electronics

MDPI

Article

CT and MRI Image Fusion via Coupled Feature-Learning GAN

Qingyu Mao [1], Wenzhe Zhai [2], Xiang Lei [3], Zenghui Wang [2] and Yongsheng Liang [1,4,*]

[1] College of Electronics and Information Engineering, Shenzhen University, Shenzhen 518060, China; qingyu.mao@outlook.com
[2] School of Electrical and Electronic Engineering, Shandong University of Technology, Zibo 255000, China; wenzhezhai@163.com (W.Z.); sdut_zenghuiwang@163.com (Z.W.)
[3] Zhiyang Innovation Co., Ltd., Jinan 250101, China; shadyatscu@gmail.com
[4] College of Big Data and Internet, Shenzhen Technology University, Shenzhen 518118, China
* Correspondence: liangys@szu.edu.cn

Abstract: The fusion of multimodal medical images, particularly CT and MRI, is driven by the need to enhance the diagnostic process by providing clinicians with a single, comprehensive image that encapsulates all necessary details. Existing fusion methods often exhibit a bias towards features from one of the source images, making it challenging to simultaneously preserve both structural information and textural details. Designing an effective fusion method that can preserve more discriminative information is therefore crucial. In this work, we propose a Coupled Feature-Learning GAN (CFGAN) to fuse the multimodal medical images into a single informative image. The proposed method establishes an adversarial game between the discriminators and a couple of generators. First, the coupled generators are trained to generate two real-like fused images, which are then used to deceive the two coupled discriminators. Subsequently, the two discriminators are devised to minimize the structural distance to ensure the abundant information in the original source images is well-maintained in the fused image. We further empower the generators to be robust under various scales by constructing a discriminative feature extraction (DFE) block with different dilation rates. Moreover, we introduce a cross-dimension interaction attention (CIA) block to refine the feature representations. The qualitative and quantitative experiments on common benchmarks demonstrate the competitive performance of the CFGAN compared to other state-of-the-art methods.

Keywords: image fusion; CT/MRI image; generative adversarial network; coupled network

Citation: Mao, Q.; Zhai, W.; Lei, X.; Wang, Z.; Liang, Y. CT and MRI Image Fusion via Coupled Feature-Learning GAN. *Electronics* **2024**, *13*, 3491. https://doi.org/10.3390/electronics13173491

Academic Editor: Luca Mesin

Received: 10 July 2024
Revised: 28 August 2024
Accepted: 30 August 2024
Published: 3 September 2024

1. Introduction

Medical images have been widely employed in healthcare systems as these have notably facilitated the development of many medical applications, e.g., surgical navigation, clinical diagnosis, and radiation surgery [1,2]. Therein, computed tomography (CT) and magnetic resonance imaging (MRI) are two principal medical images. CT images provide precise locations of the dense structures, such as skeletal tissues, while MRI images are better at reflecting the soft tissue details, e.g., blood vessels [3]. However, relying on a single modality often proves insufficient in providing comprehensive diagnostic information [4]. CT and MRI image fusion offers a promising solution to this limitation by harnessing the strengths of both imaging techniques. Through the fusion process, the complementary information from CT and MRI images can be effectively integrated, resulting in a more informative and comprehensive fused image. This synergistic combination of structural and soft tissue details provides clinicians with a powerful tool to enhance diagnostic accuracy and support precise treatment planning and guidance [5].

Various schemes have been exploited for image fusion in the literature, including three main categories: traditional methods, CNN-based methods, and GAN-based methods [6–9]. Traditional methods [10–12] are usually time-consuming due to the complex fusion strategies that are designed manually. Recently, CNN-based methods [13–19] have been proposed

for image fusion, owing to their superior ability to extract high-level features and generate high-quality fused images. However, the absence of ground-truth fused images impedes the direct optimization of CNNs using conventional supervised learning techniques. The generative adversarial network (GAN) [20] has been widely employed to generate images with favorable visual effects without the need for ground truth of the fusion image. Nonetheless, the fusion results of some GAN-based approaches tend to overemphasize one source image while neglecting the other, thereby resulting in the loss of valuable details [21,22]. Zhou et al. [5] indicate that this is due to the instability inherent in the single adversarial learning process. Ma et al. [21] built the dual-discriminator conditional GAN (DDcGAN), which utilized a generator and dual-discriminators to establish a generative adversarial relationship and expects the fused image to retain the most crucial feature information from various source images. Yang et al. [23] proposed a structure similar to DDcGAN, utilizing image differences as inputs for two discriminators, while simultaneously enhancing the ability of both the generator and the discriminators. Although these methods have achieved good results in terms of visual perception, the instability inherent in the single adversarial learning process still results in information loss or texture blurring.

In this paper, we propose a Coupled Feature-Learning GAN (CFGAN) model to fuse multimodal medical images with rich information. The proposed CFGAN is expressed as a specific adversarial process within a coupled neural network, comprising two generators and two discriminators, which guarantee that the fused image simultaneously retains significant information in CT and MRI images. Specifically, the coupled generators extract details to fuse meaningful information by sharing the same high-level information and utilizing the diverse underlying details. We embedded the discriminative feature extraction (DFE) block and the cross-dimension interaction attention (CIA) block in the generators to enable generators to preserve their robustness against various scales. The DFE block employs three dilated convolutional filters to enlarge scale diversity and receptive fields, while the CIA block extracts salient information from the feature tensor across the dimensions. In addition, we employ pre-fused images as guidance for coupled generators during the training phase. The coupled discriminators pull each other on the distribution of the generated data attained by the generators so that the fused image saves the most prominent features from both CT and MRI images. The proposed CFGAN is an end-to-end model without requiring any pre-defined fusion rules or ground truth fused images. All in all, the contributions of the paper are as follows:

1. We propose an end-to-end deep learning-based fusion model termed Coupled Feature-Learning GAN (CFGAN) for preserving the locational information of dense structures, as well as soft tissue details in multi-source images.
2. We introduce the discriminative feature extraction (DFE) block with various dilation rates to improve the robustness of generators at diverse scales.
3. We design a cross-dimension interaction attention (CIA) block for the coupled generators, integrating the salient information of cross-dimensional features to refine the feature representations.

The remainder of this paper is structured as follows. Section 2 presents the relevant work in image fusion. Section 3 introduces the details of CFGAN. Comparative experiments are conducted in Section 4. The conclusion is derived in Section 5.

2. Related Work

2.1. Traditional-Based Methods

Traditional fusion methods can be classified into two types: transform domain-based and spatial domain-based methods [24,25]. In transform domain-based methods, Zhang et al. [26] introduced an idea based on the non-subsampled contourlet transform to solve the fusion problem of multifocus images. Chen et al. [27] presented the Intensity-Hue-Saturation model, which uses the log-Gabor wavelet transform method to fuse high-frequency and low-frequency sub-bands. For spatial fusion methods, Li et al. [10] separated the source images into two scales and combined spatial-domain context for image fusion.

Kumar et al. [11] proposed fusing the source images by weighted average, where the weights are calculated from the detail images extracted from the source images with cross bilateral filters. Li et al. [12] described a spatial domain method to solve the problem of multimodal image fusion using the structure-preserving filter. However, decomposing the transform and spatial domain components in the traditional fusion methods mentioned above is time-consuming. Besides, these methods rely on considerably intricate manually-designed fusion regulations. As a result, it is challenging to convert them into practical application tools [8].

2.2. CNN-Based Methods

CNN has succeeded extensively in image processing and gradually established a critical branch of image fusion due to its powerful feature expression capability [28–31]. CNN-based methods are widely adopted to extract image features for image fusion. For example, Liu et al. [13] built a deep CNN to generate activity level measurement and fusion rule jointly. Li et al. [14] utilized an encoder to extract the grayscale feature, and the decoder is utilized to generate a fused image. Zhang et al. [15] employed a fully convolutional neural network to reconstruct the input image, named IFCNN. This method combined an applicable fusion rule to select the type of input images. Xu et al. [16] utilized a unified densely connected network combining weight blocks to obtain retention degrees of features in different source images. Xu et al. [17] trained the U2Fusion network to maintain the adaptive similarity between the fused result and the source images. Liu et al. [32] employed a coupled contrastive constraint and a multi-level attention module to simultaneously retain complementary features from both modalities. Mu et al. [33] proposed an Auto-searching Light-weighted Multi-source Fusion network (ALMFnet), which incorporates both software and hardware knowledge in a network architecture searching manner. Li et al. [34] proposed a flexible semantically guided architecture network with a mask optimization framework to efficiently preserve unique features from different modalities.

The existing CNN-based image fusion methods heavily rely on the supervised learning of the network, with the strong assumption that the ground truth has been provided. Although the ground truth is well-defined for multimodal medical image fusion, it is not realistic to define such criteria (both dense structure and soft tissue) for fusing images in the task of CT and MRI image fusion. For example, while multimodal image fusion tasks such as pansharpening requires a crisp image with no dim parts or a multispectral image with the same resolution as the panchromatic image, CT and MRI image fusion relies on the manual design of complex fusion rules. The existing CT and MRI CNN-based methods assess the smoothness of each patch in the source images by learning a depth model and compute the corresponding weight map to produce the ultimate fused image.

2.3. GAN-Based Methods

The conception of Generative Adversarial Networks (GAN) was proposed by Goodfellow et al. [20]. The original GAN comprises two adversarial networks: a generator and a discriminator. The generator learns the data distribution and constructs a simulated image that looks real. The generator aims to minimize the data distribution gap between the generated and real images until the discriminator is unable to distinguish them.

