensing image registration methods: A review, datasets, and future perspectives. *arXiv preprint arXiv:2502.01002*. <https://doi.org/10.48550/arXiv.2502.01002>
- [22] Li, H., Wu, X.J. (2019). DenseFuse: A fusion approach to infrared and visible images. *IEEE Transactions on Image Processing*, 28(5): 2614-2623. <https://doi.org/10.1109/TIP.2018.2887342>
- [23] Prabhakar, K.R., Srikar, V.S., Babu, R.V. (2017). DeepFuse: A deep unsupervised approach for exposure fusion with extreme exposure image pairs. *arXiv preprint arXiv:1712.07384*. <https://doi.org/10.48550/arXiv.1712.07384>
- [24] Zhang, H., Xu, H., Tian, X., Jiang, J.J., Ma, J.Y. (2021). Image fusion meets deep learning: A survey and perspective. *Information Fusion*, 76: 323-336. <https://doi.org/10.1016/j.inffus.2021.06.008>
- [25] Li, H., Wu, X.J., Durrani, T. (2020). NestFuse: An infrared and visible image fusion architecture based on nest connection and spatial/channel attention models. *IEEE Transactions on Instrumentation and Measurement*, 69(12): 9645-9656. <https://doi.org/10.1109/TIM.2020.3005230>
- [26] Lian, Z.L., Zhan, Y.L., Zhang, W.H., Wang, Z.J., Liu, W.B., Huang, X.H. (2025). Recent advances in deep learning-based spatiotemporal fusion methods for remote sensing images. *Sensors*, 25(4): 1093. <https://doi.org/10.3390/s25041093>
- [27] Lin, T.Y., Maire, M., Belongie, S., Bourdev, L., et al. (2014). Microsoft COCO: Common objects in context. *arXiv preprint arXiv:1405.0312*. <https://doi.org/10.48550/arXiv.1405.0312>
- [28] Wang, K.P., Zheng, M.Y., Wei, H.Y., Qi, G.Q., Li, Y.Y. (2020). Multi-modality medical image fusion using convolutional neural network and contrast pyramid. *Sensors*, 20(8): 2169. <https://doi.org/10.3390/s20082169>
- [29] Zhang, H., Xu, H., Xiao, Y., Guo, X.J., Ma, J.Y. (2020). Rethinking the image fusion: A fast unified image fusion network based on proportional maintenance of gradient and intensity. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(7): 12794-12804. <https://doi.org/10.1609/aaai.v34i07.6975>
- [30] Ma, J.Y., Yu, W., Liang, P.W., Li, C., Jiang, J.J. (2019). FusionGAN: A generative adversarial network for infrared and visible image fusion. *Information Fusion*, 48: 11-26. <https://doi.org/10.1016/j.inffus.2018.09.004>
- [31] Iervolino, P., Guida, R., Riccio, D., Rea, R. (2019). A novel multispectral, panchromatic and SAR data fusion for land classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(10): 3966-3979. <https://doi.org/10.1109/JSTARS.2019.2945188>
- [32] Mao, R., Fu, X.S., Niu, P.J., Wang, H.Q., Pan, J., Li, S.S. (2018). Multi-directional Laplacian pyramid image fusion algorithm. In *2018 3rd International Conference on Mechanical, Control and Computer Engineering (ICMCCE)*, Huhhot, China, pp. 568-572. <https://doi.org/10.1109/ICMCCE.2018.00125>
- [33] Li, J.J., Zhang, J.C., Yang, C., Liu, H.Y., Zhao, Y.G., Ye, Y.X. (2023). Comparative analysis of pixel-level fusion algorithms and a new high-resolution dataset for SAR and optical image fusion. *Remote Sensing*, 15(23): 5514. <https://doi.org/10.3390/rs15235514>
- [34] Zhang, H., Shen, H.F., Yuan, Q.Q., Guan, X.B. (2022). Multispectral and SAR image fusion based on Laplacian pyramid and sparse representation. *Remote Sensing*, 14(4): 870. <https://doi.org/10.3390/rs14040870>
- [35] Chen, J.Y., Zhang, L., Lu, L., Li, Q.L., Hu, M.F., Yang, X.M. (2021). A novel medical image fusion method based on Rolling Guidance Filtering. *Internet of Things*, 14: 100172. <https://doi.org/10.1016/j.iot.2020.100172>
- [36] Lewis, J.J., O'Callaghan, R.J., Nikolov, S.G., Bull, D.R., Canagarajah, N. (2007). Pixel- and region-based image fusion with complex wavelets. *Information Fusion*, 8(2): 119-130. <https://doi.org/10.1016/j.inffus.2005.09.006>
- [37] Yin, H.P., Li, Y.X., Chai, Y., Liu, Z.D., Zhu, Z.Q. (2016). A novel sparse-representation-based multi-focus image fusion approach. *Neurocomputing*, 216: 216-229. <https://doi.org/10.1016/j.neucom.2016.07.039>
- [38] Li, S.T., Yin, H.T., Fang, L.Y. (2013). Remote sensing image fusion via sparse representations over learned dictionaries. *IEEE Transactions on Geoscience and Remote Sensing*, 51(9): 4779-4789. <https://doi.org/10.1109/TGRS.2012.2230332>
- [39] Yang, B., Li, S.T. (2010). Multifocus image fusion and restoration with sparse representation. *IEEE Transactions on Instrumentation and Measurement*, 59(4): 884-892. <https://doi.org/10.1109/TIM.2009.2026612>
- [40] Zhang, Q., Shen, X.Y., Xu, L., Jia, J.Y. (2014). Rolling guidance filter. In *Computer Vision – ECCV 2014*. *ECCV 2014. Lecture Notes in Computer Science*, pp. 815-830. https://doi.org/10.1007/978-3-319-10578-9_53
- [41] Jiang, Y., Wang, M.H. (2014). Image fusion using multiscale edge-preserving decomposition based on weighted least squares filter. *IET Image Processing*, 8(3): 183-190. <https://doi.org/10.1049/iet-ipr.2013.0429>
- [42] Duan, C.W., Wang, Z.S., Xing, C.D., Lu, S.S. (2021). Infrared and visible image fusion using multi-scale edge-preserving decomposition and multiple saliency features. *Optik*, 228: 165775. <https://doi.org/10.1016/j.ijleo.2020.165775>
- [43] Wang, P.S., Fu, X.M., Liu, Y., Tong, X., Liu, S.L., Guo, B.N. (2015). Rolling guidance normal filter for geometric processing. *ACM Transactions on Graphics*, 34(6): 1-9. <https://doi.org/10.1145/2816795.2818068>
- [44] Kaplan, N.H., Erer, I. (2021). Scale aware remote sensing image enhancement using rolling guidance. *Journal of Visual Communication and Image Representation*, 80: 103315. <https://doi.org/10.1016/j.jvcir.2021.103315>
- [45] Lin, Y.C., Cao, D.X., Zhou, X.C. (2022). Adaptive infrared and visible image fusion method by using rolling guidance filter and saliency detection. *Optik*, 262: 169218. <https://doi.org/10.1016/j.ijleo.2022.169218>
- [46] Liu, Y., Liu, S.P., Wang, Z.F. (2015). A general framework for image fusion based on multi-scale

