



Feature extraction of multimodal medical image fusion using novel deep learning and contrast enhancement method

Jameel Ahmed Bhutto¹ · Jiang Guosong¹ · Ziaur Rahman¹ · Muhammad Ishfaq¹ · Zhengzheng Sun² · Toufique Ahmed Soomro³

Accepted: 29 March 2024 / Published online: 1 May 2024

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2024

Abstract

The fusion of multimodal medical images has garnered painstaking attention for clinical diagnosis and surgical planning. Although various scholars have designed numerous fusion methods, the challenges of extracting substantial features without introducing noise and non-uniform contrast hindered the overall quality of fused photos. This paper presents a multimodal medical image fusion (MMIF) using a novel deep convolutional neural network (D-CNN) along with preprocessing schemes to circumvent the mentioned issues. A non-linear average median filtering (NL-AMF) and multiscale improved top-hat (MI-TH) approach are utilized at the preprocessing stage to remove noise and improve the contrast of images. The non-linear anisotropic diffusion (NL-AD) scheme is employed to split the photos into base and detailed parts. The fusion of base parts is accomplished by a dimension reduction method to retain the energy information. In contrast, the detailed parts are fused by novel D-CNN to preserve the enriched detailed features effectively. The simulation results demonstrate that the proposed method produces better brightness contrast and more image details than existing methods by acquiring 0.7649 to **0.8986**, 0.3520 to **0.4783**, 0.7639 to **0.9056**, 68.8932 to **81.0487** gain for quality transfer ratio from source photo to a generated photo (Q_G^{AB}), feature mutual information (FMI), structural similarity index (SSIM), and average pixel intensity (API) respectively.

Keywords Image fusion · Feature extraction · Convolution neural network · Computed tomography · And magnetic resonance imaging

1 Introduction

The MMIF has received painstaking attention due to its vast applications that assist doctors in diagnosing and treating diseases accurately [1]. It has opened a new research path for various medical applications, including Alzheimer's disease (AD), coronary artery heart disease, COVID-19 detection, brain tumor disease [2], and so on [3]. The modern advancements in multispectral, high-resolution, reliable, and cost-effective image sensor design technology are the primary

motivation for the inspiration in MMIF research [4]. With the introduction of these multisensory imaging techniques in the last few decades, MMIF has been an emerging field of research worldwide. Most medical diagnoses are made by computers or doctors who look at patient photos. Medical images are produced using a variety of imaging modalities, each with a unique imaging mechanism and emphasis on describing the human body [5]. Computed tomography (CT) photos are utilized to identify bone fractures, tumor locations, cardiac tissues, pulmonary emboli, skulls, and brain lesions but are unable to reveal information about the soft tissues. On the other hand, magnetic resonance imaging (MRI) photos provide excellent data regarding the soft tissues and blood flow of the brain but cannot record details regarding the hard bones and activity of the brain [6, 7].

Each modality provides doctors with limited medical information, requiring additional time and effort to diagnose the patient's diseases. In addition, gathering data from multiple modalities can lead to the potential loss of critical diagnostic information regarding the patient's

✉ Jiang Guosong
jiangguosong@hgnu.edu.cn

¹ Department of Computer Science, Huanggang Normal University, Huanggang 438000, China

² School of Automation Science and Engineering, South China University of Technology, Guangzhou 510640, China

³ Electronic Engineering Department, University of Larkano, Larkana, Sindh, Pakistan

ailment, resulting in diminished diagnostic precision. The ideal way to address this issue is to generate a single photo by merging multiple modality photos of the client's body from the same spot, referred to as MMIF. The MMIF is a process of feature extraction from multiple photos to generate one output photo that preserves all salient features. It not only boosts the efficiency of medical professionals to save time by minimizing their workload but also facilitates accurate diagnosis of medical diseases [8].

Most MMIF methods typically involve three steps: first, the photo pixels are mapped into the transform domain, then fusion is performed on the transform coefficients, and finally, the fused photo is produced by applying the inverse transformation to fused coefficients. The basic procedure for MMIF is presented in Fig. 1.

The photo registration is employed to geometrically well-align two photos, and then fusion can be performed on both photos. The fusion is used by overlapping the multiple source photos to obtain one fused photo with complementary features. The ideal fusion is achieved when all the substantial features in the source photos are present in the fused photo, and the fused photo does not contain extra information that is not available in the source photos.

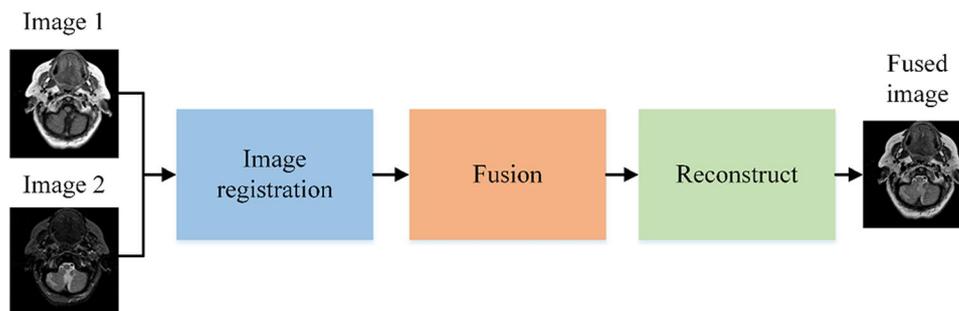
The MMIF is achieved at the pixel, feature, and decision levels based on specific applications and the nature of the input photos [9]. The most used method is pixel-level, which is accomplished by directly mapping the pixels of multiple photos. On the other side, the feature-level method is achieved by retrieving relevant features by amalgamating features from both source photos to create new features with more substantial information than any of the individual source photos. The region-based approach is one of the common examples of feature-level fusion, which numerous scholars employ. The decision-level approach is performed by acquiring substantive features from individual source photos and then determining to amalgamate the retrieved features depending on the defined specifications. This approach is accomplished by two phases: retrieving features and then amalgamating the features. First, it retrieves all the features obtained by all decisions from each photo using defined criteria. Then,

the final decision is made to amalgamate the significant features acquired from all decisions [10].