Mathematically, the generative model G is designed to generate images with a distribution that attempts to approximate the distribution of the real training data (P_{data}). The generator G and discriminator D build a minimax two-player game, formulated as:

$$\min_G \max_D V(D,G) = \mathbb{E}_{x\sim P_{\text{data}}(x)}[\log(D(x))] + \mathbb{E}_{z\sim P_z(z)}[\log(1-D(G(z)))]. \quad (1)$$

GAN has been extensively adopted in image fusion and has achieved remarkable results. For instance, Ma et al. presented FusionGAN [35], in which the generator can directly generate a fused image with prominent structures and plentiful textures.

Xu et al. [36] adopted the self-attention scheme in the generator to retain and fuse local details. Fu et al. [22] designed a generator network based on a convolutional network with dense blocks to enrich the characteristic information. However, the fused images of the GAN-based methods often suffer from image blur, loss of details, and poor perception. We hypothesize that some valuable source image features are missing during the fusion process, and we need a more rational architecture to preserve those features.

3. Proposed Method

3.1. Overview

The diagram of the CFGAN is illustrated in Figure 1, which consists of coupled generators and discriminators to effectively assemble the typical information in CT and MRI images. Initially, the multimodal medical images I_{mm} are fed into a pair of generators. The generator G_1 is devoted to intensifying the dense structure information of the CT image in the generated image $G_1(I_{mm})$. The discriminator D_1 is designed to distinguish the relative offset of the generated image from the CT image. Similarly, the second generator G_2 attempts to inject gradient information from the MRI image into the generated image $G_2(I_{mm})$. The discriminator D_2 measures the offset of the second generated image relative to the MRI image. With training iterations, the two coupled generators are able to attain reliable images that preserve both structural information of CT images and textural information of MRI images.

However, in the dual-branch structure, each generated image may be biased toward its corresponding specific source image. To mitigate this bias, the two generated images are averaged to form the fused image. The final fused image compensates for the limitations of the two generated images while leveraging their respective strengths. Additionally, the training phase incorporates pre-fused images I_{pf} as guidance for the coupled generators to avoid blurring and detail loss [8]. The overall learning process is depicted in Algorithm 1.

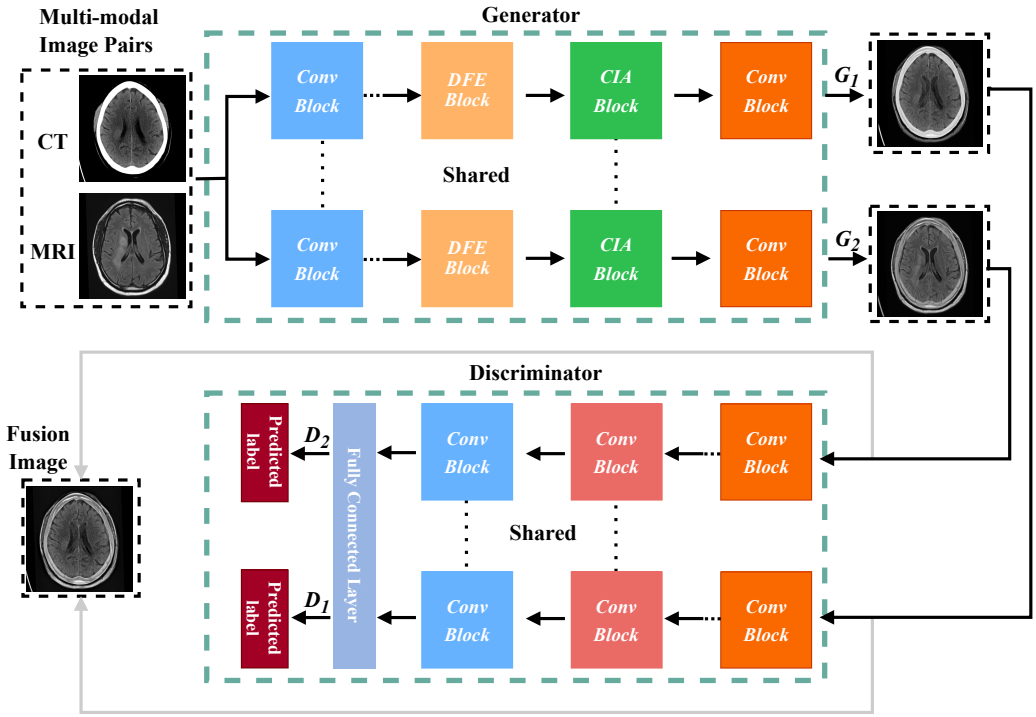

Figure 1. Diagram of the CFGAN for CT and MRI image fusion.

Algorithm 1: Training algorithm for CFGAN.

Input: The multimodal medical images **I**(**CT**, **MRI**)
Output: Fusion images.
for *1:epoch number* **do**
  for *1:iteration number* **do**
    Train the coupled discriminators
    Sample n images generated by first generator $G_1(I^1_{mm}), \ldots, G_1(I^n_{mm})$.
    Sample n CT images $I^1_{ct}, \ldots, I^n_{ct}$.
    Update the first discriminator with Adam optimizer.
    Sample n images generated by second generator $G_2(I^1_{mm}), \ldots, G_2(I^n_{mm})$.
    Sample n MRI images $I^1_{mri}, \ldots, I^n_{mri}$.
    Update the second discriminator with Adam optimizer.
    Train the coupled generators
    Sample n images generated by first generator $G_1(I^1_{mm}), \ldots, G_1(I^n_{mm})$.
    Sample n pre-fused images $I^1_{pf}, \ldots, I^n_{pf}$.
    Sample n CT images $I^1_{ct}, \ldots, I^n_{ct}$.
    Update the first generator with Adam optimizer.
    Sample n images generated by second generator $G_2(I^1_{mm}), \ldots, G_2(I^n_{mm})$.
    Sample n MRI images $I^1_{mri}, \ldots, I^n_{mri}$.
    Update the second generator with Adam optimizer.
  end
end

3.2. Generators Architecture

3.2.1. Network Design

Figure 2 depicts the architecture of the coupled generators. It is based on a siamese convolutional neural network combined with a discriminative feature extraction (DFE) block and cross-dimension interaction attention (CIA) block. The first three convolutional blocks have shared weights, where the 5×5 filters are utilized in the first and second blocks, and the 3×3 filters are set in the third block. The large convolutional filters obtain large receptive fields directly from feature maps of input multimodal image pairs, and the small convolutional filters optimize the feature maps efficiently. Then, the DFE block is utilized to sample the varied scale information densely. Next, we combine a CIA block to capture the different spatial directions and precise positional information. The convolution kernels of the last two layers are 3×3 and 1×1, respectively. The 3×3 kernel condenses the output feature, while the 1×1 filter reduces the dimension to achieve feature fusion, enabling end-to-end generation of the fused image. In addition, the convolutional block contains:

- A Batch-Normalization (BN) layer follows each layer.
- A LReLU [37] activation function in the first four layers.
- A Tanh activation function in the fifth layer.

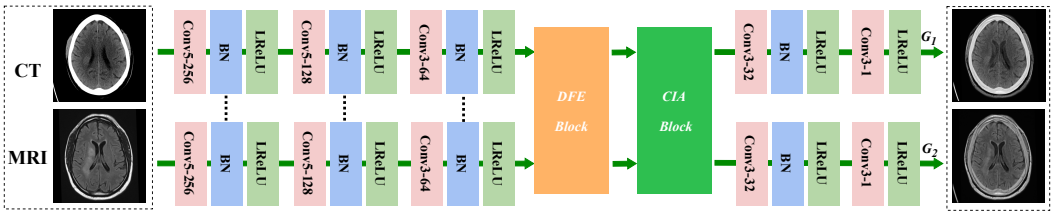

Figure 2. Architecture of the generator. Conv(k-n) indicates the convolutional layer with k filter sizes and n channels. BN represents the Batch Normalization, and FC indicates fully connected layer.

The stride is set to 1, and the padding is set to the 'SAME' for all convolution operations. The number of channels is set to 256, 128, 64, 32, and 1, respectively. The two generators

share the weights of the first three convolution blocks for the coupling design. This shared structure allows the shallow layers to extract preliminary information common to the multi-modal images, facilitating the learning of joint distributions while reducing the number of parameters.

3.2.2. Discriminative Feature Extraction Block

The discrimination of multi-scale features is an essential factor in image fusion. Notably, a convolutional kernel of a single size is limited to capturing information within a fixed receptive field. Thus, the contextual information across different ranges cannot be effectively extracted. To address this limitation, we design a discriminative feature extraction (DFE) block to enlarge the diversity and receptive field. The architecture of the DFE block is illustrated in Figure 3.

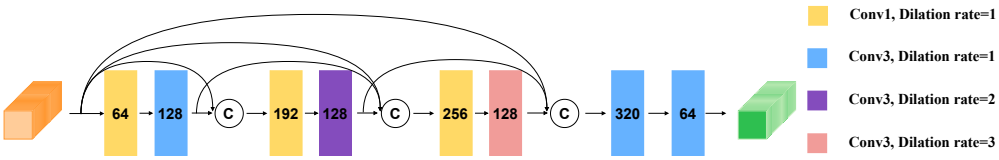

Figure 3. Architecture of the discriminative feature extraction (DFE) block.

