

Article

Multi-Sensor Medical-Image Fusion Technique Based on Embedding Bilateral Filter in Least Squares and Salient Detection

Jiangwei Li ¹, Dingan Han ¹, Xiaopan Wang ², Peng Yi ³, Liang Yan ³ and Xiaosong Li ^{1,*} 

¹ Guangdong-Hong Kong-Macao Joint Laboratory for Intelligent Micro-Nano Optoelectronic Technology, School of Physics and Optoelectronic Engineering, Foshan University, Foshan 528225, China

² Guangdong Province Graduate Joint Training Base (Foshan), Foshan University, Foshan 528225, China

³ Jiangsu Shuguang Photoelectric Co., Ltd., Yangzhou 225009, China

* Correspondence: lixiaosong@buaa.edu.cn

Abstract: A multi-sensor medical-image fusion technique, which integrates useful information from different single-modal images of the same tissue and provides a fused image that is more comprehensive and objective than a single-source image, is becoming an increasingly important technique in clinical diagnosis and treatment planning. The salient information in medical images often visually describes the tissue. To effectively embed salient information in the fused image, a multi-sensor medical image fusion method is proposed based on an embedding bilateral filter in least squares and salient detection via a deformed smoothness constraint. First, source images are decomposed into base and detail layers using a bilateral filter in least squares. Then, the detail layers are treated as superpositions of salient regions and background information; a fusion rule for this layer based on the deformed smoothness constraint and guided filtering was designed to successfully conserve the salient structure and detail information of the source images. A base-layer fusion rule based on modified Laplace energy and local energy is proposed to preserve the energy information of these source images. The experimental results demonstrate that the proposed method outperformed nine state-of-the-art methods in both subjective and objective quality assessments on the Harvard Medical School dataset.

Keywords: medical-image fusion; embedding bilateral filter; salient detection



Citation: Li, J.; Han, D.; Wang, X.; Yi, P.; Yan, L.; Li, X. Multi-Sensor Medical-Image Fusion Technique Based on Embedding Bilateral Filter in Least Squares and Salient Detection. *Sensors* **2023**, *23*, 3490. <https://doi.org/10.3390/s23073490>

Academic Editors: Alessandro Bevilacqua and Margherita Mottola

Received: 8 February 2023

Revised: 21 March 2023

Accepted: 23 March 2023

Published: 27 March 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The technique of image fusion integrates multiple images generated by different sensors with different descriptions of the same scene to produce an image with more compatible and accurate information [1]. The main image fusion technologies include multi-focus image fusion, medical image fusion, infrared and visible image fusion, remote sensing image fusion, etc. This technique has been widely applied in the fields of surveillance, clinical diagnostics, automation, national defense, biometrics, and remote sensing. Medical image fusion, which integrates all the useful complementary information from different medical images into a fused image, as a branch of image fusion, occupies a crucial position in research. The use of a single sensor-formed image as a basis for judgment has limitations while describing the health status of tissues (for example, using computed tomography (CT) that detects only dense structures such as bones and implants; magnetic resonance imaging (MRI) that provides soft tissue information; positron emission tomography (PET), which reflects the biological activity of cells and molecules; and single-photon emission computed tomography (SPECT), which reflects the blood flow through tissues and organs). Fused images can provide a more comprehensive, reliable, and better description of lesions, thereby making a significant contribution to biomedical research and clinical

diagnosis techniques, such as surgical navigation, radiotherapy planning, and future-health prediction methods [2].

In recent years, multi-sensor medical-image fusion techniques have been developed. The four main types include: spatial domain-based (SDB), transform domain-based (TDB) [3], sparse representation-based (SRD) [4], and deep learning-based (DLB) methods [5].

The TDB method typically consists of three steps: multiscale decomposition, fusion, and multiscale reconstruction. This kind of method decomposes images into different scales, which is analogous to the process of human eyes dealing with visual information ranging from coarse to fine, enabling a better signal-to-noise ratio [6]. Zhu et al. [7] decomposed images using a non-subsampled contourlet transform (NSCT) and fused the corresponding high and low frequencies using phase congruency (PC) and local Laplacian energy (LLE), respectively. Li et al. [8] used the Laplacian redecomposition (LRD) scheme to decompose images and fused subbands based on overlapping and non-overlapping domains. In addition to selecting transform tools, fusion rules also play an integral role in TDB methods. Because fusion rules achieve image fusion by directly processing pixels or regions that tend to be considered in terms of points or regions, they cannot solve the problem of extracting edge information correctly. Moreover, TDB methods are computationally expensive and may result in inevitable losses in details and generate artifacts during decomposition and reconstruction, which could reduce the fusion performance [9,10].

SRB methods mainly work on the concept that image signals can be considered a linear combination of atoms in an overcomplete dictionary. Most SRB methods involve the following steps [4]: (a) segmenting the source image into overlapping patches, (b) sparse encoding of the patches using a dictionary to obtain their sparse coefficients, (c) combining the sparse coefficients, and (d) reconstructing the images from the sparse coefficients and a dictionary. Common sparse representation (SR) models include the traditional SR model [11], group-sparsity SR model [12], robust SR model [13], non-negative SR model [14], and joint convolutional analysis and synthesis [15]. Li et al. [11] used low-pass filtering and structured texture filtering to decompose an image and fused high-frequency layers with sparse representation to achieve image fusion and denoising. Jie et al. [16] used sparse representation and a rolling guidance filter (RGF) to fuse texture layers using cartoon-texture image decomposition. When compared to TDB methods, SRB methods allow for a more meaningful and stable representation of the source image owing to the overcomplete dictionary containing richer basis atoms. Moreover, using a fixed step size to acquire image blocks is also effective in reducing artifacts and improving robustness to misalignment. Additionally, sparse encoding under SRB methods is usually time-consuming and complex, and important information from the source images is inevitably lost [16,17].

In recent years, DLB methods have received considerable attention owing to their powerful nonlinear fitting capability. In the convolutional neural network (CNN)-based approach proposed by Zhang et al. [18], two convolutional layers were used to achieve feature extraction and the reconstruction of images. In addition to CNNs, generative adversarial networks (GANs) have also been applied to image domains. Ma et al. [19] proposed a GAN for image fusion using a generator and a discriminator to achieve maintenance of intensity and detail information in the source image. The auto-encoder-based fusion framework for feature extraction and image reconstruction uses pre-trained autoencoders. Luo et al. [20] used a multi-branch encoder with contrast constraints to learn the public and private features of an image, fused the private features using an adaptive fusion rule based on energy, and then reconstructed the image using a decoder. The end-to-end fusion process in DLB methods effectively reduces pre-processing, parameter tuning, and post-processing. However, DLB methods are time-consuming for model training and require large datasets [21,22].

SDB methods rely on detecting pixel-level activities, which reflect features such as the level of image sharpness and structural saliency. The main steps are as follows. First, the activity of a pixel or region is detected by a specific function or algorithm to obtain the activity map of the image. Then, according to a given rule (e.g., the “maximum absolute value (Abs)” rule), it generates an active decision map. Finally, the decision map is used to reconstruct the source image to obtain a fused image. In SDB methods, image processing using edge-preserving filters has become increasingly common. The base layer of the image is obtained using an edge-preserving filter to potentially capture large variations and a set of detail layers to preserve detail at progressively more refined levels. Mo et al. [23] proposed an attribute-filter-based image fusion method wherein the prominent objects in the image were first extracted using attribute- and edge-preserving filters, and then the fusion results were obtained using a weight-based Laplacian-pyramid image-fusion strategy. Overall, SDB methods are simple and fast, but pixel-level activity detection is not an easy task, and incorrect activity detection may lead to the occurrence of blocking (region) artifacts, introduce certain spectral distortions, and degrade the sharpness of the fusion results [24].

Although existing multi-sensor medical-image fusion techniques have achieved great success, certain shortcomings still exist. For example, the atoms of the dictionary in SRB methods have a limited ability to represent salient features in the image [25]. The fusion rules in TDB and SDB methods are often based on pixels or regions without consideration of edges or structures in the image [26]. In addition, most existing methods lack attention to salient information, which is often a visual reflection of tissue health status in medical images. To retain the salient information in the source images, in this paper, we propose a medical image fusion method based on least-squares using the bilateral filter (BLF-LS) and deformation smoothness constraint (DSC) [24], which can effectively retain the salient information, edge, and energy from source images.

The BLF-LS is a recently developed edge-preserving filter. It takes advantage of bilateral filtering and the least-squares (LS) model, effectively smoothing the edges within the texture region while producing results without gradient reversals and halos; it also offers the advantage of fast operation [27]. Therefore, we introduced the BLF-LS to decompose the source image. A fusion rule combining DSC and the rolling guidance filter (RGF) [25] was designed to fuse detail layers. Saliency describes what attracts the visual attention of humans in a bottom-up manner. Saliency detection can maintain the integrity of important target regions and enables high-quality image fusion. The main contributions of this study are as follows:

1. A medical image fusion method based on the BLF-LS and saliency detection is proposed. To the best of our knowledge, this is the first time the BLF-LS has been applied in medical-image fusion. The source images are decomposed into the detail and base layers.
2. A detail-layer fusion rule based on DSC and RGF is proposed, which fully considers the low contrast between the target and background.
3. A fusion rule based on modified Laplace energy and local energy (MLEN) was designed to maintain detailed information and energy in the base layer.
4. The proposed fusion method can be effectively extended to the IR- and VIS-image fusion problem and yield competing fusion performance.

The remainder of this paper is organized as follows. In Section 2, the background of the BLF-LS and saliency detection using a DSC is briefly introduced. Section 3 explains the proposed image-fusion algorithm. The experimental results and discussion are presented in Section 4. Finally, Section 5 concludes the paper.