The main motive of any MMIF scheme is to generate a single output photo that retrieves all substantial features (energy information and precise detailed features such as textures, edges, and boundaries) from the multiple source photos without adding noise. A lot of MMIF methods have been implemented so far by scholars, which are broadly categorized as multiscale decomposition (MSD) [11], subspace-based, neural network (NN), sparse representation (SR), deep learning (DL) [12], and hybrid methods. Subspace-based methods such as principal component analysis (PCA), independent component analysis (ICA) [13], and robust PCA have been used, but these schemes fail to retain spectral features. The MSD schemes such as discrete wavelet transform (DWT) [14], discrete stationary wavelet transform (DSWT), contourlet transform, non-subsampled contourlet transform (NSCT), and non-subsampled shearlet transform (NSST) restore satisfactory spectral data (edges, contours, textures, and boundaries) but they unable to retain spatial information [15]. The edge-preserving schemes, in particular, cross bilateral filtering (CBF), guided filtering, and isotropic diffusion can achieve reasonable information, but the fused photos are affected by artifacts and halo effects. Numerous scholars have designed the MMIF method using SR, for instance, orthogonal matching pursuit (OMP), over-complete dictionary (OCD), and simultaneous OMD, but these schemes have limited ability to retain detailed features [16]. The role of deep learning algorithms in the MMIF has gotten noteworthy attention due to their outstanding performance for spectral data, but these models still lack retaining spatial information. Moreover, the direct implication of deep learning schemes such as convolutional neural network (CNN), visual geometry group (VGG-19), and residual networks (ResNet) in the spatial domain resulting substantial loss in energy information [17].

The prime motive of any multimodal medical image fusion method is to produce a single fused photo that can retain all substantial features from multiple input photos with enhanced contrast without introducing noise and artifacts. It is well-known that medical images are highly distorted by noise due to various factors, including camera circuitry,

Fig. 1 The diagram of MMIF



outside environment, and improper lighting. In addition, non-uniform contrast is another challenging task that is caused by varying contrast and uneven lighting conditions. Moreover, retaining substantial features from both images is another crucial task because the image contains spatial as well as spectral information, and one approach cannot preserve complementary information. Furthermore, the fusion of multiple photos is also the prior goal to acquire a high-quality final fused photo. Numerous scholars have applied novel fusion approaches, but their methods only focus on specific parts instead of considering all crucial issues, resulting in an image with noise, varying contrast, improper fusion strategy, and insufficient spatial or spectral features. In this regard, our proposed method addresses the technical gaps of recent works by presenting a suitable method for each specific issue to mitigate the described shortcomings. To the best of our investigation, this is the initial effort towards designing multiple new approaches for circumventing each specific issue.

The non-linear average median filtering (NL-AMF) scheme at the preprocessing stage is employed in this paper for noise reduction, which has never been used by research in the field of image fusion. This combined NL-AMF incorporates a statistical histogram that adaptively shrinks the filtering mask based on noise level to achieve a desirable reduction in noise while also reducing the complexity. The acquired image is then processed by another preprocessing multiscale improved top-hat (MI-TH) scheme that uses different-sized multiscale structuring elements that properly retrieve the region of interest (ROI), thereby producing an image with enhanced contrast. After that, the non-linear anisotropic diffusion (NL-AD) method is employed to partition the base and detail parts effectively through its intra-smoothing procedure. It is a vital objective of any fusion approach to retain substantial detailed featured information, which lies in spectral parts of an image, and deep learning has gained tremendous attention to capture substantial detailed features. This paper implements a novel deep convolutional neural network (D-CNN) that uses deconvolution layers and skipping convolution layer that has the potential to preserve all substantial detailed features without losing colors or patterns from images. In contrast, the base parts are fused by principal component analysis (PCA), which is proven to be an effective approach for preserving energy information without introducing collision.

Finally, the base parts and detailed parts are fused by the inverse transformation of the NL-AD approach. The proposed approach outclasses modern and state-of-the-art (SOTA) methods in qualitative evaluation through experts' human visual experience and quantitative analysis by mathematical computations.

The essential contributions of this research are emphasized as follows:

- The amalgamated NL-AMF mitigates the noise by adjusting the filter size according to the noise level, while the unique characteristics of the statistical histogram speedup the searching process for finding median values that reduce the computation time.
- The MI-TH using two different-sized structuring elements is employed to retrieve a substantial region of interest (ROI). Then, NL-AD using forward time central space (FTCS) is deployed to retain the significant energy information (base parts) and enriched detailed features (detailed parts).
- The detailed parts are fused by a novel D-CNN using skipping convolution and deconvolution layers, which use direct mapping and the focus map to shield the enriched detailed features without losing colors or other crucial details.
- The base parts are fused by a dimension reduction approach that projects high-dimensional data into low-dimension data while retrieving the energy information at high-resolution photo quality. The ultimate fused photo is retrieved through the superposition of the base part and detailed parts.

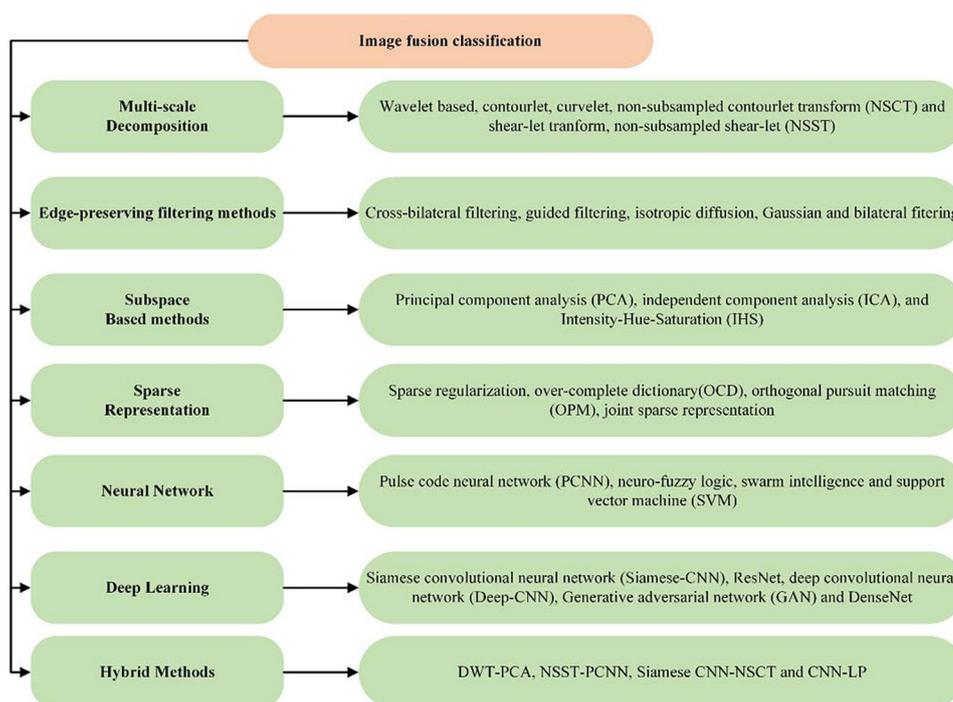
The rest of this article is arranged as follows. Section 2 highlights the detailed literature work for recent and SOTA methods. Section 3 presents our proposed methodology. The experimental results are reported in Section 4. The discussion of our findings and the conclusion of this article are elaborated in Section 5 and Section 6, respectively.