The DFE block contains three 3×3 dilated convolutional filters with various dilation rates of 1, 2, and 3 to increase the receptive field and maintain robustness at various scales. We insert a 1×1 convolutional filter before each dilation layer as a refinement unit for parameter efficiency. Furthermore, we adopt the channel concatenation to fuse multi-scale feature maps of different dimensions. Each dilated layer in the block is tightly integrated with the other layers in the DFE block, so each layer can communicate with all subsequent layers and provide information that needs to be retained. The combination of dilated convolutions and network structure offers two key advantages. First, the top layer considers all pixels in the original feature map. Second, we use the DFE block to avoid irrelevant information across large distances caused by large expansion rates in the middle layer. The design retains dense-scale information, which is critical for image fusion to extract the available multimodal features.

3.2.3. Cross-Dimension Interaction Attention Block

To improve the performance of the network, we adopt an attention mechanism to extract salient features. The structure of the proposed cross-dimension interaction attention (CIA) block is depicted in Figure 4.

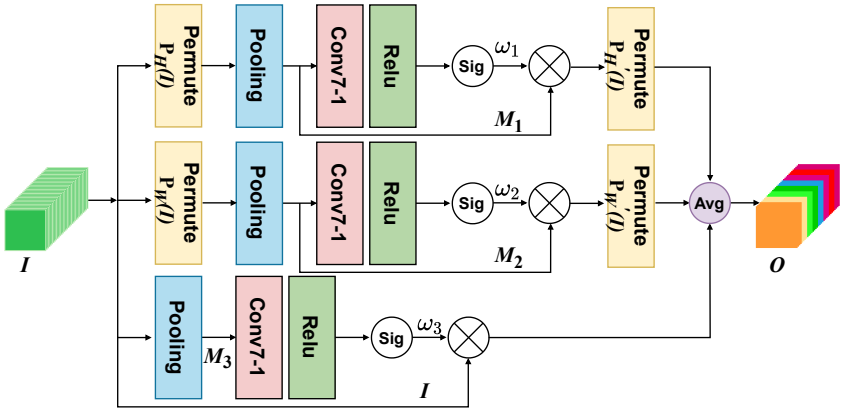

Figure 4. Architecture of the cross-dimension interaction attention (CIA) block.

The input block follows the output of the DFE block with the dimension of $I \in \mathbb{R}^{H \times W \times C}$. The CIA block contains three branches to capture dependencies between the (H, C), (W, C), and (W, H) dimensions of the input tensor, respectively. In the first branch,

we establish correlations between the height H and channel dimensions C. It is formulated as follows:

$$\omega_1 = \sigma(\mathrm{C}_7(\mathrm{Pool}(\mathrm{P}_H(I)))), \tag{2}$$

where $\mathrm{P}_H(\cdot)$ denotes the position of the permuted C and H. This output M_1 is of the shape $\mathbb{R}^{W\times H\times C}$. The Pool operation represents the concatenation of max pooling and average pooling along the channel dimension. The feature map dimension is $\mathbb{R}^{2\times H\times C}$. C_7 presents a convolution filter with the kernel size of 7×7, which provides the intermediate output of dimensions $\mathbb{R}^{1\times H\times C}$. The attention weights are then obtained via a sigmoid function σ. The first branch output is subsequently permuted to match the same shape as the input I.

In the same way, the second branch's attention weight is denoted as follows:

$$\omega_2 = \sigma(\mathrm{C}_7(\mathrm{Pool}(\mathrm{P}_W(I)))), \tag{3}$$

where $\mathrm{P}_W(\cdot)$ represents the position of the permuted C and H. The shape M_2 is updated to $\mathbb{R}^{H\times C\times W}$. After the Pool operation, the dimension of the feature map becomes $\mathbb{R}^{2\times H\times C}$. The convolutional filter 7×7 is utilized to generate a tensor of the shape $\mathbb{R}^{1\times C\times H}$. A sigmoid function σ generates the second branch attention weights. The second branch output is subsequently permuted to maintain the same shape as the input I.

For the last branch, the channels of the input tensor are reduced to two-dimension $M_3 \in \mathbb{R}^{2\times H\times W}$. The 7×7 kernel size can reduce the channel dimension. The output is passed through a sigmoid function to generate the attention weights $\omega_3 \in \mathbb{R}^{1\times H\times W}$, which is applied to the input I. It is formulated as follows,

$$\omega_3 = \sigma(\mathrm{C}_7(\mathrm{P}_W(I))). \tag{4}$$

To sum up, the process to obtain the refined attention map O from the CIA block for an input tensor $I \in \mathbb{R}^{H\times W\times C}$ can be formulated as

$$O = \frac{1}{3}[\mathrm{P}'_H(M_1\omega_1) + \mathrm{P}'_W(M_2\omega_2) + I\omega_3], \tag{5}$$

where $\mathrm{P}'_H(\cdot)$ and $\mathrm{P}'_W(\cdot)$ indicate permutation to recover the original input dimension $\mathbb{R}^{H\times W\times C}$.

3.3. Discriminator Architecture

The discriminators D_1 and D_2 share the same architecture, which is simpler than the generator architecture as depicted in Figure 5. The discriminators are intended to be adversarial to the generators. The input images of these discriminators are the generated images by the coupled generators G_1 and G_2.

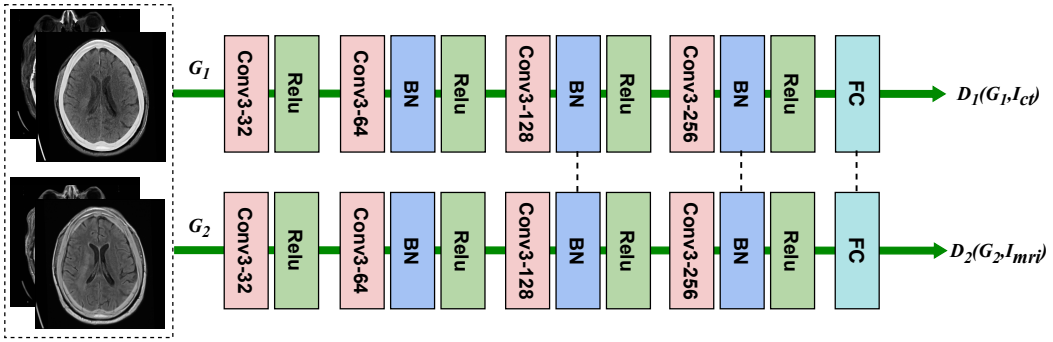

Figure 5. Architecture of the discriminator. Conv(*k*-n) denotes the convolutional layer with a *k* filter size. BN is short for batch normalization, and FC denotes the fully connected layer.

Each discriminator includes four convolutional blocks and one fully connected layer. The convolution blocks utilize 3×3 convolution filters with padding operation. The

number of channels is set to 32, 64, 128, and 256, respectively. In the first four convolutional layers, we use the LReLU activation function, and in the final layer, we use the Tanh activation function. The batch normalization layer is embedded into the middle three-layer convolutional blocks. To reduce the parameter count, the weights of the third and fourth convolutional blocks and the fully connected layer are shared. The stride is set to 2 to reduce the feature map size. The final linear layer converts the flattened feature map into one output that indicates the discriminator's assessment of the authenticity of the image generated by G.

3.4. Loss Function

3.4.1. Generator Loss Function

The first generator G_1 learns the dense structure (e.g., bones and implants) characteristic of the CT image derived from the pre-fused image. The loss function $\mathcal{L}_{G_1}$ consists of the adversarial loss $\Phi(G_1)$ and the content loss $\mathcal{L}_{\text{con1}}$ with a weight λ controlling the trade-off. It is formulated as follows:

$$\mathcal{L}_{G_1} = \Phi(G_1) + \lambda \mathcal{L}_{\text{con1}} . \tag{6}$$

The $\Phi(G_1)$ stands for the adversarial loss between generator G_1 and discriminator D_1. It is denoted as

$$\Phi(G_1) = \frac{1}{N} \sum_{n=1}^{N} (D_1(G_1(I_{mm}^n), I_{ct}^n) - a)^2, \tag{7}$$

where N denotes the number of fused images. a is the target value that the generator aims to make the discriminator believe as true for generated images.

$\mathcal{L}_{\text{con1}}$ indicates the content loss for the first generator

$$\mathcal{L}_{\text{con1}} = \frac{1}{WHN} \sum_{n=1}^{N} \left(\mu \| G_1(I_{mm}^n) - I_{ct}^n \|_2^2 + \left\| G_1(I_{mm}^n) - I_{pf}^n \right\|_2^2 \right), \tag{8}$$

where $\|\cdot\|_2$ denotes the matrix 2-norm. The width and height of the input image are indicated by W and H, respectively. The first term of the $\mathcal{L}_{\text{con1}}$ maintains bone structure information of the CT image I_{ct} in the generated image $G_1(I_{mm})$, and the second term preserves the pre-fused information contained in the pre-fused image I_{pf}. μ is utilized to coordinate the trade-off between the $\Phi(G_1)$ and $\mathcal{L}_{\text{con1}}$. Through the loss function $\mathcal{L}_{G1}$, the first generator G_1 can learn the dense structure information of the CT image I_{ct} and retain the details of the pre-fused image I_{pf}.

In the second generator G_2, we aim to integrate the gradient information of the MRI image I_{mri} into the generated image $G_2(I_{mm})$. The loss function $\mathcal{L}_{G_2}$ of the second generator is defined as

$$\mathcal{L}_{G_2} = \Phi(G_2) + \lambda \mathcal{L}_{\text{con2}} , \tag{9}$$

where λ is applied to coordinate the trade-off between $\Phi(G_2)$ and $\mathcal{L}_{\text{con2}}$.