2. Related Work

2.1. Embedding Bilateral Filter in Least Squares

Edge-preserving filters offer many advantages, such as accurately separating image structures at different scales while maintaining the spatial consistency of these structures, re-

ducing the blurring effects around edges, providing a good edge- and boundary-preserving performance, and smoothing background information. The BLF-LS is an edge-preserving filter achieved using global methods [27]. The smoothing result of this filter is free of gradient reversals and halos. Additionally, the BLF-LS runs faster because it utilizes the efficiency of bilateral filter (BLF) and the LS model. To facilitate the understanding of the BLF-LS, we first describe BLF. For a given image g , the output image μ through the BLF is computed as follows:

$$\mu_s = \frac{1}{G_{\sigma_s}(\|s - t\|)G_{\sigma_r}(\|g_s - g_t\|)} \sum_{t \in N(s)} G_{\sigma_s}(\|s - t\|)G_{\sigma_r}(\|g_s - g_t\|)g_t, \quad (1)$$

where s and t denote different pixel points, G_{σ_s} denotes the Gaussian kernel that determines the spatial support, and G_{σ_r} denotes the Gaussian kernel that controls the sensitivity to the edges. The BLF has the advantage of fast image processing. However, because the edges are sharpened in the smoothed image and boosted in the reverse direction in the enhanced image, gradient inversion and halos are produced in the result.

Suppose $f_{BLF}(\nabla g_*)$ denotes the smoothing gradients, and $*$ denotes the axis direction of the input image g with the BLF; embedding $f_{BLF}(\nabla g)$ into the LS framework achieves efficient edge-preserving smoothing. This approach allows the BLF-LS to achieve both the edge-smoothing quality of the LS and BLF models with a proper processing efficiency, as described below. Given an input image g , the output image μ with BLF-LS is:

$$\min_{\mu} \sum_s \left((\mu_s - g_s)^2 + \lambda \sum_{* \in \{x, y\}} (\nabla \mu_{*,s} - (f_{BLF}(\nabla g_*)_s))^2 \right), \quad (2)$$

where s denotes the pixel position. When the value of λ is large enough, the gradient of the image μ , that is, $\nabla \mu_{*,s}$, will resemble $f_{BLF}(\nabla g_*)(* \in \{x, y\})$, which guarantees the smooth quality of the BLF-LS. Because the LS model can be solved in the Fourier domain, the speed of the BLF-LS is guaranteed. Equation (2) can be solved as follows:

$$\mu = \mathcal{F}^{-1} \left(\frac{\mathcal{F}(g) + \lambda \sum_{* \in \{x, y\}} (\mathcal{F}(f_{BLF}(\nabla g_*)))}{\mathcal{F}(1) + \lambda \sum_{* \in \{x, y\}} \overline{\mathcal{F}(\partial_*)} \cdot \mathcal{F}(\partial_*)} \right), \quad (3)$$

where $\mathcal{F}(\cdot)$ and $\mathcal{F}^{-1}(\cdot)$ are the fast Fourier transform (FFT) and inverse fast Fourier transform (IFFT) operators, respectively; $\overline{\mathcal{F}(\cdot)}$ denotes the complex conjugate of $\mathcal{F}(\cdot)$; and $\mathcal{F}(1)$ is the FFT of the delta function. Additionally, multiplication and division are both point-wise operations.

2.2. Salient Detection via Deformed Smoothness Constraint

The DSC [28] is a propagation model that can capture significant targets when there is low contrast between the object regions and background. It comprises three main steps. First, the image is segmented using superpixels, and the segmentation result is represented as a graph. Then, a coarse map is generated via the background seeds and a deformed smoothness-based manifold ranking model, and the objectness map is built through the object proposal. Finally, the coarse and objectness maps are used to generate a refined map.

The input image I generates a significant detection map \mathbf{g} described as follows:

$$\min_{\mathbf{g}} \frac{1}{2} \left\{ \mathbf{g}^T \left[D^c - W^c + \mu \left(I - \frac{D^c}{v^c} \right) \right] \mathbf{g} + \|\mathbf{g} - M^c\|^2 + \mathbf{g}^T D^o \mathbf{g} \right\}, \quad (4)$$

where M^c is the coarse map; D^c and v^c are the degree matrix and volume of M^c respectively; W^c is a weight matrix computed by M^c ; μ is a non-negative parameter that balances the weights of the two smoothness constraints; M^o denotes the objectness map obtained for

each node using an edge box ($M^o = (m_i^o)_n$); and $D^o = (d_{ii}^o)_{n \times n}$ represents a diagonal matrix with $d_{ii}^o = \text{diag}(\exp(-m_i^o))$.

The optimal solution of Equation (4) is expressed as:

$$g = \left[D^c - W^c + \mu_2 \left(I - \frac{D^c}{v^c} \right) + D^o \right]^{-1} M^c \tag{5}$$

The elements of g were normalized to $[0, 1]$ and assigned to the corresponding super-pixels to generate a saliency detection map.

3. Proposed Method

The proposed method is illustrated in Figure 1. First, the source images are decomposed into a base layer and a detail layer via the BLF-LS. The base layer, which is obtained by decomposing the source image, is fused based on the MLEN fusion rules to retain the energy information of the source image. Moreover, the detail layers are considered a superposition of the salient regions and background information. The detail layers are decomposed into background-detail layers and salient-detail layers using a model based on the DSC and RGF. To fully retain the energy information, the background-detail layer of the fused image is obtained using the fusion rule, Abs. Regarding the salient-detail layers that contain important salient targets, the overlap between the two salient-detail layers is removed using the DSC-RGF model, and a direct summation method is used to obtain the salient-detail layer of the fused image. Finally, the fused image is obtained by reconstruction.

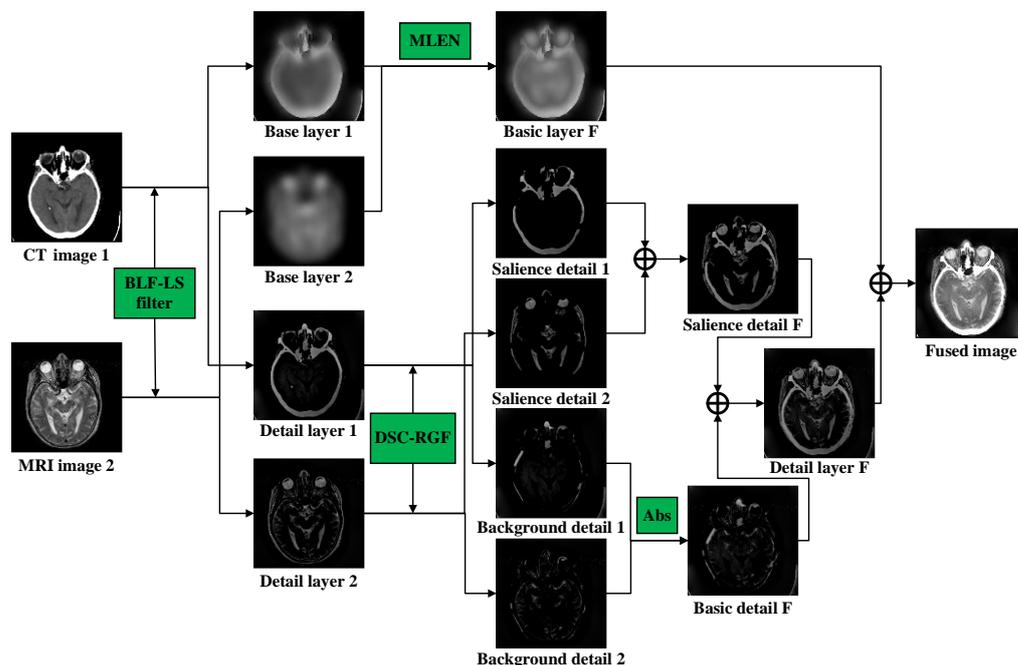


Figure 1. Flowchart of proposed image-fusion method.

Additionally, for a functional medical image fusion problem, the following conversion scheme is used: red, green, and blue (RGB) \rightarrow luma, blue projection, and red projection (YUV) \rightarrow RGB, as shown in Figure 2.

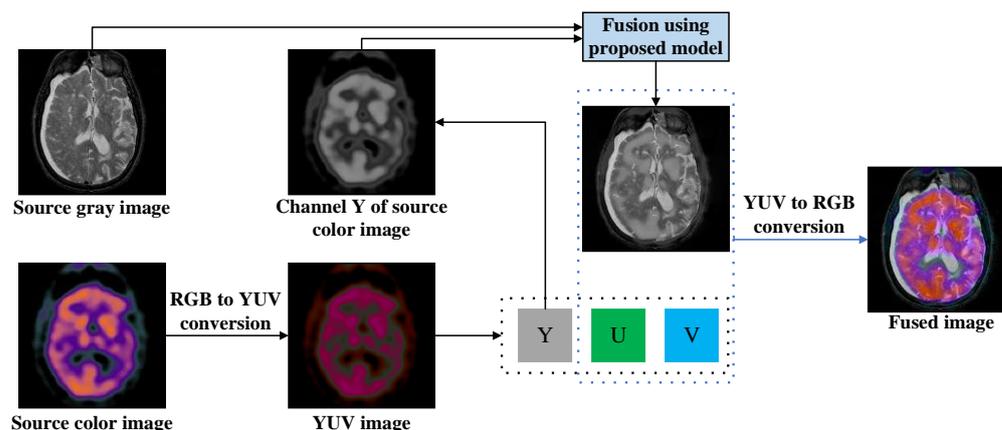


Figure 2. Scheme for fusing medical images with color.

3.1. Decomposition of Base Layer and Detail Layer

BLF-LS achieves useful edge-preserving smoothing by embedding the BLF into the LS framework. First, we employed a BLF-LS to decompose the source images into a base layer and a detail layer. The details of the base layer are as follows:

$$B_n = I_n * F_{BLF-LS} \quad (n = 1, 2), \quad (6)$$

where I_1 and I_2 denote the source images, B_1 and B_2 are the base layers obtained by decomposing I_1 and I_2 , respectively; and F_{BLF-LS} is a BLF-LS used for smoothing the images described in Equation (2). After the base layer is obtained, it can be subtracted from the source image to obtain the detail layer:

$$D_n = I_n - B_n \quad (n = 1, 2), \quad (7)$$

where D_1 and D_2 are the detail layers obtained by decomposing I_1 and I_2 , respectively. The base layer can potentially capture large variations in intensity, and the detail layer can preserve details at fine scales.

3.2. Decomposition of Detail Layer Based on DSC-RGF Algorithm

In recent years, many saliency detection methods that can detect salient visual areas or objects and easily draw visual attention have been proposed. It is typically easier to detect salient targets in the detail layer obtained using a smoothing filter. To this end, we designed a method based on the DSC and RGF to fuse the detail layer, as shown in Figure 3; this method mainly consists of the following steps. First, the initial salient decision map is obtained by applying the DSC to the detail layer. Second, the overlapping part of the two initial salient decision maps produces ghosting in the fusion results and affects the visual effect; therefore, the overlap-removal procedure is performed on the initial salient decision map. Then, in view of the edge-smoothing problem of the significant target in the salient decision map, the RGF is used to process the salient decision map to obtain the salient guided filtering (SGF) map. Finally, the detail layer is decomposed into background and salient-detail layers using this map.