2 Related work

The MMIF has received significant attention due to its growing demands in medical applications. Numerous scholars have put their efforts in the last two decades to design algorithms that can be successfully implemented in hospitals to diagnose clinical diseases precisely. The mainstream MMIF is organized into seven classes, each with several types presented in Fig. 2. Various articles have been published regarding MMIF. The literature work of classical and recent methods is presented in Fig. 2.

The DWT approach is utilized in [18], which splits the input photos into low frequency (LF) and high frequency (HF) to retain the coarse energy information and detailed features. Nonetheless, the decimation process in down-sampling distorts photo details. Another scholar in [19] has implemented discrete cosine transform (DCT) that uses correlation parameters for determining the extent of changes in the fused photo followed by LF filtration. These approaches (DWT, improved DWT, DCT, and DSWT) retain better spectral data than spatial domain approaches. Nevertheless,

Fig. 2 MMIF classification and its sub-categories



these schemes still fail to retain all substantial features due to limited directional information.

The non-subsample shearlet transform (NSST) is pinpointed in [20], which avoids the discarding procedure, and it uses an averaging fusion scheme for LF parts while larger absolute values fuse HF components. This multi-resolution and multi-directional method retrieves the enriched detailed contours, textures, and edge information in the acquired photo. Another approach using NSCT and stacked sparse auto-encoder (SSAE) is designed in [21]. The source photos are decomposed using the NSCT method, and then SSAE is implemented for feature extraction of high-frequency (HF) images. In contrast, the low-frequency (LF) images are fused by the maximum fusion strategy. This method introduces artifacts and uneven illumination. A new decomposition approach named weighted fast discrete curvelet transform (W-FDCT) along with optimized Type-2 fuzzy entropy is developed in [22]. This method splits source photos into LF and HF by W-FDCT, and then an averaging scheme is applied for the fusion of LF. The HF photos are fused by the optimized Type-2 fuzzy entropy approach, which is good at capturing directional information. Though these methods (NSCT, W-FDCT, and NSST) capture sharp edges and textures with multi-directional information, the fused photos have insufficient spatial information.

Edge-preserving multiscale decomposition methods have also received noteworthy consideration in the last decade. These filtering methods segment the photo into the base and detail parts. The base part precisely produces the impactful changes in the intensity, and the detailed layer contains the

series of photos having adjustable resolution. In [23], the researcher has proposed a cross-bilateral filtering (CBF) scheme in which weight calculation entails measuring the photo strength by subtracting the input photo. Weights are then directly multiplied and normalized to acquire the output photo. Numerous other edge-preserving methods have been used, but the artifacts and halo effects around the edges are generated that distort the photo quality.

The SR algorithms have been deployed in MMIF to learn an over-complete dictionary (OCD) from vivid images, and then a trained dictionary sparsely presents the source photos. These methods utilize a sliding window approach to partition photos into several overlapping patches to mitigate the artifacts. The amalgamated convolutional sparse representation (CSR) is implemented in [24], which splits the photo into the base and detailed sub-photos. The choose-max scheme fuses the base parts, while the detailed sub-photos are fused by sparse coefficients, which are acquired by CSR. This approach potentially retrieves enriched details than simple SR methods. Numerous efforts have been made to implement SR methods for producing a fused photo with a substantial feature. However, these approaches are highly sensitive for dictionary selection, which is a crucial challenge and requires huge computational requirements.

The hybrid methods have played a substantial role in enhancing the quality of fused photos in the last two decades. Authors in [25] present a hybrid multimodal medical image fusion approach in which they utilized the Laplacian pyramid (LP), sum-modified Laplacian (SML), and sparse representation (SR). The LP in this work acts as a decomposition

of images into low-frequency (LF) and high-frequency (HF) photos. Then, SR is used to fuse the LF photos, while SML is applied to fuse the HF photos. Another hybrid algorithm is designed in [26], which utilizes non-subsampled contourlet transform (NSCT) and guided filtering. The source photos are split into LF and HF by NSCT. After that, the LF images are fused by the average fusion strategy, while the HF photos are fused by employing a deep-guided filter approach. However, this approach is unable to provide detailed information due to non-uniform illumination, and noise is generated due to the environment and camera circuitry. The non-subsampled shearlet transform (NSST), convolutional sparse representation (CSR), and multiple fusion strategies are highlighted in [27]. The photos are decomposed in LF and HF using the NSST. Then, CSR and alternating direction multiplier method (ADMM) are utilized to train the LF and HF. The LF images are fused by combining the regional energy method and average Norm method, while the HF photos are fused by spatial frequency (SF) and average Norm method. This method has the potential to increase contrast, but the final fused image is distorted due to artifacts that are caused by complicated fusion strategies. Another NNST-based approach using CSR and multiple fusion strategies is designed in [28]. The input photos are first split into LF and HF; then, HF photos are fused by using the CSR approach. In contrast, the LF parts are fused by a mutual information correlation strategy. The threshold is set, and then the maximum fusion or weighted average fusion scheme is applied based on specific criteria. The latent low-rank representation (LatLRR) is deployed in [29], which uses LatLRR to divide source photos into saliency parts and low rank. After that, the sum strategy fuses the saliency parts, while a simple averaging scheme fuses low-rank parts. An amalgamated method is deployed in [30], which employs Discrete wavelet transform (DWT), curvelet transform, maximum fusion method, and the principal component analysis (PCA). First, the input photos are split into LF and HF using the DWT approach, and then LF parts are fused using the maximum fusion rule. The HF parts are fed to curvelet transform and then PCA fusion method is applied. This method is complicated and unable to provide sufficient information. In [31], the authors implemented another hybrid approach using DWT and convolution neural networks (CNN). The DWT approach acts as a decomposition of source photos, and then these decomposed photos are fed to CNN for feature extraction. Finally, the averaging scheme for fusion is utilized. This method fails to capture energy information, and noise is introduced in the final fused photo.