The $\Phi(G_2)$ stands for the adversarial loss between generator G_2 and discriminator D_2. It is formulated as,

$$\Phi(G_2) = \frac{1}{N} \sum_{n=1}^{N} (D_2(G_2(I_{mm}^n), I_{mri}^n) - a)^2. \tag{10}$$

The second term content denotes the content loss $\mathcal{L}_{\text{con2}}$ and it is formulated as follows:

$$\mathcal{L}_{\text{con2}} = \frac{1}{WHN} \sum_{n=1}^{N} \left(\beta \| \nabla G_2(I_{mm}^n) - \nabla I_{mri}^n \|_2^2 + \left\| G_2(I_{mm}^n) - I_{pf}^n \right\|_2^2 \right), \tag{11}$$

where $\bigtriangledown$ represents the gradient operator. The first term of the $\mathcal{L}_{\text{con2}}$ preserves the gradient information of the MRI image I_{mri} in the generated image $G_2(I_{mm}^n)$ by the second generator G_2, and the second term keeps the pre-fused information contained in the pre-fused image

I_{pf}. β is utilized to control the trade-off between the two terms. Consequently, G_2 can learn the gradient characteristics of the MRI image derived from the pre-fusion image.

The coupled generators can be regarded as optimizing the pre-fusion image along various orientations. The final result image retains both CT dense structure and MRI texture information. Hence, the fused outcome F is the mean value of the two generated images as follows:

$$F = \frac{1}{2}(G_1(I_{mm}) + G_2(I_{mm})). \tag{12}$$

3.4.2. Discriminator Loss Function

The coupled discriminators D_1 and D_2 play a role in distinguishing the source images and the generated fused image. Furthermore, through backpropagation, the fusion images incorporate the information of the corresponding opposite image. The $\mathcal{L}_{D_1}$ represents a measurement of the relative proximity of the image generated by the generator G_1 to the CT image. The loss function formula for the first discriminator D_1 is formulated as

$$\mathcal{L}_{D_1} = \mathbb{E}[-\log D_1(I_{ct})] + \mathbb{E}[-\log(1 - D_1(G_1(I_{mm})))]. \tag{13}$$

The $\mathcal{L}_{D_2}$ is dedicated to calculating the correlation of the image generated by the generator G_2 to the MRI image. The loss function of the second discriminator D_2 is formulated as

$$\mathcal{L}_{D_2} = \mathbb{E}[-\log D_2(I_{mri})] + \mathbb{E}[-\log(1 - D_2(G_2(I_{mm})))]. \tag{14}$$

4. Experimental Results and Analysis

4.1. Dataset and Training Details

In the experimental section, the CT and MRI medical images are obtained from publicly available datasets provided by the Whole Brain Atlas database of Harvard Medical School [38] and other online sites [39,40]. All the acquired data are pre-registered to ensure spatial correspondence between the CT and MRI images. We employ 50 pairs of CT and MRI scan images for the experiment, which are transformed to grayscale and resized to 256 × 256. Among these, fifteen pairs of images are presented as a test set in Figure 6.

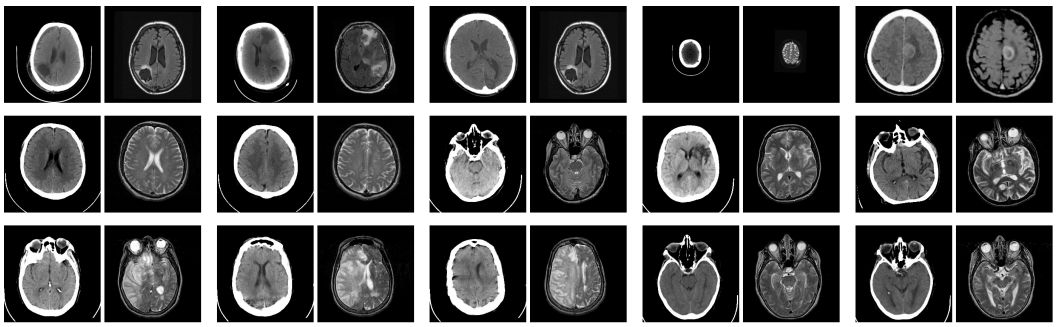

Figure 6. Fifteen pairs of CT-MRI images for evaluation. In each pair, the left is the CT image, and the right is the MRI image.

During the training procedure, to ensure a sufficient number of training samples for CFGAN, the training images are cropped to patches of size 120 × 120, and randomly flipped horizontally and vertically. The depth and width of feature maps are constricted because the coupled generator does not use padding operations. In order to maintain the output size at 120 × 120, all patch images need to be zero-padded to 132 × 132. The CT and MRI image patches are combined in a dual channel and delivered to the coupled generators. The network is trained for 100 epochs using the Adam optimizer [41], and the batch size is set to 32 by default. All the experiments are implemented in PyTorch equipped with an NVIDIA RTX 3090Ti GPU. In this study, we use the fused results of IFCNN [15] as pre-fused images.

4.2. Experiments and Analysis

In this section, comparative experiments are conducted in subjective and objective assessments to verify the effectiveness of the proposed CFGAN. The subjective map indicates the sensory quality of the fused image and the degree of retention of significant information in the source image. The objective assessment utilizes evaluation metrics to further differentiate between images with analogous sensory quality. In this work, six metrics, i.e., entropy standard deviation (SD) [42], peak signal-to-noise ratio (PSNR) [43], correlation coefficient (CC) [44], structural similarity index measure (SSIM) [45], visual information fidelity (VIF) [46], and Mutual information (MI) [47] are used for objective evaluation. Notably, these metrics are obtained by comparing the fused image with each of the source images separately and then averaging the results. The proposed CFGAN is contrasted with twelve state-of-the-art image fusion methods, i.e., GFF [10], CBF [11], CNN [13], SAIF [12], FusionGAN [35], Densefuse [14], IFCNN [15], DDcGAN [21], FusionDN [16], MEF-GAN [36], PerceptualFusion [22], and U2Fusion [17]. First, four typical case studies are presented in detail. Then, the objective evaluations of the competitors in the whole dataset are discussed.

4.2.1. Case Study

Case 1: Acute stroke presenting as speech arrest. The experimental data were obtained from a patient who was a 63-year-old right-handed male with a history of Micronase-treated adult-onset diabetes mellitus and arterial hypertension ([Online]. Available: http://www.med.harvard.edu/aanlib/cases/case2/case.html, accessed on 15 May 2023). The subjective comparison results of the first case are depicted in Figure 7. The CT image is commonly negative during the acute period of stroke, and the MRI image reveals acute cerebral infarction involving the left pre-central gyrus. Preferably, the fused image retains the bone part from CT and the textural information from MRI. Although the traditional methods (e.g., GFF [10], CBF [11], and SAIF [12]) perform well in persevering soft tissues from images, they exhibit poor results in maintaining the illumination intensity of images. The white contour in the CT image shows the skull, but the fusion of GFF and CBF results in the loss of a majority of the skull information. The CNN-based methods (e.g., CNN [13], IFCNN [15], FusionDN [16], and U2Fusion [17]) and GAN-based methods (e.g., DDcGAN [21] and PerceptualFusion [22]) have lower contrast in the skull part. The details of brain tissue are sufficiently clear, except for FusionGAN [35] and MEF-GAN [36]. Densefuse [14] and PerceptualFusion [22] lack some tissue information in the boundary between encephalic tissue and the skull in the red box, close-up. DDcGAN [21] and CFGAN retain the skeletal information of CT images more than other competitors. DDcGAN [21], FusionDN [16], and CFGAN have high contrast, as well as preserving soft tissue information. The objective results are depicted in Table 1. It illustrates that the proposed CFGAN performs best in the other four objective indicators except for the CC indicator, in which the CFGAN ranks fourth place.

Case 2: Acute stroke presenting as right body weakness. This case is from a 45-year-old female with a sudden onset of right body weakness and trouble speaking ([Online]. Available: http://www.med.harvard.edu/aanlib/cases/case20/case.html, accessed on 15 May 2023). The subjective comparison results of the second case are depicted in Figure 8. These methods (e.g., GFF [10], SAIF [12], Densefuse [14], and IFCNN [15]) preserve some tissue texture information, but the fused images have lower contrast, resulting in missing cephalometric information. The comparison between CNN [13] and FusionGAN [35] has higher contrast, but the edges are blurred, and some information about brain tissue is lost. The MEF-GAN [36], U2Fusion [17], and PerceptualFusion [22] are unable to attain distinct textures or boundaries. The superior colliculus of the fused results in the red box (e.g., DDcGAN [21], FusionDN [16], and CFGAN) has higher brightness and richer details, but the bone information of DDcGAN [21] lacks details, including edges and texture. Our proposed method can reserve more significant information, particularly gradient information, contrast, boundary, and textural details. The objective comparison results are

denoted in Table 2. It illustrates that the proposed CFGAN ranks first in SD, PSNR, SSIM, and MI. For the CC and VIF indicators, the CFGAN ranks third and second, respectively.

Table 1. The objective comparison results for the first case. The most prominent results are highlighted in **bold**.