The details of each step are as follows. First, the DSC model is used to detect the detail layer to obtain the salient information, and threshold correction is adopted to process the salient information of the initial salient decision map.

$$I_{ID}^n(x, y) = \begin{cases} 1 & D_n(x, y) * F_{DSC} \geq T \\ 0 & D_n(x, y) * F_{DSC} < T \end{cases} \quad (n = 1, 2), \quad (8)$$

where $*F_{DSC}$ denotes the salient-detection operation using the DSC model in Equation (4); T is the threshold value; and $I_{ISD}^1(x, y)$ and $I_{ISD}^2(x, y)$ are the initial salient decision maps obtained from $D_1(x, y)$ and $D_2(x, y)$, respectively.

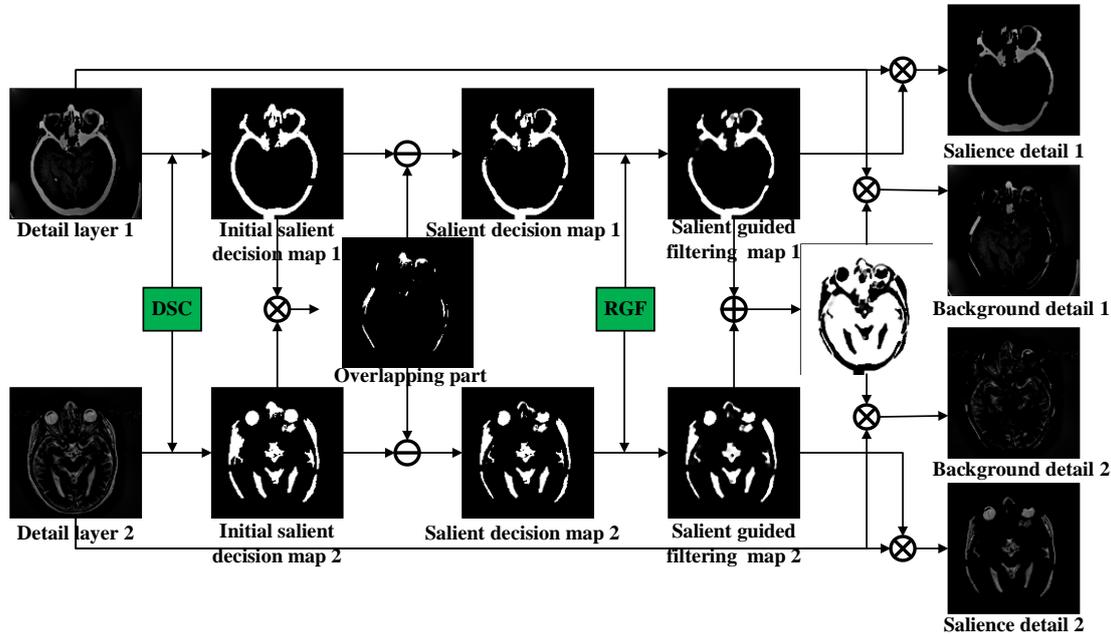


Figure 3. Deformed smoothness constraint–rolling guidance filter (DSC-RGF) algorithm.

Second, it is necessary to remove the overlapping part I_R , which is generated by multiplying I_{ISD}^1 and I_{ISD}^2 ($I_R = I_{ISD}^1 \cdot I_{ISD}^2$). Directly phasing it into the fusion result causes ghosting, which affects the visual effect.

$$I_{SD}^n = I_{ISD}^n - I_R \quad (n = 1, 2), \quad (9)$$

where I_{SD}^1 and I_{SD}^2 represent the salient decision maps obtained after removing the overlapping parts. Considering the edge-smoothing problem of the significant target in the salient decision map, we used RGF to process I_{SD}^n to obtain the SGF map.

$$I_{SGF}^n = I_{SD}^n * F_{RGF}(T', r, \epsilon) \quad (n = 1, 2), \quad (10)$$

where $F_{RGF}(\cdot)$ represents the RGF function; T' denotes the number of iterations; r denotes the filter size; ϵ denotes the degree of blur; and I_{SGF}^1 and I_{SGF}^2 denote the SGF maps used to decompose the detail-layer maps of the source image.

Finally, the salient-detail layers are obtained by multiplying the SGF maps by the detail layers, and the background-detail maps are obtained by removing the salient parts of the detail layers, as described below.

$$SD_n(x, y) = D_n(x, y) \cdot I_{SGF}^n(x, y) \quad (n = 1, 2), \quad (11)$$

$$BD_n(x, y) = D_n(x, y) \cdot (1 - I_{DGF}^1(x, y) - I_{DGF}^2(x, y)) \quad (n = 1, 2) \quad (12)$$

where $SD_1(x, y)$ and $SD_2(x, y)$ denote the salient detail layers obtained from the decomposition of $D_1(x, y)$ and $D_2(x, y)$, respectively; and $BD_1(x, y)$ and $BD_2(x, y)$ denote the background detail layers obtained from the decomposition of $D_1(x, y)$ and $D_2(x, y)$, respectively.

3.3. Fusion of Base Layer Based on MLEN

The pair of base layers obtained under the BLF-LS decomposition contain abundant energy information and little information on the detail in the source image. Therefore, we considered using local energy (LEN) to extract the energy information and sum-modified Laplacian (SML) energy to extract the detail-related information from the base layers and finally add these two types of information to obtain the fused base layer. SML is defined as follows [29]:

$$SML_n(x, y) = \sum_{m=-M}^M \sum_{n=-N}^N ML_n(x + m, y + n)^2 \quad (n = 1, 2), \quad (13)$$

where $M \times N$ denotes the window size centered at (x, y) , and $ML_n(x, y)$ denotes the modified Laplacian (ML) at point (x, y) ; $ML_n(x, y)$ is defined as follows:

$$ML_n(x, y) = |2B_n(x, y) - B_n(x - 1, y) - B_n(x + 1, y)| + |2B_n(x, y) - B_n(x, y - 1) - B_n(x, y + 1)| \quad (n = 1, 2), \quad (14)$$

where B_1 and B_2 are the base layers decomposed from I_1 and I_2 , respectively. LEN is defined as follows:

$$LEN_n(x, y) = \sum_{m=-M}^M \sum_{n=-N}^N B_n(x + m, y + n)^2 \quad (n = 1, 2), \quad (15)$$

where $M \times N$ denotes the window size centered at (x, y) , and the fusion of the base layer can be briefly described as follows:

$$B_F(x, y) = \begin{cases} B_1(x, y), & SML_1(x, y) \geq SML_2(x, y) \text{ or } LEN_1(x, y) \geq LEN_2(x, y) \\ B_2(x, y), & \text{else} \end{cases} \quad (16)$$

where $B_F(x, y)$ denotes the base layer of the fused image.

3.4. Fusion Result

$BD_1(x, y)$ and $BD_2(x, y)$ contain most of the energy information from the original image. To avoid excessive energy loss in the fused image, we used the fusion rule of taking the Abs to obtain the background-detail layer of the fused image $BD_F(x, y)$:

$$BD_F(x, y) = \begin{cases} BD_1(x, y), & BD_1(x, y) \geq BD_2(x, y) \\ BD_2(x, y), & \text{else} \end{cases}. \quad (17)$$

Because the salient-detail layers contain significant information, they are fused by direct summation, as follows:

$$SD_F(x, y) = SD_1(x, y) + SD_2(x, y). \quad (18)$$

where $SD_F(x, y)$ denotes the salient-detail layers of the fused image. Finally, the fused image is obtained by combining the base, salient-detail, and background-detail layers, as follows:

$$I_F(x, y) = B_F(x, y) + BD_F(x, y) + SD_F(x, y). \quad (19)$$

The formal mechanism of the proposed method is described in Algorithm 1.

Algorithm 1 Steps in proposed fusion method

Inputs: Medical CT Image I_1 ; Medical MRI Image I_2

Output: Fused image F

Step 1: The BLF-LS is employed to decompose I_1 and I_2 to obtain the corresponding base layers B_1 and B_2 and detail layers D_1 and D_2 (Equations (6) and (7)).

Step 2: The DSC-RGF algorithm is utilized to decompose the detail layers D_1 and D_2 to obtain the corresponding significant-detail layers SD_1 and SD_2 and background-detail layers BD_1 and BD_2 (Equations (8)–(12)).

Step 3: The fusion base layer B_F is obtained using the MLEN rule (Equations (13)–(16)). Then, the fusion Abs rule is employed to fuse BD_1 and BD_2 and thereby obtain the fusion base layer BD_F (Equation (17)). SD_1 and SD_2 are added to obtain the significant-detail layer SD_F of the fused image (Equation (18)).

Step 4: The fused image is obtained by summing GD_F , B_F , and SD_F (Equation (19)).

4. Experimental Results and Comparisons

4.1. Experimental Setup

4.1.1. Test Data

The experimental dataset was selected from 100 sets each comprising a CT-MRI, PET-MRI, and SPECT-MRI image, amounting to 300 source images for testing. The source images were images of the human brain captured by different imaging mechanisms; each image was 256×256 pixels, and each pair of images was aligned. The test images were obtained from a database of Harvard Medical School (<http://www.med.harvard.edu/aanlib/home.html>, accessed on 8 November 2022).

4.1.2. Quantitative Evaluation Metrics

Subjective quality assessment of image fusion represents human intuition but lacks quantitative description, so objective quality assessment is also needed to evaluate the performance of fusion algorithms. The metrics used for objective quality evaluation of images typically include three categories: information theory-based, image feature-based, and human perception-inspired fusion metrics. In this study, six common metrics were selected to objectively assess the fusion performance: normalized mutual information (Q_{MI}) [30], image fusion metric based on a multiscale scheme (Q_M) [31], nonlinear correlation information entropy (Q_{NCIE}) [32], metric based on phase congruency (Q_P) [33], entropy (EN) [34], and visual information fidelity (VIF) [35].