Recently, hybrid algorithms using deep learning (DL) models have become a prominent research area of interest in the field of MMIF due to their outperforming results [32]. Numerous renowned scholars have diverted their research from other domains to DL due to its ever-growing demands in MMIF

[33, 34]. Another approach is implemented in [35], which uses average weighting fusion for LP parts while high HF parts are extracted using ResNet-512. These multi-layer features are processed and fused by multiple strategies (regularization, bilinear interpolation, and choose-max) to acquire the highest weight layers, which are then multiplied by HF to generate new HF. Then, superposition is applied to reconstruct the LF and new HF parts to get a final fused photo. Another fully connected CNN algorithm in [36] is developed to decompose the photo into LF and HF by local non-subsampled shearlet transform (LNSST). CNN extracts the HF images to get weight maps, and these weight maps are fused by average gradient, while LF photos are fused by local energy.

The authors in [37] have designed a deep convolutional neural network for the restoration of the photos. This work establishes the connection between a traditional optimization-based scheme and a neural network architecture. A separable structure is utilized for robust support for reliable deconvolution to mitigate the noise and artifacts. In this paper, a deep convolution neural network is used for a single image that is blurred or degraded. The network architecture is designed to efficiently handle the process of restoring images from their blurred or degraded versions. By leveraging the hierarchical features learned through multiple layers, this method aimed to achieve superior performance in image restoration tasks compared to traditional approaches. The authors in [38] introduced a novel framework that combines Bayesian methods with multiscale convolutional neural networks (CNNs) to accurately predict local stress fields in structures containing microscale features. This approach aimed to address the challenges of predicting complex stress distributions in materials with heterogeneous microstructures by leveraging both the predictive power of CNNs and the uncertainty quantification capabilities of Bayesian methods.

Another CCN-based algorithm using hybrid optimization dynamic (HOD) is designed in the paper [6], in which photos are split into LF and HF parts. The split photos are fed to CNN for feature extraction, and HOD is utilized to improve the accuracy of weights. The authors in [10] have used the LatLRR approach to obtain various saliency parts and one low-rank part, and then VGG-19 is employed for feature extraction to generate weight maps. The low-rank parts are fused by making Hadamard products of weight maps and source photos, while saliency parts are fused by choosing the max rule. Table 1 presents the overall summary of existing methods.

3 Proposed methodology

The essential goal of any MMIF method is to merge multiple photos to generate a single output photo that possesses all the substantial detailed features and energy information

Table 1 The summary of the state-of-the-art methods in literature work

Categories	Sub-categories	Fusion method	Disadvantages
Subspace-based	Independent component analysis (ICA), PCA, Intensity hue saturation (IHS)	Averaging, weighted averaging, and choose-max	Retains insufficient spectral (edges, contours, textures) features
Multiscale decomposition (MSD)	DCT, DWT, DSWT, contourlet, NSCT and NSST	Activity level-based, region-based, weighted averaging, and averaging	It cannot achieve adequate spatial (energy) information
Edge-preserving filtering	CBF, guided filtering, Gaussian and bilateral and isotropic	Window-based, gradient, local energy, and choose-max	It generates halo effects and artifacts
Sparse representation	Joint sparsity, OMP, over-complete dictionary, and simultaneous OCD	Choose-max, sparse coefficients substitution, and averaging	It cannot produce sufficient detailed information and is highly sensitive for the selection of a dictionary
Deep learning	CNN, Residual network, VGG-19, Siamese network, and fully convolution network (FCN)	Normalization, gradient, local similarity method, and averaging	A significant loss in spatial information
Hybrid methods	DWT-PCA, contourlet with PCA, LSIST with CNN, anisotropic with VGG-19, and convolution with SR	Averaging, activity level, normalization, average gradient, and choose-max	Noise is generated, misalignment, poor contrast, and uneven lighting

from given photos without adding noise and artifacts. The hybrid algorithms using CNN have taken key attention in the MMIF area due to their outstanding performance in generating fused photos. Nonetheless, these approaches are still lacking due to non-uniform contrast and uneven lighting conditions. Furthermore, the appropriate extraction of detailed features and energy information during decomposition is another rigorous task that distorts the overall quality of the fused photo. In addition, noise is another crucial challenge in medical photos, which is added due to imaging mechanism sensors and environmental effects. In this regard, this work fills the technical gaps of recent studies by designing a novel hybrid method that circumvents the mentioned flaws. The proposed method in this work comprises several stages, as shown in Fig. 3.

The vital objective of each stage is to improve the quality of the photo by performing a particular task. Each stage of the proposed method is discussed in detail in the following subsections.

3.1 Amalgamated non-linear average median filtering (NL-AMF)

The noise is a crucial challenge that is introduced due to abrupt variations in lighting, camera circuitry, and environmental changes. Some scholars have used different approaches to reduce the noise effect, but their approaches increase the complexity. There is a need to design a technique that reduces the noise and decreases the complexity. This work presents a new amalgamated NL-AMF approach at the preprocessing stage, which also employs a statistical histogram to mitigate the noise with its fast computation. It is well-known that better photo details are achieved if the mask size is smaller, but noise cannot be removed effectively.

In contrast, if the mask size is larger, the photo details cannot be restored effectively, but the noise is effectively mitigated. This paper addresses the issue by introducing an adaptive filtering approach that resizes the mask based on the noise level. The noise reduction accomplished using filtration primarily depends on the shape and length of the mask. In contrast, the computation complexity is controlled by the rate at which it can locate the median value. The combined NL-AMF that incorporates statistical histogram adaptively shrinks the filtering mask based on noise level to achieve an exceptional reduction in noise, thereby generating processed photos with precise details.

Let n be the mask size in which \max is the maximum value of gray level intensities, avg is the average value of gray levels, \min is the minimum value of gray level intensities, med is the median value, and $f(i,j)$ is the central mask value. Then, the adaptive filtering is acquired in two stages. The first stage is to resize the mask and initialize