- transform and sparse representation. *Information Fusion*, 24: 147-164.
<https://doi.org/10.1016/j.inffus.2014.09.004>
- [47] Gong, X.Q., Hou, Z.Y., Ma, A.L., Zhong, Y.F., Zhang, M., Lv, K.Y. (2023). An adaptive multiscale gaussian co-occurrence filtering decomposition method for multispectral and SAR image fusion. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 16: 8215-8229.
<https://doi.org/10.1109/JSTARS.2023.3296505>
- [48] Yin, M., Liu, X.N., Liu, Y., Chen, X. (2019). Medical image fusion with parameter-adaptive pulse coupled neural network in nonsubsampling Shearlet transform domain. *IEEE Transactions on Instrumentation and Measurement*, 68(1): 49-64.
<https://doi.org/10.1109/TIM.2018.2838778>
- [49] Demirel, H., Anbarjafari, G. (2011). Discrete wavelet transform-based satellite image resolution enhancement. *IEEE Transactions on Geoscience and Remote Sensing*, 49(6): 1997-2004.
<https://doi.org/10.1109/TGRS.2010.2100401>
- [50] Adeel, H., Tahir, J., Riaz, M.M., Ali, S.S. (2022). Siamese networks based deep fusion framework for multi-source satellite imagery. *IEEE Access*, 10: 8728-8737. <https://doi.org/10.1109/ACCESS.2022.3143847>
- [51] Li, J., Shi, X.Q., Li, Y.N., Zhou, H.B. (2025). Exploring fusion domain: Advancing infrared and visible image fusion via IDFFN-GAN. *Neurocomputing*, 611: 128647.
<https://doi.org/10.1016/j.neucom.2024.128647>

NOMENCLATURE

F	Image in JBF Filter related operation
S	Multiscale decomposed layer image
A	Approximation layer
D	Contour and detail layer
I'	Processed I component

Greek symbols

σ	Gaussian filter parameter
α, β	Tuneable weights
ω_1, ω_2	Calculated weights in LSEM model
E_1, E_2	Edge strengths of both input images
μ	Mean of image

Subscripts

in, out	Input and output image
r	Gaussian structure scale Controlling parameter
s	Gaussian distribution
d	Detail layer
c	Contour layer
a	Approximation layer
F	Fused image
SAR	SAR image
MS	MS image
new	Layer-wise fused image
CNN	CNN method fused image
$diff$	Absolute difference image
$source$	Input image
$fused$	Fused image