Q_{MI} is a quality index that describes the quantity of information conveyed from the source image to the fused image; Q_{NCIE} is used to display the nonlinear correlation degree of the concerned multivariable dataset; EN measures the amount of information contained in the fused image; Q_{MI} , Q_{NCIE} , and EN are evaluation metrics based on information theory. Q_M evaluates the retention value of the edge information in fused images from multiple scales; Q_P is defined by the maximum and minimum moments of phase coherence and is used to evaluate the angle and edge information measures; Q_M and Q_P are evaluation metrics based on image features. The VIF metric measures the information fidelity of the fused image, and the distortions of the images include additive noise, blur, and global or local changes in contrast. VIF is a fusion measure inspired by human perception. Table 1 shows a summary of these six metrics. A comprehensive and objective evaluation of the fused image quality is achieved by considering these metrics, and the larger the value of all these metrics, the better the quality of the fused image [36].

Table 1. Quantitative evaluation metrics used in the experiments.

| Metric | Mathematical Expression | Definition | Best Value Outcomes |
|--------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------|---------------------|
| 1 | $Q_{MI} = 2 \left[\frac{MI(A,F)}{H(A)+H(F)} + \frac{MI(B,F)}{H(B)+H(F)} \right]$ | Measure of retention value of edge information | Higher |
| 2 | $Q_{NCIE} = 1 + \sum_{i=1}^3 \frac{\lambda_i}{3} \log_2 \frac{\lambda_i}{3}$ | Measure of nonlinear correlation information entropy | Higher |
| 3 | $Q_M = \prod_{s=1}^N \left(Q_s^{\frac{AB}{F}} \right)^{\alpha_s}$ | Measure of retention value of edge information | Higher |
| 4 | $Q_P = (P_p)^\alpha (P_M)^\beta (P_m)^\gamma,$ | Measure of phase congruency | Higher |
| 5 | $EN = - \sum_{l=0}^{L-1} p_l \log_2 p_l$ | Evaluation metrics based on information theory | Higher |
| 6 | $VIF = \frac{\sum_{j \in \text{subbands}} I(C^{\rightarrow N_j} ; F^{\rightarrow N_j} S^{N_j})}{\sum_{j \in \text{subbands}} I(C^{\rightarrow N_j} ; E^{\rightarrow N_j} S^{N_j})}$ | A fusion measure inspired by human perception | Higher |

4.1.3. Methods Compared with Proposed Methods

To verify the effectiveness of the proposed methods, our results were compared with those of nine state-of-the-art fusion algorithms, including two SD methods, namely information of interest in local Laplacian filtering (ILLF) [37] and LRD [8]; two TDB methods, namely non-subsampled shearlet transform–pulse coupled neural network (NSST-PCNN) [38] and NSCT-PCLLE [7]; and five DL methods, namely zero-learning fast (Zero-LF) [39], image fusion framework based on CNN (IFCNN) [18], squeeze-and-decomposition network (SDNet) [40], enhanced medical image fusion network (EMFusion) [41], and unified unsupervised image fusion network (U2Fusion) [42]. For a fair comparison, the parameter settings of all the methods were consistent with the original text. The methods ILLF, LRD, TDB, NSST-PCNN, and NSCT-PCLLE were implemented in MATLAB 2019a, methods Zero-LF, IFCNN, SDNet, EMFusion, and U2Fusion were implemented in PyCharm 2022. All of the fusion methods were implemented on a PC with Intel® Core™; i7-5500U 2.40 GHz-CPU (2394 MHz) and 12-GB RAM.

4.2. Parameter Analysis

Different parameters determine the performance of the algorithm. In the proposed method, the parameter T in Equation (8) plays a decisive role in the performance of the algorithm, mainly because after processing the detail layers using saliency detection to obtain salient information, a decision map (SGF) was needed to extract this information to the salient detail layer. It was found using Equation (11) that there was more salient information when there was more favorable SGF to the salient information layer. According to Equation (8), the threshold value T was the key factor affecting the initial salient decision maps, influencing the SGF through Equations (9) and (10). For smaller values of T , the SGF was more favorable to the salient detail layers, and more salient information was contained in the salient detail layer. However, when T was too small, the noise in the salient detail layer could not be reduced effectively, so we needed a reasonable size of T . The selection of the parameter T in the proposed model is discussed here. We selected five sets of source images and set the variation range of T to 0.01–0.09 data (eight sets in total) because it was difficult to clearly distinguish the differences in quality of these fusion results using only subjective quality assessment. Six metrics were used to evaluate the fusion results, and the average value of the objective evaluation of the five sets of images was obtained, as shown in Figure 4. Q_{MI} has a large value when T is 0.01, Q_M , Q_P , and VIF have large values when T is 0.08, Q_{NICE} has a large value when T is 0.07, and EN has a large value when T is 0.04. Considering these indicators, we set T to 0.07.

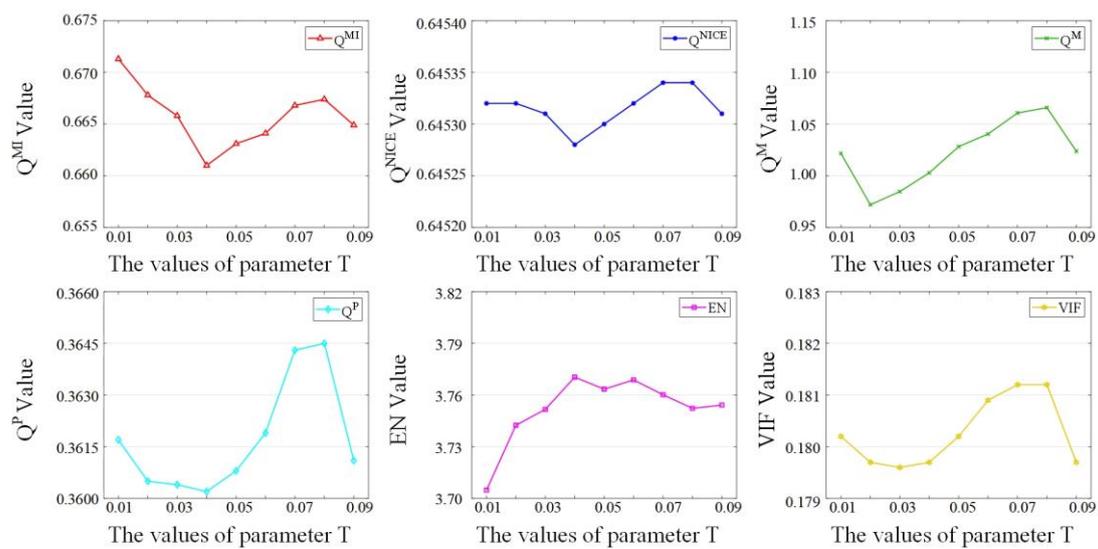


Figure 4. Fusion performance under different values of parameter T .

For the other parameters, the GRF filter was set to [filter size $r = 3$, blur degree $\varepsilon = 0.3$, iteration number $T' = 4$]. Based on previous suggestions [27], the BLF-LS was set to [$\sigma_S = 12$, $\sigma_r = 0.02$], and the window size for the SML in Equation (13) and LEN in Equation (15) were set to 3×3 .

4.3. Subjective Quality Assessment

For conciseness, we have only shown the results of three sets of images in the subjective evaluation. Figures 5–7 show the fusion results of different types of medical images obtained by different image fusion algorithms.

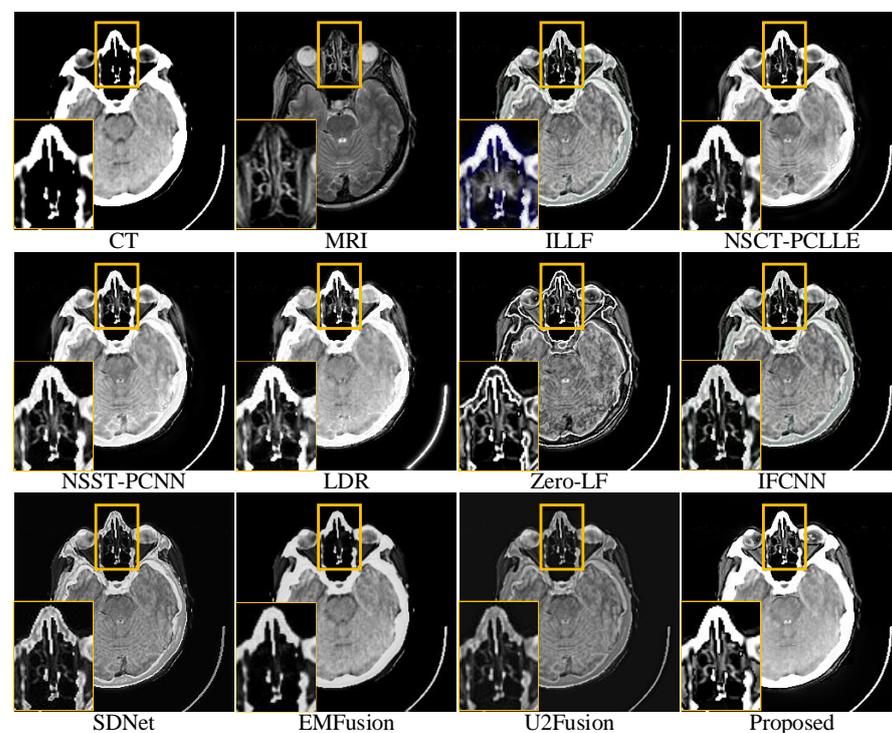


Figure 5. Comparison of performances of various methods on computed tomography magnetic resonance imaging (CT-MRI) source images. For a clear comparison, we select a same region (i.e., the yellow box) in each image and zoom in it in the bottom left corner.