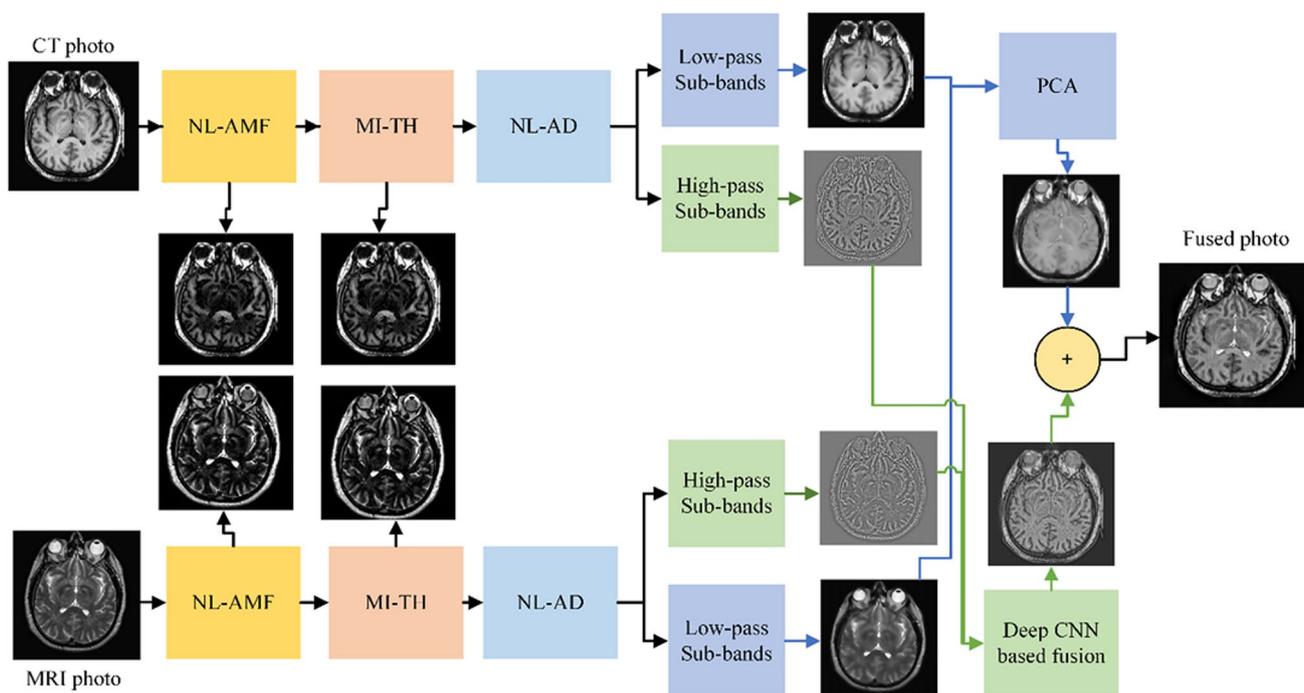


Fig. 3 The block diagram for the proposed MMIF

this mask by $n=3$. Then we compute $A_1 = med - min$ and $A_2 = med - max$. After that, verify if $A_1 > 0$ and $A_2 < 0$ then employ the second step of applying median filtering to compute median values; however, if any of these conditions are not correct, then increase the filter size by $n = n + 2$.

Figure 4 presents the flow chart that further helps in understanding this working phenomenon. Each pixel is examined to reduce the noise effect. If the value of the pixel exceeds the average value in the mask, it is assumed to be distorted by noise, then substitute that pixel with the median value according to the mask’s length. On the other hand, we keep the value of the pixel unchanged if its value is less than or the same as an average value.

This approach not only retains the substantial details in the photo but also decreases the computation time due to the use of statistical histogram in the search process of finding median values. After substituting the original value of a pixel with its median value within the mask, the subsequent calculation of the average value can utilize the updated pixel value. Forming the iterative process not only reduces the computation time but also improves the noise-mitigation effect.

3.2 Multiscale improved top-hat (MI-TH) transform

Contrast enhancement is one of the most critical challenges in medical photos because uneven lighting results in poor contrast in a photo. Numerous scholars have deployed the

Top-hat-bottom-hat (THBH) approach for medical photos due to its satisfactory results. Nonetheless, this approach uses only a single structuring element, which still fails to produce a desirable contrast in the photo. This paper designs a novel MI-TH approach that uses two distinct size structuring elements referred to as inner structuring elements (S_i) and outer structuring elements (S_o). These two structuring elements (S_i and S_o) are formed, whereas the size of S_o is typically larger than S_i . The MI-TH using improved white top-hat (MI-WTH) and improved black top-hat (MI-BTH) are calculated by Eqs. 1 and 2:

$$MI - WTH(x, y) = f(x, y) - ((f \oplus \Delta S) \ominus S_b) \tag{1}$$

$$MI - BTH(x, y) = ((f \ominus \Delta S) \oplus S_b) - f(x, y) \tag{2}$$

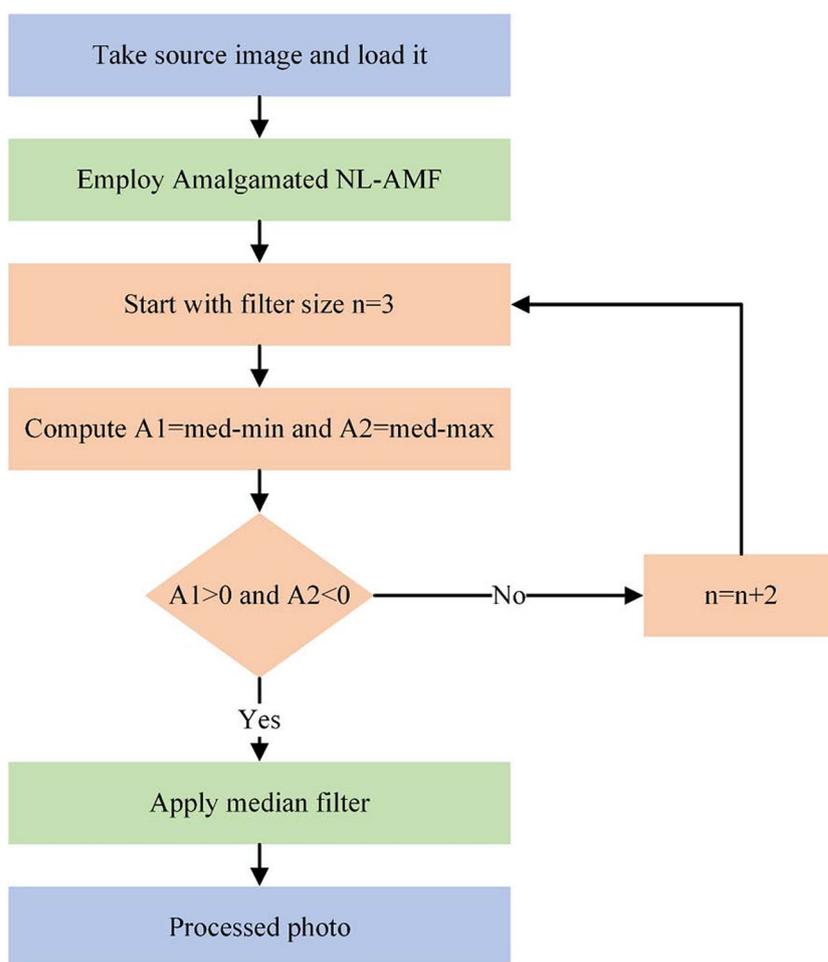
where $\Delta S = S_o - S_i$ indicates the marginal structuring element for the region between ROI and the surrounding region, whereas S_b represents a region that is in the ROI. To eliminate the potential negative values, Eqs. 1 and 2 are further derived to acquire negative gray levels, which are computed by Eqs. 3 and 4:

$$MI - WTH(x, y) = f(x, y) - \min((f \oplus \Delta S) \ominus S_b, f(x, y)) \tag{3}$$

$$MI - BTH(x, y) = \max((f \ominus \Delta S) \oplus S_b, f(x, y)) - f(x, y) \tag{4}$$