Methods	SD	PSNR	CC	SSIM	VIF	MI
GFF [10]	10.5126	15.7217	0.7924	0.6542	0.5450	3.0995
CBF [11]	10.4616	15.2092	0.7695	0.6506	0.4582	3.2984
CNN [13]	10.5146	15.3438	0.7643	0.6614	0.5721	3.3877
SAIF [12]	10.6046	14.8446	0.7587	0.6575	0.5750	3.3796
FusionGAN [35]	8.9678	12.4935	0.7900	0.2610	0.4436	3.2101
Densefuse [14]	9.7168	14.1000	0.7258	0.1653	0.2010	2.5983
IFCNN [15]	10.6539	16.0550	0.8132	0.6584	0.4731	3.1806
DDcGAN [21]	10.5925	12.2912	0.7916	0.2341	0.3456	3.0899
FusionDN [16]	10.5334	11.5345	0.7938	0.2742	0.4315	3.2145
MEF-GAN [36]	10.5422	14.1312	0.7878	0.6377	0.4211	3.0332
PerceptualFusion [22]	10.6509	12.5574	**0.8209**	0.2889	0.4393	3.2912
U2Fusion [17]	10.4145	16.2216	0.8094	0.3732	0.3993	3.1125
Ours	**10.6910**	**16.5646**	0.7953	**0.6836**	**0.5759**	**3.4058**

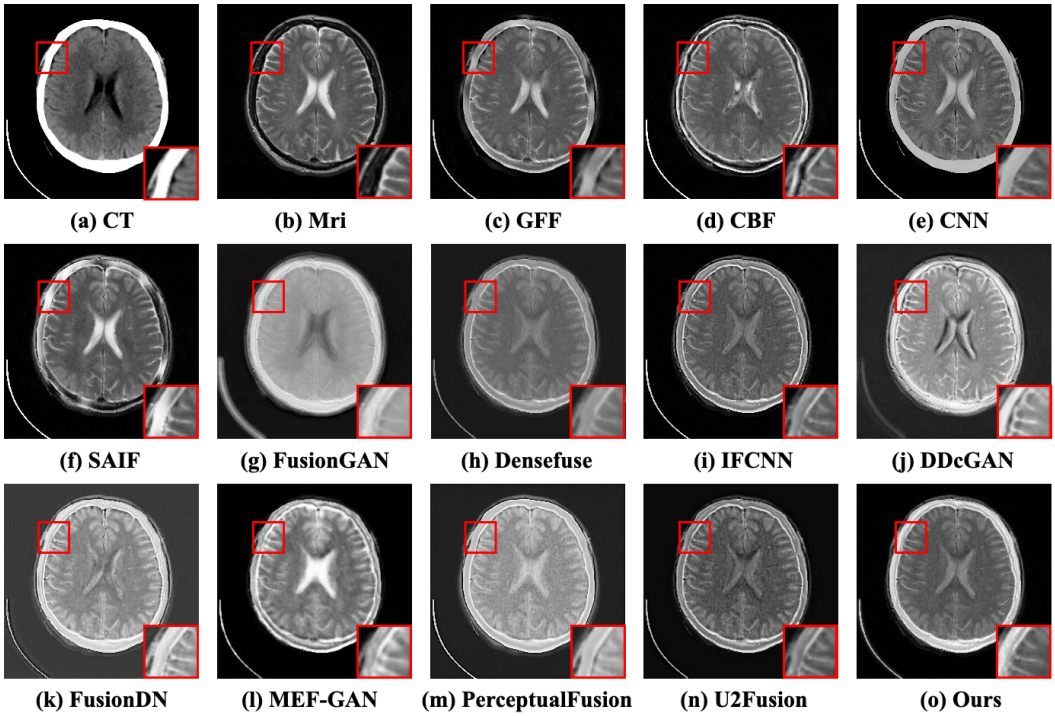

Figure 7. The subjective comparison results of the first case. (**a**,**b**) represent the CT and MRI images. (**c**–**n**) denote the fused results of the competitors. (**o**) is the result of CFGAN. In this and following figures, at the bottom right of each subfigure, we show the highlighted image in red box.

Case 3: Multiple infarctions. The third case is a 55-year-old male who suffered multiple refractory focal seizures in the setting of pulmonary empyema ([Online]. Available: http://www.med.harvard.edu/aanlib/cases/case34/case.html, accessed on 15 May 2023). The fused results of the subjective comparison for the third case are illustrated in Figure 9. The source MRI has more clarity and more tissue detail than the CT image, and it makes sense for both of the above pieces of information to be retained in the fused image. Nevertheless, these methods (e.g., GFF [10], CBF [11], and FusionGAN [35]) have weak visual contrast. There are distortions in the contours of images (e.g., SAIF [12], DDcGAN [21], IFCNN [15], and MEF-GAN [36]). In general, SAIF [12], DDcGAN [21], and CFGAN achieve superior perceived quality. Viewing the atrium in the red box, we can observe

that CBF [11], SAIF [12], IFCNN [15], DDcGAN [21], U2Fusion [17], and CFGAN reserve more texture details. The six objective evaluation indicators are presented in Table 3. The proposed method CFGAN is proved to show better performance compared to the other twelve methods.

Table 2. The objective comparison results for the second case. The most prominent results are highlighted in **bold**.

Methods	SD	PSNR	CC	SSIM	VIF	MI
GFF [10]	9.4583	14.8625	0.8124	0.7089	0.6221	2.7335
CBF [11]	9.2601	13.8956	0.7903	0.7014	0.4652	2.7443
CNN [13]	9.2192	18.2652	0.7777	0.7408	0.5749	3.2288
SAIF [12]	9.2985	13.6398	0.7526	0.7253	**0.6994**	2.9468
FusionGAN [35]	8.0180	13.1487	0.8258	0.1782	0.4517	2.7466
Densefuse [14]	9.3583	11.6880	0.6073	0.0601	0.0659	1.8011
IFCNN [15]	9.4512	15.7070	0.8453	0.6850	0.5271	2.8248
DDcGAN [21]	9.4113	10.4151	0.8007	0.1418	0.2751	2.5835
FusionDN [16]	9.1808	10.0367	0.7838	0.2000	0.4092	2.6537
MEF-GAN [36]	9.2921	14.8140	0.8339	0.6550	0.4572	2.8217
PerceptualFusion [22]	9.5938	12.9201	**0.8544**	0.2183	0.4480	2.7561
U2Fusion [17]	9.2378	15.1122	0.8453	0.2495	0.4489	2.7753
CFGAN (Ours)	**9.5857**	**15.9410**	0.8350	**0.7410**	0.6541	**3.0756**

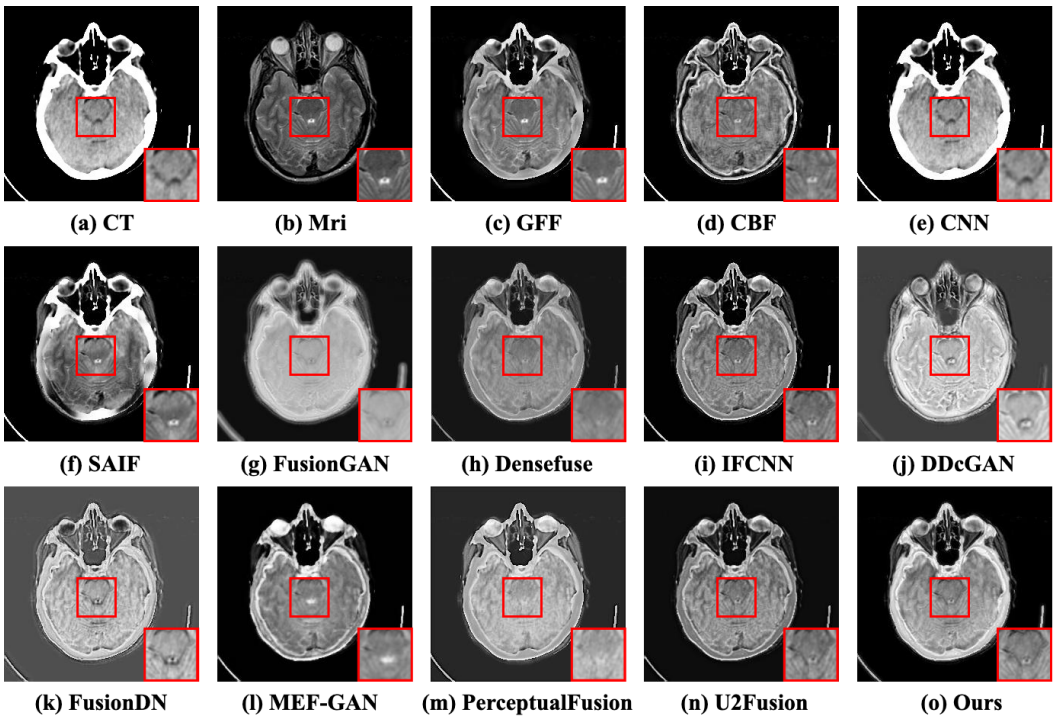

Figure 8. The subjective comparison results of the second case. (**a**,**b**) represent the CT and MRI images. (**c**–**n**) denote the fused results of the competitors. (**o**) is the result of CFGAN.

Case 4: Fatal stroke. The experimental case was collected from a patient who developed a sudden onset of left-sided hemiparesis, muteness, and bilateral ptosis ([Online]. Available: http://www.med.harvard.edu/aanlib/cases/case37/case.html, accessed on 15 May 2023). The cerebral infarct lesion showed abrupt contrast variation in both CT and MRI images as illustrated in Figure 10a,b. GFF [10], CBF [11], and SAIF [12] have inferior performance in retaining the profile information of the CT image. IFCNN [15], Densefuse [14], and U2Fusion [17] show lower contrasted images. MEF-GAN [36] and CFGAN preserve more cranial information from CT images and more tissue information from MRI images compared with U2Fusion [17]. Still, some illumination information is

lost at the contour by MEF-GAN [36]. The objective comparison results are tabulated in Table 4. It demonstrates that the CFGAN performs best in the SD, PSNR, SSIM, VIF, and MI indicators.