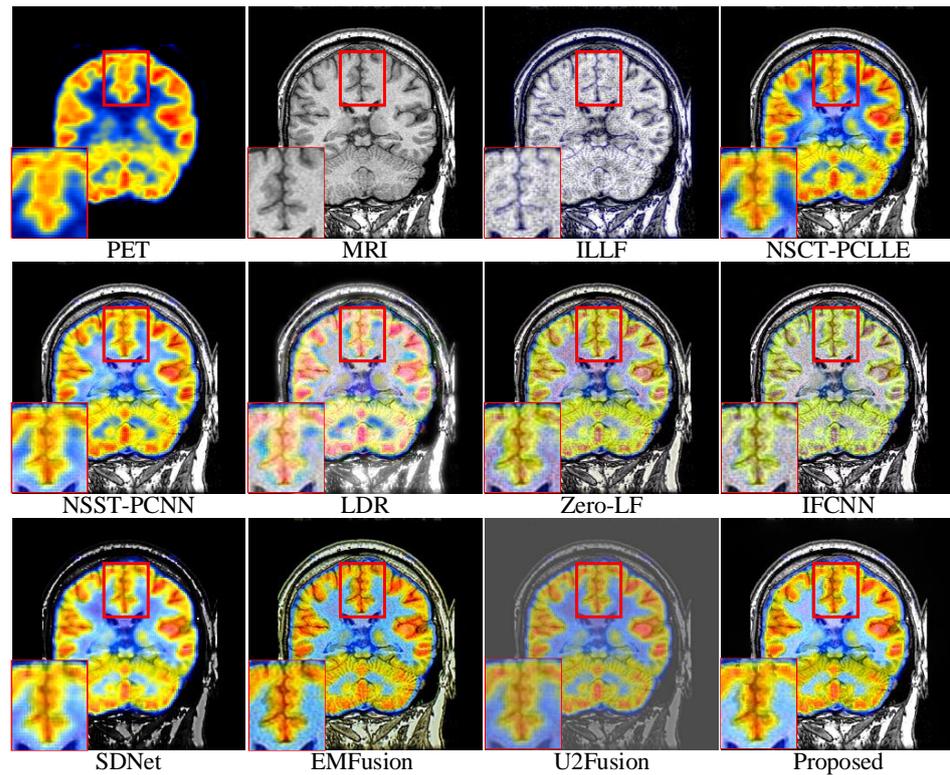


Figure 6. Comparison of performances of various methods on the positron emission tomography (PET)-MRI source images. For a clear comparison, we select a same region (i.e., the red box) in each image and zoom in it in the bottom left corner.

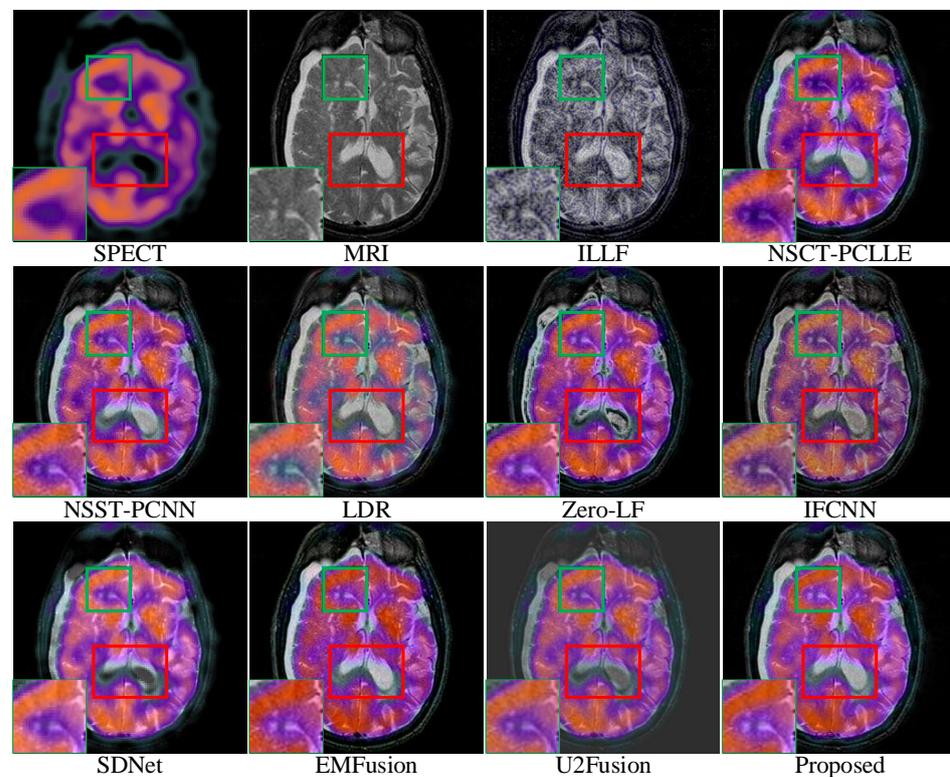


Figure 7. Comparison of performances of various methods on the single-photon emission computed tomography (SPECT)-MRI images. For a clear comparison, we select two same regions (i.e., the green and red boxes) in each image, and the green boxes are zoomed in the bottom left corner.

The fusion results of the different methods on CT-MRI medical images are shown in Figure 5. The local areas are marked by colored rectangles, which are enlarged in the lower left corners for better comparison. All the methods retained the main information and features, as shown in Figure 5; however, there were still significant differences regarding the features. ILLF showed color distortion, which led to the introduction of speckles in the fusion results. Zero-LF, IFCNN, SDNet, EMFusion, and U2Fusion could not completely retain the energy in the CT images, leading to low brightness and contrast in the fusion results. Second, NSCT-PCLLE, NSST-PCNN, LDR, SDNet, EMFusion, and U2Fusion were unable to retain the detail information in the MRI images (yellow part of the magnification area). Figure 5 shows that the proposed method outperformed the other methods in terms of the energy retention of CT-MRI medical-source images. It also successfully preserved information such as the details and structures in the source images without artifacts and brightness distortion.

Figure 6 shows a set of PET-MRI images fused by different methods. The fusion results of ILLF, LDR, Zero-LF, IFCNN, and U2Fusion show their insufficient ability in retaining the color in the PET images, which led to color distortion in the fusion results. LDR, SDNet, EMFusion, and U2Fusion performed poorly in retaining the luminance information of the MRI images; the luminance-oversaturation phenomenon occurred in the fusion results under LDR. Under SDNet, EMFusion, and U2Fusion, most of the energy from the MRI images was lost, particularly under SDNet, which led to a low overall illuminance in the images. The enlarged portion in the lower left corners shows that our method was able to retain the detailed part of the MRI image. Additionally, under our method, jagged edges can be observed in the fused image, while these edges were slightly missing under the other methods, demonstrating the superior performance of our method. Overall, Figure 6 indicates that our method could retain the structural information of the source image and outperformed the other methods in expressing intensity-based features.

In Figure 7, under ILLF, LDR, IFCNN, and EMFusion, there are deviations in color in the source SPECT image. Regarding the source image, the ILLF results show grayscale information, the LDR and IFCNN results show lighter colors, and the EMFusion results show color enhancement. NSCT-PCLLE, NSST-PCNN, Zero-LF, SDNet, and U2Fusion did not completely capture the luminance information of the MRI images, and this is represented by a small black shading in the marked red area. The fusion results of NSCT-PCLLE, NSST-PCNN, Zero-LF, SDNet, and U2Fusion show black blocks. As shown by the green enlarged area in the lower left corner of the images, ILLF, NSST-PCLLE, LDR, IFCNN, SDNet, and U2Fusion were not completely capable of retaining the details in the source image, while our method was able to retain them well. Figure 7 shows that the images fused under our proposed method are more informative, clearer, and have a higher contrast than those under the existing methods.

4.4. Objective Quality Assessment

Tables 2–4 show the objective evaluations of the different methods. Table 2 shows the objective evaluation results of the CT-MRI images. Our proposed method ranked first for the indicators Q_{MI} , Q_{NCIE} , Q_M , Q_P and VIF . This shows that our method obtained good results regarding the amount of information it could transfer from the source image to the fused image, the degree of nonlinear correlation, edge information, phase consistency, and information fidelity. Although it did not rank high for EN, the difference between its value and the highest value was small; therefore, we concluded that the proposed method was able to produce good results for the CT-MRI images under an objective evaluation assessment.

Table 2. Comparison of performance of proposed fusion method with those of nine existing methods on CT-MRI images; the best results are shown in bold.

| | Methods | Q_{MI} | Q_{NCIE} | Q_M | Q_P | EN | VIF |
|------------------------------------------------------------|---------------|---------------|------------|---------------|---------------|---------------|---------------|
| Objective evaluation of different fused images in Figure 5 | ILLF | 0.5554 | 0.8054 | 0.1583 | 0.2049 | 5.7517 | 0.1330 |
| | NSST-PCNN | 0.6502 | 0.8050 | 0.6791 | 0.3444(3) | 4.7473 | 0.2048 |
| | NSCT-PCLLE | 0.6232 | 0.8047 | 0.6962(3) | 0.2792 | 4.6168 | 0.2194(3) |
| | LRD | 0.7302 | 0.8056(3) | 0.2218 | 0.3292 | 4.5160 | 0.1702 |
| | Zero-LF | 0.7918(3) | 0.8056 | 1.5481 | 0.1172 | 4.0112 | 0.0874 |
| | IFCNN | 0.6690 | 0.8047 | 0.1469 | 0.2907 | 4.2017 | 0.1669 |
| | SDNet | 0.6513 | 0.8051 | 0.1084 | 0.2840 | 4.8154(2) | 0.1712 |
| | EMFusion | 0.8345(2) | 0.8063(2) | 0.1309 | 0.5829 | 4.2328 | 0.3247 |
| | U2Fusion | 0.6059 | 0.8045 | 0.0849 | 0.2925 | 4.4836 | 0.1862 |
| Proposed | 0.8443 | 0.8070 | 1.1038(2) | 0.4424(2) | 4.7689(3) | 0.2293(2) | |
| Average evaluation mean of 100 groups of images | ILLF | 0.7265 | 0.8048 | 0.1635 | 0.2890 | 4.1261 | 0.2087 |
| | NSST-PCNN | 0.7532 | 0.8049 | 0.6157 | 0.2896 | 3.9475 | 0.2488 |
| | NSCT-PCLLE | 0.7319 | 0.8048 | 0.7197(3) | 0.2726 | 3.9428 | 0.2736(3) |
| | LRD | 0.7861 | 0.8052 | 0.3345 | 0.2548 | 3.8813 | 0.1994 |
| | Zero-LF | 0.8749(3) | 0.8057(3) | 1.6937 | 0.1465 | 3.6424 | 0.0993 |
| | IFCNN | 0.7512 | 0.8047 | 0.1797 | 0.2904 | 3.7007 | 0.1915 |
| | SDNet | 0.7578 | 0.8052 | 0.1282 | 0.2713 | 4.1732(2) | 0.1928 |
| | EMFusion | 0.8933(2) | 0.8057(2) | 0.1510 | 0.4295 | 3.6298 | 0.3504 |
| | U2Fusion | 0.6976 | 0.8045 | 0.1138 | 0.3130 | 3.9285 | 0.2249 |
| Proposed | 0.8935 | 0.8064 | 1.1555 (2) | 0.3820 (2) | 4.0829(3) | 0.2903(2) | |