The key purpose of MI-WTH and MI-BTH is to retrieve the bright and light regions, respectively, and MI-TH can achieve it

Fig. 4 Flowchart of NL-AMF



with the incremental size of structuring elements. Let n scales of structuring elements be determined, and the size of every scale s is $(1 \leq s \leq n)$ then the size of the square structuring element for each scale (nL_s) and the size of square marginal structuring element for each scale (nW_s) are obtained by Eqs. 5 and 6, respectively.

$$nL_s = nL + s \times nS \tag{5}$$

$$nW_s = nW + s \times nS \tag{6}$$

where nS indicates the incremental in size for the n th scale. The MI-TH executes the extraction for regions of size until $nW - 2 \times nM$ from the ΔS and S_b . The nM provides the margin size in the ΔS .

Therefore, the dark and light medical photo regions are retrieved for MI-WTH and MI-BTH at scale s are calculated by Eqs. 7 and 8:

$$MI - WTH(x, y) = f(x, y) - \min((f \oplus \Delta S_s) \ominus S_{b_s}, f(x, y)) \tag{7}$$

$$MI - BTH(x, y) = \max((f \ominus \Delta S_s) \oplus S_{b_s}, f(x, y)) - f(x, y) \tag{8}$$

The size of S_{b_s} and ΔS_s are determined based on nL_s and nW_s respectively. Therefore, the dark and light regions are accomplished from the retrieved multiscale photo regions for final medical images by Eqs. 9 and 10 as:

$$RW = \max(MI - WTH_1, MI - WTH_2, \dots, MI - WTH_n) \tag{9}$$

$$RB = \max(MI - BTH_1, MI - BTH_2, \dots, MI - BTH_n) \tag{10}$$

where R indicates the region. The next step is to improve the contrast between dark and bright regions, which is accomplished by powered photo enhancement (E) by Eq. 11:

$$E = f \times w_1 + RW \times w_2 - RB \times w_3 \tag{11}$$

where f is the input photo, E is the enhanced photo, and w_1, w_2 and w_3 are the weight for adjusting the contrast enhancement. It will brighten the light regions than the original photo by enlarging the w_2 and w_3 .

3.3 Non-linear Anisotropic Diffusion (NL-AD)

The source photo contains energy information and detailed features such as textures, contours, edges, and boundaries. The appropriate fusion scheme is employed for each part because a single fusion strategy cannot provide energy information and detailed features. The NL-AD multiscale decomposition (MSD) approach is utilized in this paper to split each source photo into the base (energy information) part and the detailed (features, textures, edges, and contours) part. The NL-AD holds considerable potential for preserving photo features while eliminating noise. The proposed approach utilizes an intra-region smoothing procedure and the use of partial differential equations (PDE) to retain sharp textures while also precisely smoothing out homogenous regions. The NL-AD for photo diffusion is computed by Eq. 12, which utilizes the flux function:

$$P_t = R(x, y, t)\Delta P + \nabla R \cdot \nabla P \tag{12}$$

In Eq. 12, P_t indicates the photo diffusion, Δ = gradient operator, Δ = Laplacian operator, $R(x, y, t)$ = diffusion rate, and t is the number of iterations. It is vital to keep the stability of diffusion, and this can be achieved using forward time central space (FTCS), which is accomplished by maximum principle [39]. It can be computed by Eq. 13:

$$P_{ij}^{t+1} = P_{ij}^t + \lambda[R_N \cdot \bar{\nabla}_N P_{ij}^t + R_S \cdot \bar{\nabla}_S P_{ij}^t + R_W \cdot \bar{\nabla}_W P_{ij}^t + R_E \cdot \bar{\nabla}_E P_{ij}^t] \tag{13}$$

The P_{ij}^{t+1} shows coarse resolution of photo in the forward time $t + 1$, λ indicates the stability constant in the value of $0 \leq \lambda \leq 0.25$. The $\bar{\nabla}_W, \bar{\nabla}_S, \bar{\nabla}_E$ and $\bar{\nabla}_N$ represents the neighboring derivates in all four directions. Equation 13 can further be computed by Eq. 14 as follows:

$$\left\{ \begin{array}{l} \bar{\nabla}_N P_{ij} = P_{i-1,j} - P_{i,j}, \\ \bar{\nabla}_S P_{ij} = P_{i+1,j} - P_{i,j}, \\ \bar{\nabla}_E P_{ij} = P_{i,j+1} - P_{i,j}, \\ \bar{\nabla}_W P_{ij} = P_{i,j-1} - P_{i,j} \end{array} \right. \tag{14}$$

The flux functions are updated for all iterations in east, west, north, and south directions by Eq. 15:

$$\left\{ \begin{array}{l} R_{N_{ij}}^t = h(\|\nabla P\|_{i+1/2,j}^t) = h\left(\left|\bar{\nabla}_N P_{ij}^t\right|\right), \\ R_{S_{ij}}^t = h(\|\nabla P\|_{i-1/2,j}^t) = h\left(\left|\bar{\nabla}_S P_{ij}^t\right|\right), \\ R_{E_{ij}}^t = h(\|\nabla P\|_{i,j+1/2}^t) = h\left(\left|\bar{\nabla}_E P_{ij}^t\right|\right), \\ R_{W_{ij}}^t = h(\|\nabla P\|_{i,j-1/2}^t) = h\left(\left|\bar{\nabla}_W P_{ij}^t\right|\right), \end{array} \right. \tag{15}$$

The R_N, R_S, R_E and R_W represent the flux functions. The $h(\cdot)$ is function which decreases monotonically with $h(0) = 1$. It can be calculated by Eqs. 16 and 17:

$$h(\nabla P) = e^{-\left(\frac{\|\nabla P\|}{z}\right)^2} \tag{16}$$

$$h(\nabla P) = e^{-\left(\frac{\|\nabla P\|}{z}\right)^2} \tag{17}$$

The z in Eqs. 16 and 17 indicate a free parameter that evaluates the validity of the boundary according to the intensity level of edges.