Table 3. The objective comparison results for the third case. The most prominent results are highlighted in **bold**.

Methods	SD	PSNR	CC	SSIM	VIF	MI
GFF [10]	10.3692	14.3832	0.7663	0.6704	0.5638	3.0355
CBF [11]	10.3748	13.9841	0.7378	0.6702	0.4505	3.2002
CNN [13]	9.8755	14.0121	0.7290	0.6909	0.4346	3.3365
SAIF [12]	9.6132	13.1604	0.7198	0.6771	0.5445	3.2127
FusionGAN [35]	8.0868	12.2890	0.7386	0.1762	0.4057	3.0279
Densefuse [14]	9.6829	12.4549	0.6183	0.0708	0.1048	2.3603
IFCNN [15]	10.4489	14.6458	0.7562	0.6482	0.4790	3.1004
DDcGAN [21]	10.5841	11.4985	0.7390	0.1692	0.3079	2.8855
FusionDN [16]	10.3115	11.0779	0.7782	0.2351	0.4106	3.0387
MEF-GAN [36]	10.2738	12.8709	0.7750	0.5986	0.4383	3.1169
PerceptualFusion [22]	10.6757	12.2461	**0.7912**	0.2437	0.4488	3.0562
U2Fusion [17]	9.9803	14.9898	0.7804	0.2786	0.3976	3.0455
CFGAN (Ours)	**10.5600**	**15.2686**	0.7650	**0.6931**	**0.5838**	**3.3411**

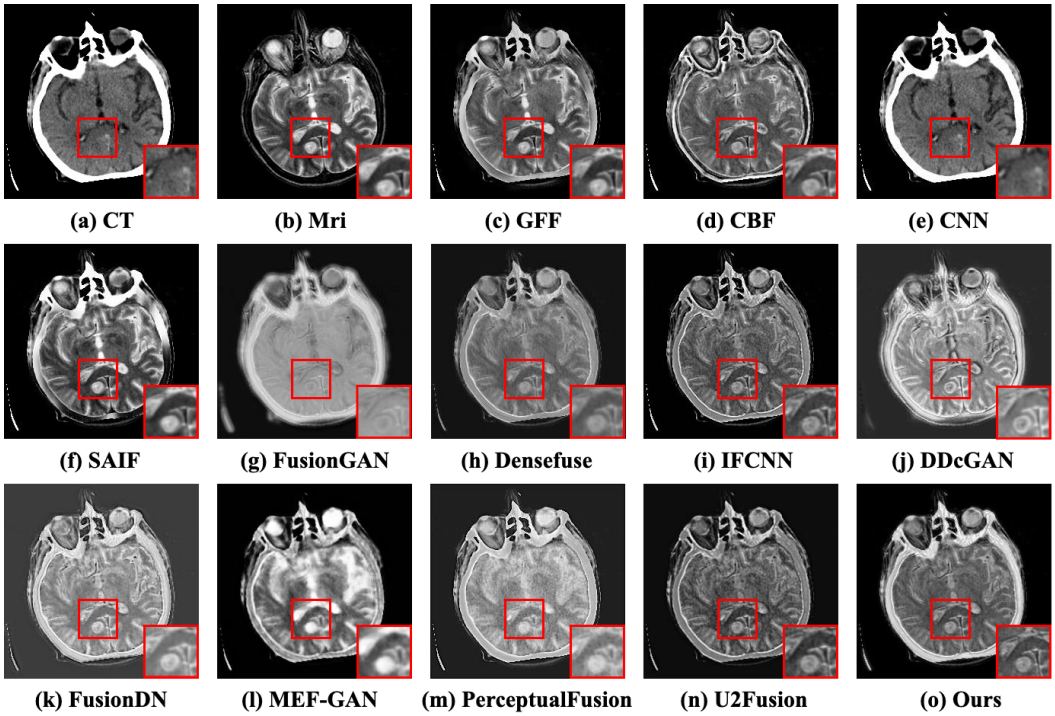

Figure 9. The subjective comparison results of the third case. (**a**,**b**) represent the CT and MRI images. (**c**–**n**) denote the fused results of the competitors. (**o**) is the result of CFGAN.

Table 4. The objective comparison results for the fourth case. The most prominent results are highlighted in **bold**.

Methods	SD	PSNR	CC	SSIM	VIF	MI
GFF [10]	10.2951	14.8623	0.8270	0.7053	0.4982	2.9309
CBF [11]	9.9778	14.0533	0.7987	0.7050	0.4317	3.2012
CNN [13]	10.1570	15.0837	0.7961	0.7254	0.5409	3.2684
SAIF [12]	9.9229	13.5342	0.7783	0.7139	0.5776	3.2500
FusionGAN [35]	9.2530	13.5302	0.8135	0.2090	0.4482	3.0701
Densefuse [14]	9.4289	11.4488	0.6030	0.0730	0.0791	2.0421
IFCNN [15]	10.3208	15.2448	0.8529	0.6836	0.4638	3.0585
DDcGAN [21]	10.2889	11.7470	0.8019	0.1795	0.2906	2.8856
FusionDN [16]	10.1830	11.8313	0.8305	0.2442	0.3776	2.9502
MEF-GAN [36]	10.2693	13.4217	0.8490	0.6324	0.4237	3.0947
PerceptualFusion [22]	10.3464	13.3984	**0.8571**	0.2551	0.4208	3.0005
U2Fusion [17]	10.1849	15.2428	0.8521	0.2948	0.3827	2.9726
CFGAN (Ours)	**10.3870**	**15.7633**	0.8353	**0.7330**	**0.5850**	**3.3249**

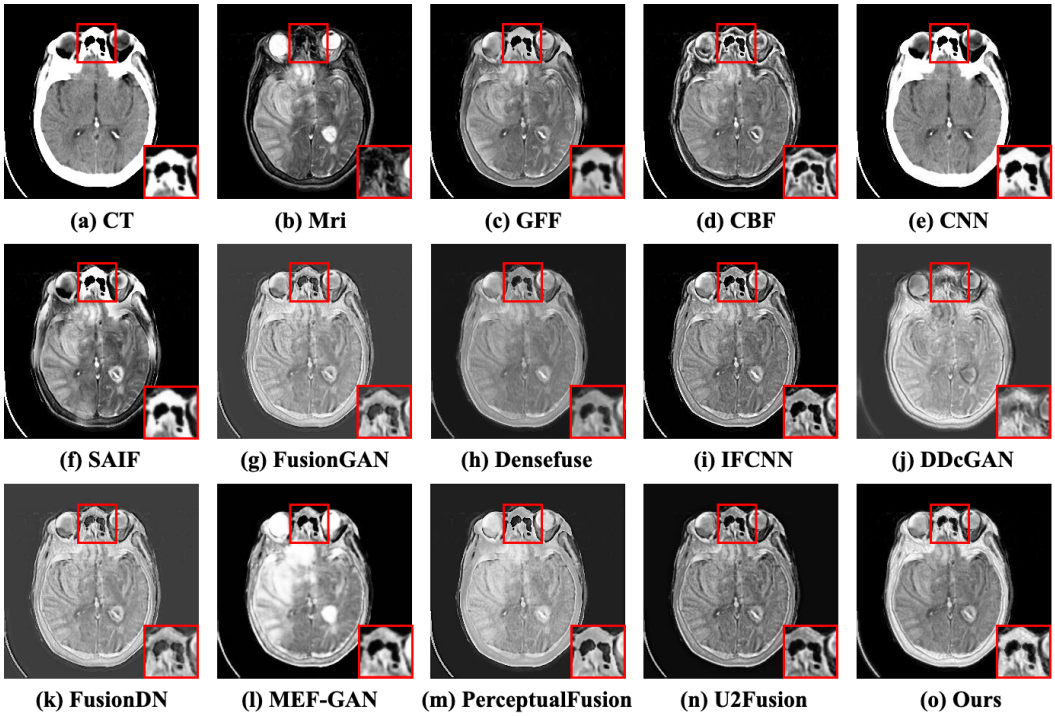

Figure 10. The subjective comparison results of the fourth case. (**a**,**b**) represent the CT and MRI images. (**c**–**n**) denote the fused results of the competitors. (**o**) is the result of CFGAN.

4.2.2. Qualitative Comparisons

Figure 11 provides the quantitative comparisons on 15 test image pairs. The proposed scheme has the most prominent values regarding the four evaluation metrics (i.e., SD, PSNR, SSIM, and VIF) for pairs 7, 13, 13, and 7 of the 15 test set image pairs. Moreover, the proposed method also shows competitive results in the CC and MI metrics. The experimental results indicate that the CFGAN can retain the source image pair feature information to the maximum extent. This means that the fused images have high contrast, rich edges, and detailed information so that the results of CFGAN are considerably similar to the source image.