Table 3. Comparison of performance of proposed fusion method with those of nine methods on PET-MRI images; the best results are shown in bold.

| | Methods | Q_{MI} | Q_{NCIE} | Q_M | Q_P | EN | VIF |
|------------------------------------------------------------|---------------|---------------|---------------|-----------|---------------|---------------|---------------|
| Objective evaluation of performance for images in Figure 6 | ILLF | 0.3732 | 0.8035 | 0.0547 | 0.0167 | 4.6800 | 0.0095 |
| | NSST-PCNN | 0.6509(2) | 0.8078(2) | 1.4767(2) | 0.4576 | 5.8536 | 0.2977 |
| | NSCT-PCLLE | 0.6040 | 0.8069 | 1.3982(3) | 0.4276 | 5.9460(3) | 0.2982(3) |
| | LRD | 0.5225 | 0.8060 | 0.2486 | 0.3397 | 6.4027 | 0.1808 |
| | Zero-LF | 0.6229(3) | 0.8074(3) | 0.3195 | 0.4543 | 5.6813 | 0.2352 |
| | IFCNN | 0.6073 | 0.8072 | 0.2155 | 0.4949(2) | 5.8007 | 0.2461 |
| | SDNet | 0.6062 | 0.8057 | 0.1035 | 0.2804 | 5.0403 | 0.1451 |
| | EMFusion | 0.5906 | 0.8072 | 0.2717 | 0.5669 | 5.8329 | 0.2688 |
| | U2Fusion | 0.5653 | 0.8055 | 0.0879 | 0.3689 | 5.0805 | 0.3071(2) |
| Proposed | 0.8319 | 0.8129 | 1.9020 | 0.4902(3) | 6.2604(2) | 0.3170 | |
| Average evaluation mean of 100 groups of images | ILLF | 0.3448 | 0.8035 | 0.1230 | 0.0956 | 5.1814 | 0.0503 |
| | NSST-PCNN | 0.7019(2) | 0.8088(3) | 1.6547(2) | 0.5141 | 5.7223 | 0.3330 |
| | NSCT-PCLLE | 0.6635(3) | 0.8081 | 1.6272(3) | 0.4939 | 5.7743 | 0.3369 |
| | LRD | 0.5533 | 0.8064 | 0.3332 | 0.3583 | 6.2651 | 0.1920 |
| | Zero-LF | 0.6558 | 0.8077 | 0.6523 | 0.4809 | 5.6436 | 0.2434 |
| | IFCNN | 0.6187 | 0.8072 | 0.3018 | 0.5403(2) | 5.6759 | 0.2572 |
| | SDNet | 0.6235 | 0.8057 | 0.1288 | 0.2815 | 4.9340 | 0.1528 |
| | EMFusion | 0.6160 | 0.8242 | 0.3305 | 0.6479 | 5.8601(3) | 0.2819 |
| | U2Fusion | 0.5753 | 0.8055 | 0.1161 | 0.4202 | 4.9477 | 0.3341(3) |
| Proposed | 0.8836 | 0.8134(2) | 2.0628 | 0.5148(3) | 5.9378(2) | 0.3342(2) | |

Table 4. Comparison of performance of proposed fusion method with those of nine methods on MRI-SPECT images; the best results are shown in bold.

| | Methods | Q_{MI} | Q_{NCIE} | Q_M | Q_P | EN | VIF |
|------------------------------------------------------------|---------------|---------------|---------------|-----------|---------------|---------------|---------------|
| Objective evaluation of different fused images in Figure 7 | ILLF | 0.5339 | 0.8063 | 0.2813 | 0.2664 | 5.8953 | 0.1818 |
| | NSST-PCNN | 0.6030 | 0.8073 | 1.0805(3) | 0.3283 | 6.0048(2) | 0.3171 |
| | NSCT-PCLLE | 0.5940 | 0.8072 | 1.0262 | 0.3452(3) | 5.9899(3) | 0.3561(2) |
| | LRD | 0.6423 | 0.8081 | 0.7759 | 0.3395 | 6.1174 | 0.3051 |
| | Zero-LF | 0.6972(3) | 0.8087(3) | 1.3326(2) | 0.3233 | 5.7236 | 0.2191 |
| | IFCNN | 0.5995 | 0.8069 | 0.5536 | 0.3206 | 5.6857 | 0.2846 |
| | SDNet | 0.6525 | 0.8074 | 0.2431 | 0.2395 | 5.6691 | 0.2214 |
| | EMFusion | 0.7402(2) | 0.8102(2) | 0.6409 | 0.5500 | 5.6414 | 0.3927 |
| | U2Fusion | 0.5883 | 0.8063 | 0.2151 | 0.2791 | 5.2268 | 0.3342(3) |
| Proposed | 0.8526 | 0.8132 | 1.4879 | 0.4283(2) | 5.9386(4) | 0.3226(4) | |
| Average evaluation mean of 100 groups of images | ILLF | 0.5761 | 0.7898 | 0.2838 | 0.3237 | 5.1492 | 0.1933 |
| | NSST-PCNN | 0.6856 | 0.7914 | 1.1394(3) | 0.4271 | 5.1601(3) | 0.3365 |
| | NSCT-PCLLE | 0.6908 | 0.7915 | 1.0820 | 0.4422(3) | 5.1117 | 0.3657(3) |
| | LRD | 0.6687 | 0.7912 | 0.6880 | 0.4002 | 5.3850 | 0.2965 |
| | Zero-LF | 0.7677(2) | 0.7921(3) | 1.3337(2) | 0.3971 | 4.8929 | 0.2486 |
| | IFCNN | 0.6542 | 0.7903 | 0.5443 | 0.4083 | 4.8430 | 0.2917 |
| | SDNet | 0.7127 | 0.7904 | 0.2070 | 0.2964 | 4.5649 | 0.2427 |
| | EMFusion | 0.7638(3) | 0.7927(2) | 0.5768 | 0.6211 | 4.8813 | 0.3799 |
| | U2Fusion | 0.6307 | 0.7896 | 0.1951 | 0.3483 | 4.4321 | 0.3703(2) |
| Proposed | 0.8736 | 0.7955 | 1.4704 | 0.4701(2) | 5.2540(2) | 0.3194(5) | |

Tables 3 and 4 show the objective evaluation results of the nine methods on the PET-MRI and MRI-SPECT color images, respectively. Our method did not rank first in certain metrics, but its overall ranking was at the top. It also achieved good results for the PET-MRI and MRI-SPECT images under the objective evaluation assessment.

Based on the above subjective visual evaluation and objective metrics analysis, we concluded that the fusion performance of our method was the highest of all methods. This was mainly due to the good decomposition of the details and basis values of the images from using the BLF-LS, the effective preservation of the significant structure and edge information of the source image in the fused image using saliency detection, and the processing of the weight map using the RGF, which makes full use of the strong correlation between the neighboring pixels.

4.5. Discussion on Time Efficiency

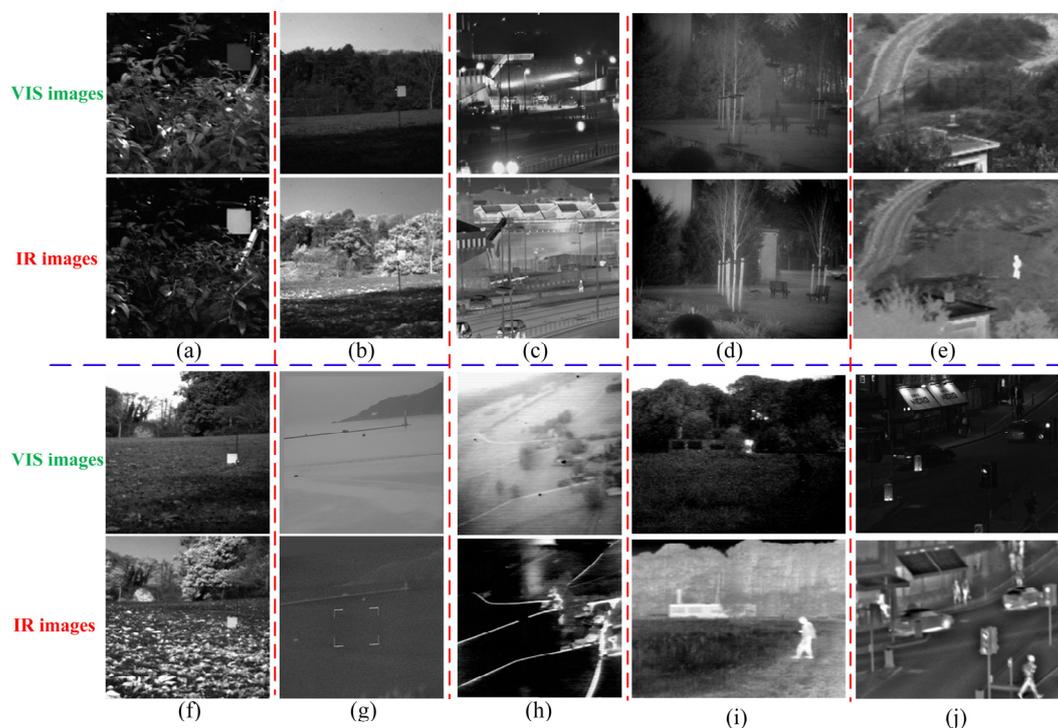
In this section, we compared the time efficiency of the proposed method with those of the other nine methods on grayscale images. As shown by the results in Table 5, the Zero-LF, IFCNN, SDNet, EMFusion, and U2Fusion DL methods trained the models in advance, allowing them to process the images quickly. The ILLF method had the longest running time because the ILLF filter was not as fast as the other multi-scale tools and computed the decomposition of the image at different scales. The LRD algorithm took too long in the gradient-domain image enhancement owing to its over-reliance on the fitting function. NSST-PCNN also required more time than our method because of the PCNN iterations that were involved. Although the proposed method was not the fastest, considering its high performance, it was still effective. Moreover, we believe that if we fully optimize the code behind the working of our method and convert it to increase its efficiency using tools such as the graphical processing unit (GPU) and C++, the time required to execute our method will be significantly shorter, enabling the method to satisfy the requirements of more applications.