Let the source photo $P_n(x, y)_{n=1}$ with the size of $s \times q$ are fed to NL-AD process for obtaining the base layer, then it is computed by Eq. 18:

$$B_n(x, y) = A_{df}(P_n(x, y)) \tag{18}$$

In Eq. 18, the $A_{df}(P_n(x, y))$ presents the NL-AD procedure for $n - th$ base layer, and $B_n(x, y)$ shows the obtained base layer; then, the final detailed part is computed by Eq. 19:

$$D_n(x, y) = P_n(x, y) - B_n(x, y) \tag{19}$$

After splitting both source photos into the base and detailed parts by NL-AD, the next step is to apply a fusion strategy to both parts.

3.4 Feature extraction of base parts by dimension reduction method

The energy information is extracted from base layers using appropriate fusion strategies. Various authors have attempted to extract information from base layers by distinct approaches but failed to preserve sufficient energy information. This paper employs the dimension reduction-based fusion strategy, which not only preserves sufficient energy information but also reduces the complexity and addresses the overfitting issues. This approach transforms the correlated variables into independent ones, resulting in a concise representation from both base layers. Moreover, it is an expedient dimension reduction approach that swiftly chooses the highest Eigenvector as a principal component. Each component is mutually perpendicular to other components, aiding in removing redundant data. The procedure for a fusion of base layers by this method is elaborated as follows:

- Let $B_1(x, y)$ and $B_2(x, y)$ represent the two base layers that are achieved from the NL-AD. The column matrix Z is computed for these base layers. After that, the variance vr and covariance cr are computed from Z by Eq. 20:

$$Z = \begin{bmatrix} vr_1 & cr(1, 2) \\ cr(1, 2) & vr_2 \end{bmatrix} \tag{20}$$

$$cr = \frac{(\sum(x_1 - \bar{x}_1) \times (x_2 - \bar{x}_2))}{Tr - 1} \tag{21}$$

Here, Tr indicates the iteration number, vr indicates the variance, and cr indicates the covariance.

- The next step is to compute the Eigenvalues λ by Eq. 22:

$$Det|Z - \lambda I| = 0 \tag{22}$$

$$((vr_1 - \lambda) \times (vr_2 - \lambda)) \times cr(1, 2)^2 = 0 \tag{23}$$

Here Det is abbreviated for Determinant, while I is the identity matrix.

- The Highest Eigenvectors δ are computed after acquiring the Eigenvalues by Eq. 24:

$$\delta_1 = \begin{bmatrix} \delta_1(1) \\ \delta_2(2) \end{bmatrix} \text{ and } \delta_2 = \begin{bmatrix} \delta_1(1) \\ \delta_2(2) \end{bmatrix} \tag{24}$$

- Then, independent variables I_{c1} and I_{c2} corresponding to match highest $\lambda(\lambda_{max} = \max(\lambda_1, \lambda_2))$. Since λ_{max} is the highest Eigenvalues and δ_{max} is the highest Eigenvector, then I_{c1} and I_{c2} are acquired by Eq. 25:

$$I_{c1} = \frac{\delta_{max}(1)}{\sum_n \delta_{max}(n)}, I_{c2} = \frac{\delta_{max}(2)}{\sum_n \delta_{max}(n)} \tag{25}$$

- The last step is to compute the fusion of base layers which is mathematically obtained by Eq. 26:

$$\bar{B}(x, y) = I_{c1}B_1(x, y) + I_{c2}B_2(x, y) \tag{26}$$

The fusion of both layers preserves sufficient energy information by fast computation while also avoiding over-fitting issues.

3.5 Feature extraction of detailed parts by a novel Deep CNN

It is the prime motive of any fusion method to preserve all substantial detailed features extracted from the detailed parts. In recent years, numerous fusion strategies have been designed to extract features such as contours, edges, boundaries, and textures. Nevertheless, the produced photo is unable to preserve significant features. To the best of our knowledge, this paper is the first attempt that uses a novel deep convolution neural network (D-CNN) by employing the skipping convolutional layers and deconvolutional layers for enhancing the fusion effect and retaining the true colors, textures, and edges in the detailed parts. The proposed novel D-CNN uses convolutional layers, skipping convolution layers, pooling layers, rectified linear units (ReLU), and deconvolutional layers, as depicted in Fig. 5. The proposed model supports end-to-end learning, meaning that the entire network can be trained in a single optimization process. This simplifies the training procedure and allows the network to

learn effective feature representations specifically tailored to the multimodal medical image fusion task. Figure 5 shows that five stages are applied to extract multi-layer features. The prime purpose of using D-CNN is its potential to extract multi-layer features from detail parts with better visual perception. The other main reason for using this model is its feasibility in which we have set a fixed size of all convolutional layers 3×3 with a stride of 1, while the size of pooling layers is also fixed 2×2 with a stride of 2. The ReLU function is utilized after each convolutional layer to speed up the training, thereby saving time. The pooling layer is employed in stages 1 and 4 after the ReLU function. The pooling layers act as sub-sampling that decreases the size of features.

Since D-CNN retains the color, textures, and edges, however some edges, textures, and colors are lost during fusion. Therefore, two 1×1 convolutional skip layers are used to increase the fusion effect. The features that are sub-sampled by pooling layers are up-sampled by two deconvolutional layers to match the dimensions employed after stage 5. These multi-layer features are fused, which enhances the overall quality of fusion.

These enhanced features are then applied to the SoftMax operator, and the stochastic gradient descent (SGD) is used in this paper to optimize the loss functions. The softmax loss function is computed in this paper by 27:

$$\zeta = (y, \hat{y}) = - \sum_{i=1}^2 y_i \log(\hat{y}_i) \tag{27}$$

where $\hat{y} = (\hat{y}_1, \hat{y}_2)$ defines the output vector probability and $y = (y_1, y_2)$ shows the corresponding label vector of the class. The patch size is fixed to 128 with a momentum of 0.9, having a weight decay of 0.0005, while the loss function in the training phase is reduced by SGD. Therefore, the weight using updating rule is obtained by Eq. 28:

$$w_i + 1 = w_i(0.9.u_i - 0.0003.b.w_i - b. \frac{\partial \zeta}{\partial w_i}) \tag{28}$$

In Eq. 28, w indicates the weight, ζ defines the loss function, u shows the momentum, i presents the $i - th$ iteration, and β highlights the learning rate, while $\frac{\partial \zeta}{\partial w_i}$ is derivative of the SoftMax loss function. The Xavier approach is employed for all the convolutional layers to initialize the weight, with bias having an initialization value of 0. Finally, the mean gradient (MG) is utilized for the fusion of both detailed parts, which is computed by 29:

$$D(x, y) = \beta_A f(x, y) + \beta_B f(x, y) \tag{29}$$

where $\beta_A \beta_B$ are fused coefficients of the MG of detailed photos.