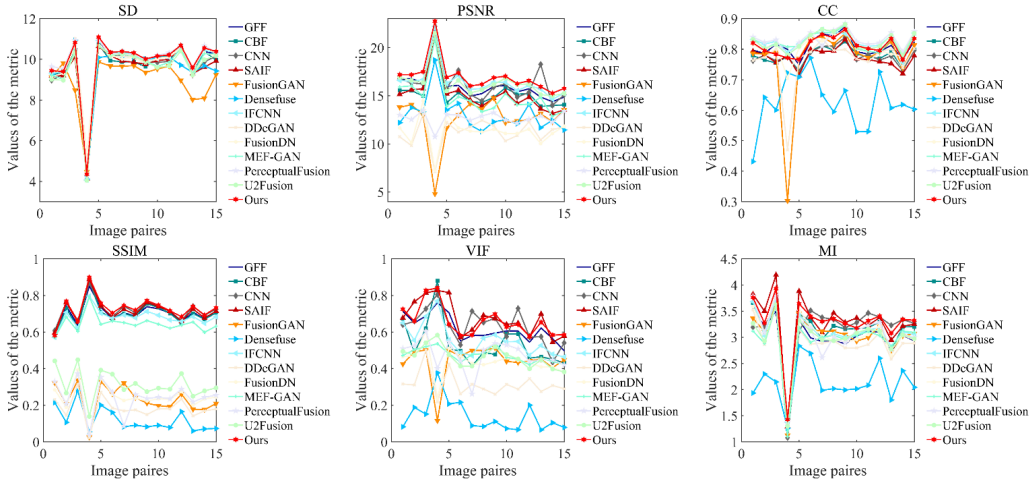

Figure 11. Quantitative comparison with SOTA competitors.

4.3. Ablation Study

To comprehensively evaluate the effectiveness of the proposed DFE and CIA blocks, we conduct a series of ablation experiments with the following detailed network configurations:

1. "Baseline" refers to the vanilla generator model without any component.
2. "Baseline + DFE" denotes the baseline model with a single DFE block.
3. "Baseline + CIA" represents the baseline model with a single CIA block.
4. "Baseline + CIA_DFE" refers to the baseline model with the CIA block and DFE block sequentially connected.
5. "Baseline + DFE_CIA" refers to the baseline model with the DFE block and CIA block sequentially connected.

The objective comparison results are shown in Table 5. The results prove that the DFE block and CIA block in the generators contribute to substantial improvements in the baseline method. The "Baseline" achieves the lowest performance. Compared with the "Baseline", the "Baseline + DFE", and "Baseline + CIA" synergize multi-scale information and salient information, which facilitates an improvement in the objective indicators of the generated images. Specifically, the "Baseline + DFE" achieves 35.5%, 24.9%, 3.8%, and 4.6% improvements in PSNR, SSIM, VIF, and MI, respectively. The "Baseline + CIA" method achieves 31.7%, 18.4%, 4.4%, and 4.6% improvements in PSNR, SSIM, VIF, and MI, respectively. Meanwhile, when the DFE and CIA blocks are simultaneously incorporated into the baseline, the improvement is more obvious. Between these two configurations, the "Baseline + DFE_CIA" is better than "Baseline+CIA_DFE". The final CFGAN with "Baseline + DFE_CIA" boosts the baseline by 1.3%, 41.2%, 1.6%, 201.2%, 18.1%, and 7.1% in terms of SD, PSNR, CC, SSIM, VIF, and MI, respectively.

The subjective results with different network configurations are illustrated in Figure 12. The example images of four cases are illustrated in Figure 12. Figure 12a,b are the source images of the CT and MRI, respectively. Figure 12c represents the fusion image generated by the baseline method. The CIA block extracts the salient features, including dense structures from CT images and soft tissue detail from MRI images as depicted in Figure 12d. The DFE block can supplement the detailed features of the source image as shown in Figure 12e. Both the compound modes of 'Baseline + CIA_DFE' (Figure 12f) and 'Baseline + DFE_CIA' (Figure 12g) can preserve more significant information, particularly skull information and details of brain tissue.

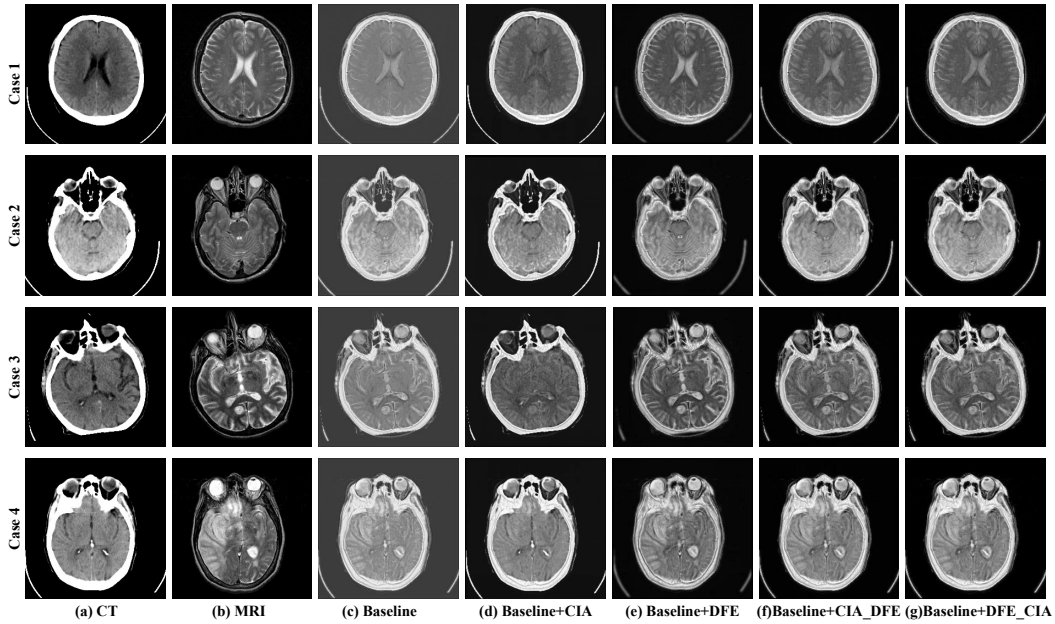

Figure 12. The subjective comparison results of different network configurations.

Table 5. Ablation analysis on the key components in CFGAN. (The best results are marked in **bold**).

Methods	SD	PSNR	CC	SSIM	VIF	MI
Baseline	9.7235	12.0252	0.7976	0.2417	0.5585	3.0487
Baseline + DFE	9.7658	16.3007	0.8066	0.3019	0.5798	3.1881
Baseline + CIA	9.6916	15.8922	0.8023	0.2862	0.5828	3.1567
Baseline + CIA_DFE	**9.8536**	16.8894	**0.8136**	0.5758	0.6096	3.2192
Baseline + DFE_CIA	9.8524	**16.9842**	0.8105	**0.7281**	**0.6597**	**3.2662**

5. Conclusions

This paper proposes a Coupled Feature-Learning GAN (CFGAN) for CT and MRI image fusion. The coupled generators and discriminators are designed to fully exploit the discriminative information in CT and MRI images. The discriminators are trained to form an adversarial relationship by distinguishing between real source images and fused images generated by the generators based on a specifically designed content loss. Meanwhile, we creatively develop a DFE block and a CIA block in the generators to expand the receptive field and facilitate the extraction of salient features. Notably, the entire model is trained in an end-to-end manner without the need for ground-truth images. Experimental results prove that the proposed method achieves competitive performance compared to other SOTA methods.

6. Future Work

The current study has demonstrated the potential of using GANs for multi-modal image fusion. However, several challenges remain to be addressed in future research. One key concern is the lack of explicit regularization to control the contributions of different modalities. While ideally, the bone information from CT and soft tissue information from MRI should be preserved in the fused image, it is not always the case, as not all CT and MRI images necessarily contain such information. To address this limitation, future work should focus on developing advanced regularization techniques and architectures that can effectively guide the fusion process to preserve modality-specific information when available. This may involve incorporating prior knowledge about the anatomical structures and their corresponding modalities into the loss functions or network architectures. By explicitly guiding the fusion process to preserve bone details from CT and soft tissue details from MRI when available, the fused images can provide more accurate and comprehensive

representations of the underlying anatomy. Furthermore, flexibly identifying and fusing important information from source images remains a challenge. Future work should explore advanced techniques for adaptively determining the relevant information to fuse from each modality based on the specific characteristics of the input images. By addressing these challenges, future studies can build upon the current findings and develop more accurate and reliable multi-modal fusion methods.

Author Contributions: Conceptualization, methodology, software, and original draft preparation: Q.M.; validation, visualization, and formal analysis: W.Z.; investigation, data curation, and resources: Z.W.; writing—review, and editing: X.L.; supervision and funding acquisition: Y.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Natural Science Foundation of China (Grant No. 62031013) and by the Guangdong Province Key Construction Discipline Scientific Research Capacity Improvement Project (Grant No. 2022ZDJS117).