Table 5. Running times of different methods for fusing two source images of 256×256 pixels each.

| Methods | ILLF | NSST-PCNN | NSCT-PCLLE | LRD | Zero-LF |
|---------|--------|-----------|------------|----------|----------|
| Time | 161.51 | 10.01 | 3.26 | 126.64 | 2.45 |
| Methods | IFCNN | SDNet | EMFusion | U2Fusion | Proposed |
| Time | 0.21 | 0.16 | 0.57 | 0.36 | 4.24 |

4.6. Extension to Infrared (IR) and Visible (VIS) Image Fusion

To justify the ability of our proposed method to generalize, we tested the fusion ability of our method on ten sets of IR-VIS images (shown in Figure 8). Six advanced fusion methods for IR-VIS images were selected for comparison: visual saliency map and weighted least square optimization (VSWL) [43], Gaussian curvature filtering (GCF) [44], IFCNN [18], SDNet [40], U2Fusion [42], and SwinFusion [45].

**Figure 8.** Source images. (a–j) are the ten pairs of VIS/IF images.

As shown in Figure 9, although all seven methods could retain the energy in the IR image and the details in the VIS image, differences still existed. The red box with the pedestrian in the lower right corner of Figure 9 shows that although all seven methods obtained the detailed and contour information of the person in the source image, the overall brightness, specifically under the SDNet, U2Fusion, and SwinFusion methods, was low. In the fusion results of the proposed methods, the person's edges did not appear as black shadows owing to the smoothing of the edges of the significant target using RGF. Second, regarding the poster board framed in green in the lower-left corner of the image in Figure 9, the VSWL, GCF, IFCNN, SDNet, U2Fusion, and SwinFusion methods compared the light map without retaining the overall luminance information of the light sign. The above analysis proves that our algorithm had the best detail retention and color fidelity and was more consistent with the subjective vision for processing object edges in an image.

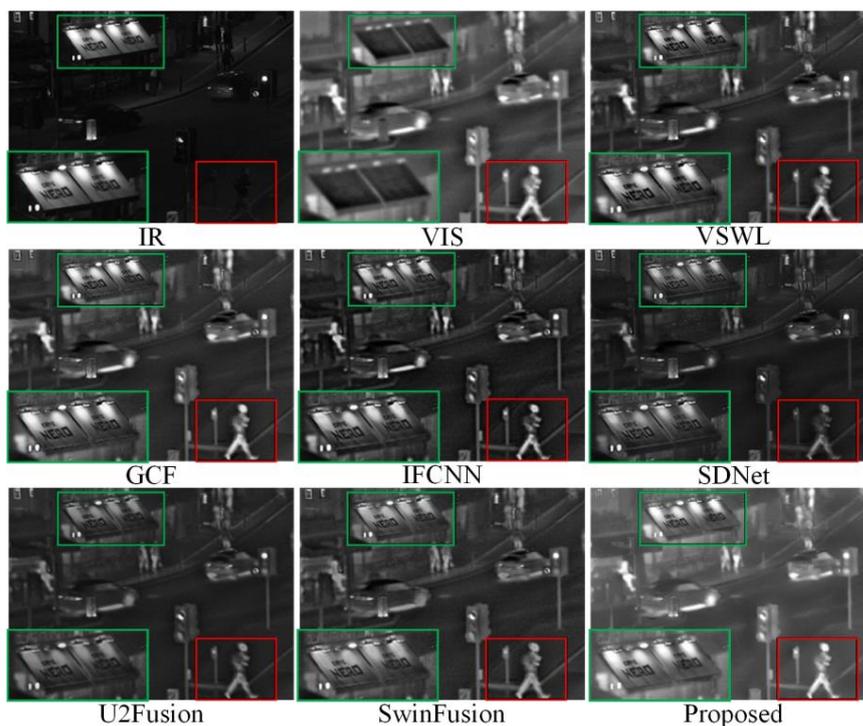


Figure 9. Comparison of performance of various methods on source images of Queens Road, Bristol. For a clear comparison, we select two same regions (i.e., the green and red boxes) in each image, and the green boxes are zoomed in the bottom left corner.

Figure 10 shows the objective evaluation results for the “Queens Road, Bristol” image, and the average value of the objective evaluation of the 10 IR-VIS images (the 10 sets of images shown in Figure 8). The horizontal and vertical coordinates represent the different methods and the values of different evaluation metrics, respectively. The red line shows the objective evaluation of the different methods on the image “Queens Road, Bristol”, and the blue line shows the average value of the objective evaluation of the different methods in Figure 8. Regarding the objective evaluation assessment, our method ranked first in Q_{MI} , Q_{NCIE} , Q_M , EN , and the average VIF index for the 10 IR-VIS images. Although the result for Q_P was not the highest, its difference from the best value was not pronounced. Thus, the validity of the proposed method in terms of objective assessment is confirmed. The above evaluation shows that our method can be effectively extended to IR-VIS image fusion.

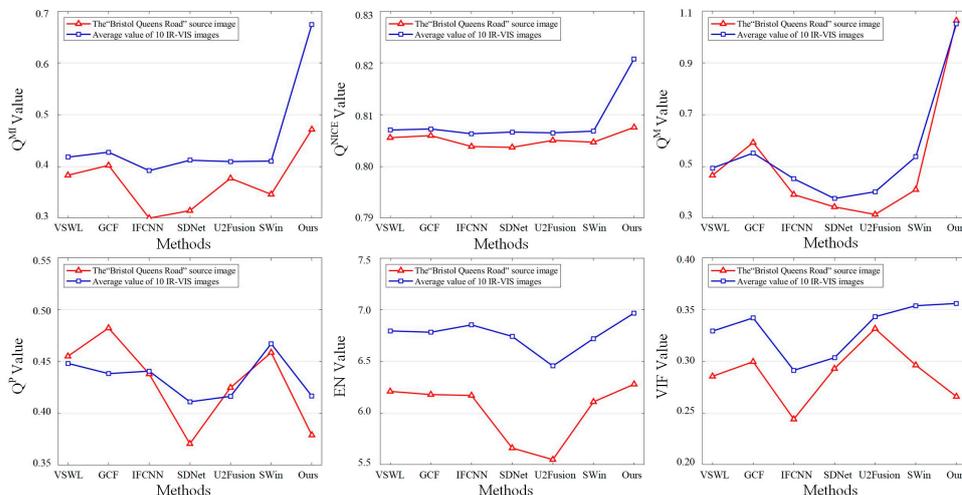


Figure 10. Objective evaluation of 7 methods on IR-VIS image fusion.

5. Conclusions

In this study, we proposed a multi-sensor medical-image fusion method based on the BLF-LS and DSC. First, the image decomposition of the base layer using the BLF-LS was simple and effective as it potentially captured large changes in intensity. The detail layer preserved details such as the structure, texture, and edges of the original image efficiently. The DCS model effectively detected the salient information, and the GRF made full use of the strong correlation between the neighboring pixels for weight optimization, allowing the fused detail layer to effectively retain the salient structure and edge information in the source image. Finally, the base-layer fusion rules based on the MLEN effectively preserved the energy information of the source images.

The fusion results of different methods on CT-MRI, PET-MRI and MRI-SPECT images were demonstrated. The experimental results showed the advantages of the proposed method in both subjective visual and objective quantitative evaluations. Compared to the nine state-of-the-art methods used in the study, the proposed medical image fusion algorithm can provide fusion images with clearer edge details, complete salient information, more brightness, and superior colors. Additionally, this method is also applicable to IR-VIS image fusion. However, the proposed fusion method is easily affected by noise because the inputs are in alignment pairs. In the future, we will work on solving the effect of noise on images, thus bridging the gap between medical image fusion and actual clinical applications.

Author Contributions: Methodology, J.L. and X.L.; software, J.L.; writing—original draft preparation, J.L.; writing—review and editing, X.W. and X.L.; supervision, D.H.; data curation, P.Y.; visualization, L.Y.; funding acquisition, D.H. and X.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (Grant Nos. 62201149, 62271148), the Ji Hua Laboratory of Guangdong province, China (Grant No. X200051UZ200), the Research and Development Projects in Key Areas of Guangdong Province (2020B1111040001), and the Research Fund of Guangdong–Hong Kong–Macao Joint Laboratory for Intelligent Micro-Nano Optoelectronic Technology (No. 2020B1212030010).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available upon request from the authors.

Acknowledgments: The authors would like to thank the editors and anonymous reviewers for their constructive and valuable comments and suggestions.

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

References

1. Goyal, B.; Dogra, A.; Khoond, R.; Gupta, A.; Anand, R. Infrared and Visible Image Fusion for Concealed Weapon Detection using Transform and Spatial Domain Filters. In Proceedings of the 2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Noida, India, 3–4 September 2021; pp. 1–4.
2. Hermessi, H.; Mourali, O.; Zagrouba, E. Multimodal medical image fusion review: Theoretical background and recent advances. *Signal Process.* **2021**, *183*, 108036. [[CrossRef](#)]
3. Li, S.T.; Kang, X.D.; Fang, L.Y.; Hu, J.W.; Yin, H.T. Pixel-level image fusion: A survey of the state of the art. *Inf. Fusion* **2017**, *33*, 100–112. [[CrossRef](#)]
4. Zhang, Q.; Liu, Y.; Blum, R.S.; Han, J.G.; Tao, D.C. Sparse representation based multi-sensor image fusion for multi-focus and multi-modality images: A review. *Inf. Fusion* **2018**, *40*, 57–75. [[CrossRef](#)]
5. Zhang, H.; Xu, H.; Tian, X.; Jiang, J.J.; Ma, J.Y. Image fusion meets deep learning: A survey and perspective. *Inf. Fusion* **2021**, *76*, 323–336. [[CrossRef](#)]
6. Li, X.S.; Zhou, F.Q.; Tan, H.S.; Chen, Y.Z.; Zuo, W.X. Multi-focus image fusion based on nonsubsampling contourlet transform and residual removal. *Signal Process.* **2021**, *184*, 108062. [[CrossRef](#)]
7. Zhu, Z.Q.; Zheng, M.G.; Qi, G.Q.; Wang, D.; Xiang, Y. A Phase Congruency and Local Laplacian Energy Based Multi-Modality Medical Image Fusion Method in NSCT Domain. *IEEE Access* **2019**, *7*, 20811–20824. [[CrossRef](#)]