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

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

Conflicts of Interest: Author Xiang Lei was employed by the Zhiyang Innovation Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Tawfik, N.; Elnemr, H.A.; Fakhr, M.; Dessouky, M.I.; El-Samie, A.; Fathi, E. Survey study of multimodality medical image fusion methods. *Multimed. Tools Appl.* **2021**, *80*, 6369–6396. [CrossRef]
2. Du, J.; Li, W.; Lu, K.; Xiao, B. An overview of multi-modal medical image fusion. *Neurocomputing* **2016**, *215*, 3–20. [CrossRef]
3. Yin, M.; Liu, X.; Liu, Y.; Chen, X. Medical image fusion with parameter-adaptive pulse coupled neural network in nonsubsampled shearlet transform domain. *IEEE Trans. Instrum. Meas.* **2018**, *68*, 49–64. [CrossRef]
4. Huang, B.; Yang, F.; Yin, M.; Mo, X.; Zhong, C. A review of multimodal medical image fusion techniques. *Comput. Math. Methods Med.* **2020**, *2020*, 8279342. [CrossRef]
5. Zhou, T.; Li, Q.; Lu, H.; Cheng, Q.; Zhang, X. GAN review: Models and medical image fusion applications. *Inf. Fusion* **2023**, *91*, 134–148. [CrossRef]
6. Mao, Q.; Yang, X.; Zhang, R.; Jeon, G.; Hussain, F.; Liu, K. Multi-focus images fusion via residual generative adversarial network. *Multimed. Tools Appl.* **2022**, *81*, 12305–12323. [CrossRef]
7. Huang, Y.; Li, W.; Gao, M.; Liu, Z. Algebraic multi-grid based multi-focus image fusion using watershed algorithm. *IEEE Access* **2018**, *6*, 47082–47091. [CrossRef]
8. Li, Q.; Lu, L.; Li, Z.; Wu, W.; Liu, Z.; Jeon, G.; Yang, X. Coupled GAN with relativistic discriminators for infrared and visible images fusion. *IEEE Sens. J.* **2019**, *21*, 7458–7467. [CrossRef]
9. Zhai, W.; Song, W.; Chen, J.; Zhang, G.; Li, Q.; Gao, M. CT and MRI image fusion via dual-branch GAN. *Int. J. Biomed. Eng. Technol.* **2023**, *42*, 52–63. [CrossRef]
10. Li, S.; Kang, X.; Hu, J. Image fusion with guided filtering. *IEEE Trans. Image Process.* **2013**, *22*, 2864–2875.
11. Shreyamsha Kumar, B. Image fusion based on pixel significance using cross bilateral filter. *Signal Image Video Process.* **2015**, *9*, 1193–1204. [CrossRef]
12. Li, W.; Xie, Y.; Zhou, H.; Han, Y.; Zhan, K. Structure-aware image fusion. *Optik* **2018**, *172*, 1–11. [CrossRef]
13. Liu, Y.; Chen, X.; Peng, H.; Wang, Z. Multi-focus image fusion with a deep convolutional neural network. *Inf. Fusion* **2017**, *36*, 191–207. [CrossRef]
14. Li, H.; Wu, X.J. DenseFuse: A fusion approach to infrared and visible images. *IEEE Trans. Image Process.* **2018**, *28*, 2614–2623. [CrossRef]
15. Zhang, Y.; Liu, Y.; Sun, P.; Yan, H.; Zhao, X.; Zhang, L. IFCNN: A general image fusion framework based on convolutional neural network. *Inf. Fusion* **2020**, *54*, 99–118. [CrossRef]
16. Xu, H.; Ma, J.; Le, Z.; Jiang, J.; Guo, X. Fusiondn: A unified densely connected network for image fusion. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 12484–12491.
17. Xu, H.; Ma, J.; Jiang, J.; Guo, X.; Ling, H. U2Fusion: A Unified Unsupervised Image Fusion Network. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *44*, 502–518. [CrossRef]
18. Song, W.; Zeng, X.; Abdelmoniem, A.M.; Zhang, H.; Gao, M. Cross-Modality Interaction Network for Medical Image Fusion. *IEEE Trans. Consum. Electron.* **2024**. [CrossRef]

19. Song, W.; Zeng, X.; Li, Q.; Gao, M.; Zhou, H.; Shi, J. CT and MRI image fusion via multimodal feature interaction network. *Netw. Model. Anal. Health Inform. Bioinform.* **2024**, *13*, 13. [CrossRef]
20. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. *Adv. Neural Inf. Process. Syst.* **2014**, *27*.
21. Ma, J.; Xu, H.; Jiang, J.; Mei, X.; Zhang, X.P. DDcGAN: A dual-discriminator conditional generative adversarial network for multi-resolution image fusion. *IEEE Trans. Image Process.* **2020**, *29*, 4980–4995. [CrossRef]
22. Fu, Y.; Wu, X.J.; Durrani, T. Image fusion based on generative adversarial network consistent with perception. *Inf. Fusion* **2021**, *72*, 110–125. [CrossRef]
23. Yang, Z.; Chen, Y.; Le, Z.; Ma, Y. GANFuse: A novel multi-exposure image fusion method based on generative adversarial networks. *Neural Comput. Appl.* **2021**, *33*, 6133–6145. [CrossRef]
24. RattÁ, G.; Vega, J.; Murari, A.; Contributors, J. Image fusion: Advances in the state of the art. *Inf. Fusion* **2007**, *8*, 114–118.
25. Mitianoudis, N.; Stathaki, T. Pixel-based and region-based image fusion schemes using ICA bases. *Inf. Fusion* **2007**, *8*, 131–142. [CrossRef]
26. Zhang, Q.; Guo, B.l. Multifocus image fusion using the nonsubsampled contourlet transform. *Signal Process.* **2009**, *89*, 1334–1346. [CrossRef]
27. Chen, C.I. Fusion of PET and MR brain images based on IHS and log-Gabor transforms. *IEEE Sens. J.* **2017**, *17*, 6995–7010. [CrossRef]
28. Jian, L.; Yang, X.; Liu, Z.; Jeon, G.; Gao, M.; Chisholm, D. SEDRFuse: A symmetric encoder-decoder with residual block network for infrared and visible image fusion. *IEEE Trans. Instrum. Meas.* **2020**, *70*, 1–15. [CrossRef]
29. Wang, X.; Jiang, J.; Gao, M.; Liu, Z.; Zhao, C. Activation ensemble generative adversarial network transfer learning for image classification. *J. Electron. Imaging* **2021**, *30*, 013016. [CrossRef]
30. Song, W.; Zhai, W.; Gao, M.; Li, Q.; Chehri, A.; Jeon, G. Multiscale aggregation and illumination-aware attention network for infrared and visible image fusion. *Concurr. Comput. Pract. Exp.* **2024**, *36*, e7712. [CrossRef]
31. Song, W.; Gao, M.; Li, Q.; Guo, X.; Wang, Z.; Jeon, G. Optimizing Nighttime Infrared and Visible Image Fusion for Long-haul Tactile Internet. *IEEE Trans. Consum. Electron.* **2024**, *70*, 4277–4286. [CrossRef]
32. Liu, J.; Lin, R.; Wu, G.; Liu, R.; Luo, Z.; Fan, X. Coconet: Coupled contrastive learning network with multi-level feature ensemble for multi-modality image fusion. *Int. J. Comput. Vis.* **2024**, *132*, 1748–1775. [CrossRef]
33. Mu, P.; Wu, G.; Liu, J.; Zhang, Y.; Fan, X.; Liu, R. Learning to Search a Lightweight Generalized Network for Medical Image Fusion. *IEEE Trans. Circuits Syst. Video Technol.* **2023**, *34*, 5921–5934. [CrossRef]
34. Li, J.; Liu, J.; Zhou, S.; Zhang, Q.; Kasabov, N.K. Gesenet: A general semantic-guided network with couple mask ensemble for medical image fusion. *IEEE Trans. Neural Netw. Learn. Syst.* **2023**. [CrossRef] [PubMed]
35. Ma, J.; Yu, W.; Liang, P.; Li, C.; Jiang, J. FusionGAN: A generative adversarial network for infrared and visible image fusion. *Inf. Fusion* **2019**, *48*, 11–26. [CrossRef]
36. Xu, H.; Ma, J.; Zhang, X.P. MEF-GAN: Multi-exposure image fusion via generative adversarial networks. *IEEE Trans. Image Process.* **2020**, *29*, 7203–7216. [CrossRef]
37. Maas, A.L.; Hannun, A.Y.; Ng, A.Y. Rectifier nonlinearities improve neural network acoustic models. In Proceedings of the International Conference on International Conference on Machine Learning (ICML), Citeseer, Atlanta, GA, USA, 16 June–21 June 2013; Volume 30, p. 3.
38. Johnson, K.A.; Becker, J.A. The Whole Brain Atlas database of Harvard Medical School. Available online: http://www.med.harvard.edu/aanlib/home.html (accessed on 15 May 2023).
39. Parekh, A.; Patil, N.; Biju, R.; Shah, A. Multimodal Medical Image Fusion to Detect Brain Tumors. 2020. Available online: https://github.com/ashna111/multimodal-image-fusion-to-detect-brain-tumors (accessed on 15 May 2023).
40. Bavirisetti, D.P.; Kollu, V.; Gang, X.; Dhuli, R. Fusion of MRI and CT images using guided image filter and image statistics. *Int. J. Imaging Syst. Technol.* **2017**, *27*, 227–237.
41. Da, K. A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
42. Eskicioglu, A.M.; Fisher, P.S. Image quality measures and their performance. *IEEE Trans. Commun.* **1995**, *43*, 2959–2965. [CrossRef]
43. Wang, Z.; Li, Q. Information content weighting for perceptual image quality assessment. *IEEE Trans. Image Process.* **2010**, *20*, 1185–1198. [CrossRef]
44. Mukaka, M.M. A guide to appropriate use of correlation coefficient in medical research. *Malawi Med J.* **2012**, *24*, 69–71.
45. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [CrossRef]
46. Han, Y.; Cai, Y.; Cao, Y.; Xu, X. A new image fusion performance metric based on visual information fidelity. *Inf. Fusion* **2013**, *14*, 127–135. [CrossRef]
47. Qu, G.; Zhang, D.; Yan, P. Information measure for performance of image fusion. *Electron. Lett.* **2002**, *38*, 313–315. [CrossRef]