8. Li, X.; Guo, X.; Han, P.; Wang, X.; Li, H.; Luo, T. Laplacian Redecomposition for Multimodal Medical Image Fusion. *IEEE Trans. Instrum. Meas.* **2020**, *69*, 6880–6890. [[CrossRef](#)]
9. Khan, H.; Sharif, M.; Bibi, N.; Usman, M.; Haider, S.A.; Zainab, S.; Shah, J.H.; Bashir, Y.; Muhammad, N. Localization of radiance transformation for image dehazing in wavelet domain. *Neurocomputing* **2020**, *381*, 141–151. [[CrossRef](#)]
10. Juneja, S.; Anand, R. *Contrast Enhancement of an Image by DWT-SVD and DCT-SVD*; Springer: Singapore, 2018; pp. 595–603.
11. Li, X.S.; Zhou, F.Q.; Tan, H.S. Joint image fusion and denoising via three-layer decomposition and sparse representation. *Knowl.-Based Syst.* **2021**, *224*, 107087. [[CrossRef](#)]
12. Li, S.T.; Yin, H.T.; Fang, L.Y. Group-Sparse Representation With Dictionary Learning for Medical Image Denoising and Fusion. *IEEE Trans. Biomed. Eng.* **2012**, *59*, 3450–3459. [[CrossRef](#)]
13. Zhang, Q.; Levine, M.D. Robust Multi-Focus Image Fusion Using Multi-Task Sparse Representation and Spatial Context. *IEEE Trans. Image Process.* **2016**, *25*, 2045–2058. [[CrossRef](#)] [[PubMed](#)]
14. Wang, J.; Peng, J.Y.; Feng, X.Y.; He, G.Q.; Fan, J.P. Fusion method for infrared and visible images by using non-negative sparse representation. *Infrared Phys. Technol.* **2014**, *67*, 477–489. [[CrossRef](#)]
15. Gu, S.; Meng, D.; Zuo, W.; Zhang, L. Joint Convolutional Analysis and Synthesis Sparse Representation for Single Image Layer Separation. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 1717–1725.
16. Jie, Y.; Zhou, F.; Tan, H.; Wang, G.; Cheng, X.; Li, X. Tri-modal medical image fusion based on adaptive energy choosing scheme and sparse representation. *Measurement* **2022**, *204*, 112038. [[CrossRef](#)]
17. Muhammad, N.; Bibi, N.; Jahangir, A.; Mahmood, Z. Image denoising with norm weighted fusion estimators. *Pattern Anal. Appl.* **2018**, *21*, 1013–1022. [[CrossRef](#)]
18. Zhang, Y.; Liu, Y.; Sun, P.; Yan, H.; Zhao, X.L.; Zhang, L. IFCNN: A general image fusion framework based on convolutional neural network. *Inf. Fusion* **2020**, *54*, 99–118. [[CrossRef](#)]
19. Ma, J.Y.; Yu, W.; Liang, P.W.; Li, C.; Jiang, J.J. FusionGAN: A generative adversarial network for infrared and visible image fusion. *Inf. Fusion* **2019**, *48*, 11–26. [[CrossRef](#)]
20. Luo, X.; Gao, Y.; Wang, A.; Zhang, Z.; Wu, X.J. IFSepR: A general framework for image fusion based on separate representation learning. *IEEE Trans. Multimed.* **2021**, *25*, 608–623. [[CrossRef](#)]
21. Zhu, Z.Q.; He, X.Y.; Qi, G.Q.; Li, Y.Y.; Cong, B.S.; Liu, Y. Brain tumor segmentation based on the fusion of deep semantics and edge information in multimodal MRI. *Inf. Fusion* **2023**, *91*, 376–387. [[CrossRef](#)]
22. Liu, Y.; Chen, X.; Wang, Z.F.; Wang, Z.J.; Ward, R.K.; Wang, X.S. Deep learning for pixel-level image fusion: Recent advances and future prospects. *Inf. Fusion* **2018**, *42*, 158–173. [[CrossRef](#)]
23. Mo, Y.; Kang, X.D.; Duan, P.H.; Sun, B.; Li, S.T. Attribute filter based infrared and visible image fusion. *Inf. Fusion* **2021**, *75*, 41–54. [[CrossRef](#)]
24. Wang, G.F.; Li, W.S.; Du, J.; Xiao, B.; Gao, X.B. Medical Image Fusion and Denoising Algorithm Based on a Decomposition Model of Hybrid Variation-Sparse Representation. *IEEE J. Biomed. Health Inform.* **2022**, *26*, 5584–5595. [[CrossRef](#)] [[PubMed](#)]
25. Xu, G.X.; Deng, X.X.; Zhou, X.K.; Pedersen, M.; Cimmino, L.; Wang, H. FCFusion: Fractal Componentwise Modeling With Group Sparsity for Medical Image Fusion. *IEEE Trans. Ind. Inform.* **2022**, *18*, 9141–9150. [[CrossRef](#)]
26. Li, X.S.; Zhou, F.Q.; Tan, H.S.; Zhang, W.N.; Zhao, C.Y. Multimodal medical image fusion based on joint bilateral filter and local gradient energy. *Inf. Sci.* **2021**, *569*, 302–325. [[CrossRef](#)]
27. Liu, W.; Zhang, P.P.; Chen, X.G.; Shen, C.H.; Huang, X.L.; Yang, J. Embedding Bilateral Filter in Least Squares for Efficient Edge-Preserving Image Smoothing. *IEEE Trans. Circuits Syst. Video Technol.* **2020**, *30*, 23–35. [[CrossRef](#)]
28. Wu, X.; Ma, X.; Zhang, J.; Wang, A.; Jin, Z. Salient Object Detection Via Deformed Smoothness Constraint. In Proceedings of the 2018 25th IEEE International Conference on Image Processing (ICIP), Athens, Greece, 7–10 October 2018; pp. 2815–2819.
29. Huang, W.; Jing, Z. Evaluation of focus measures in multi-focus image fusion. *Pattern Recognit. Lett.* **2007**, *28*, 493–500. [[CrossRef](#)]
30. Qu, G.H.; Zhang, D.L.; Yan, P.F. Information measure for performance of image fusion. *Electron. Lett.* **2002**, *38*, 313–315. [[CrossRef](#)]
31. Wang, P.W.; Liu, B. A Novel Image Fusion Metric Based on Multi-Scale Analysis. In Proceedings of the 9th International Conference on Signal Processing, Beijing, China, 26–29 October 2008; pp. 965–968.
32. Wang, Q.; Shen, Y.; Zhang, J.Q. A nonlinear correlation measure for multivariable data set. *Phys. D-Nonlinear Phenom.* **2005**, *200*, 287–295. [[CrossRef](#)]
33. Zhao, J.Y.; Laganieri, R.; Liu, Z. Performance assessment of combinative pixel-level image fusion based on an absolute feature measurement. *Int. J. Innov. Comput. Inf. Control* **2007**, *3*, 1433–1447.
34. Ma, J.Y.; Ma, Y.; Li, C. Infrared and visible image fusion methods and applications: A survey. *Inf. Fusion* **2019**, *45*, 153–178. [[CrossRef](#)]
35. Sheikh, H.R.; Bovik, A.C. Image information and visual quality. *IEEE Trans. Image Process.* **2006**, *15*, 430–444. [[CrossRef](#)]
36. Liu, Z.; Blasch, E.; Xue, Z.Y.; Zhao, J.Y.; Laganieri, R.; Wu, W. Objective Assessment of Multiresolution Image Fusion Algorithms for Context Enhancement in Night Vision: A Comparative Study. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 94–109. [[CrossRef](#)] [[PubMed](#)]
37. Du, J.; Li, W.S.; Xiao, B. Anatomical-Functional Image Fusion by Information of Interest in Local Laplacian Filtering Domain. *IEEE Trans. Image Process.* **2017**, *26*, 5855–5866. [[CrossRef](#)] [[PubMed](#)]

38. Yin, M.; Liu, X.N.; Liu, Y.; Chen, X. Medical Image Fusion With Parameter-Adaptive Pulse Coupled Neural Network in Nonsampled Shearlet Transform Domain. *IEEE Trans. Instrum. Meas.* **2019**, *68*, 49–64. [[CrossRef](#)]
39. Lahoud, F.; Süssstrunk, S. Zero-Learning Fast Medical Image Fusion. In Proceedings of the 2019 22th International Conference on Information Fusion (FUSION), Ottawa, ON, Canada, 2–5 July 2019; pp. 1–8.
40. Zhang, H.; Ma, J.Y. SDNet: A Versatile Squeeze-and-Decomposition Network for Real-Time Image Fusion. *Int. J. Comput. Vis.* **2021**, *129*, 2761–2785. [[CrossRef](#)]
41. Xu, H.; Ma, J.Y. EMFusion: An unsupervised enhanced medical image fusion network. *Inf. Fusion* **2021**, *76*, 177–186. [[CrossRef](#)]
42. Xu, H.; Ma, J.Y.; Jiang, J.J.; Guo, X.J.; Ling, H.B. U2Fusion: A Unified Unsupervised Image Fusion Network. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *44*, 502–518. [[CrossRef](#)]
43. Ma, J.L.; Zhou, Z.Q.; Wang, B.; Zong, H. Infrared and visible image fusion based on visual saliency map and weighted least square optimization. *Infrared Phys. Technol.* **2017**, *82*, 8–17. [[CrossRef](#)]
44. Tan, W.; Zhou, H.X.; Song, J.L.Q.; Li, H.; Yu, Y.; Du, J. Infrared and visible image perceptive fusion through multi-level Gaussian curvature filtering image decomposition. *Appl. Opt.* **2019**, *58*, 3064–3073. [[CrossRef](#)]
45. Ma, J.Y.; Tang, L.F.; Fan, F.; Huang, J.; Mei, X.G.; Ma, Y. SwinFusion: Cross-domain Long-range Learning for General Image Fusion via Swin Transformer. *IEEE-CAA J. Autom. Sin.* **2022**, *9*, 1200–1217. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.