The reconstruction $F(x, y)$ of the base $B(x, y)$ and detailed $D(x, y)$ photos are achieved by 1.30:

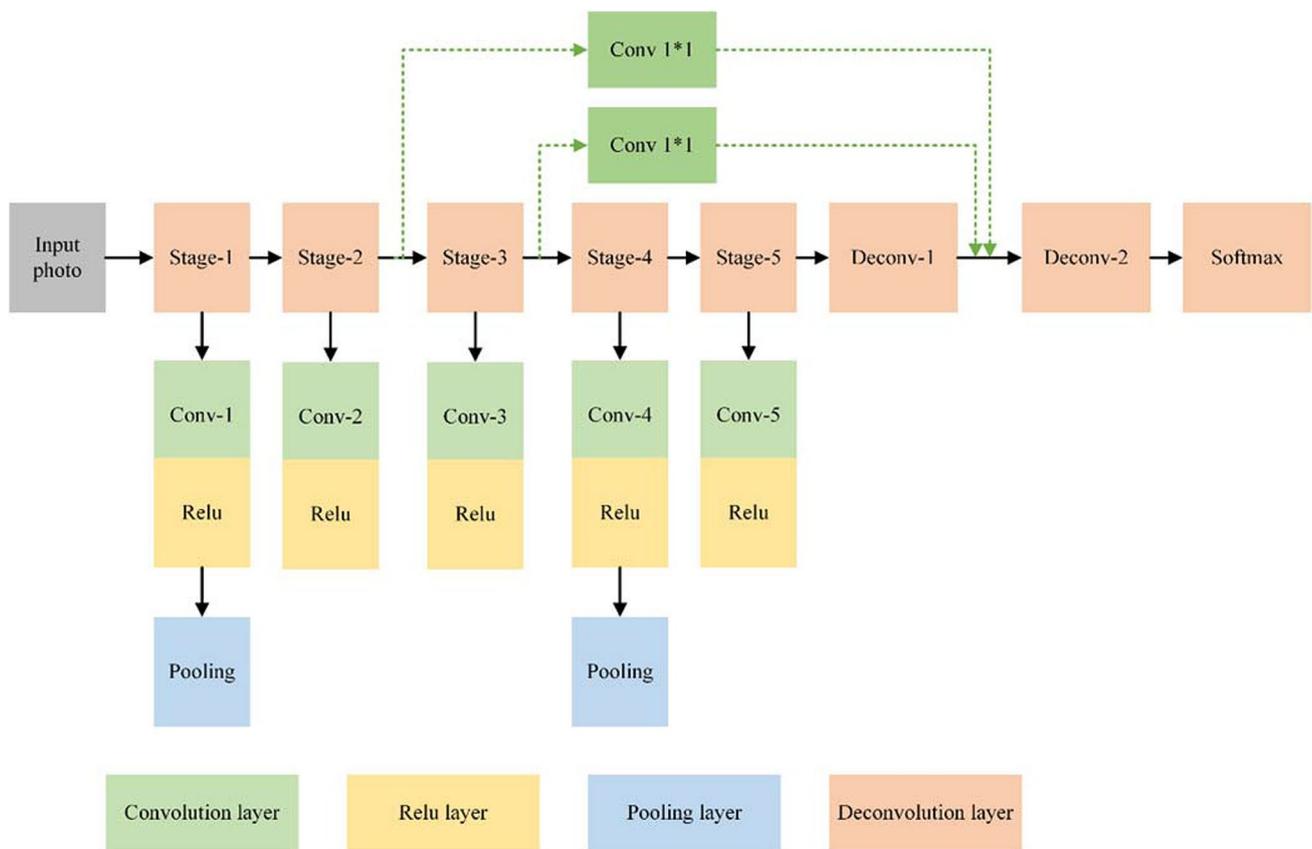


Fig. 5 The architecture of D-CNN for a fusion of detailed parts

$$F(x, y) = B(x, y) + D(x, y) \tag{30}$$

3.5.1 Parameters adjustment

In multimodal medical image fusion, the setting of parameters for deep learning has an important impact on training and testing the model. Therefore, it is important to set all the parameters appropriately. There are many parameters that need to be set, including batch size, epoch, balance parameters, learning rate, and so on.

- a. **Batch size:** The batch is part of the data that is used for training, and batch size is the total number of training samples for each batch. It is vital to choose the batch size carefully to maintain a balance between memory capacity and memory efficiency to optimize the speed and performance of the proposed model. The batch size should be set based on the actual situation of the memory, according to Leslie’s theory. By keeping this in mind, the batch size to 32.
- b. **Epoch:** The epoch term defines the number of times the learning algorithm works during the whole training set. The epoch value is related to the stability of the network

- c. **Learning rate, momentum, and weight decay:** In this work, the learning rate was adjusted using an adaptive strategy. By considering the D-CNN convergence and loss curves, proper adjustment was applied for the learning rate to avoid overfitting and provide optimal performance during training. The proposed model achieves optimal performance by setting a learning rate of 0.00001. In addition, the momentum and weight decay are set at 0.9 and 0.0005, respectively.

4 Experimental results, datasets, and evaluation of fusion schemes

Several experiments have been conducted on medical photos to prove the outstanding results of the proposed methods over SOTA methods. The qualitative and quantitative evaluation is carried out to compare all the methods with the proposed methods. Numerous scholars have used different quantitative approaches to evaluate the quality of fused photos, making it complicated to compare the different MMIF methods. By considering this flaw, this paper uses the same quantitative parameters used in comparing methods for the fair trial to demonstrate the supremacy of the proposed method. Five different medical imaging datasets for this paper are used, which are open access series of imaging studies (OASIS), image fusion database (IFD), BrainWeb Atlas (AANLIB), Alzheimer's disease neuroimaging initiative (ADNI), and medical image dataset annotation service (MIDAS). The images have been registered to produce numerous aligned cross-model pair photos and scanned into multimodal photos at a resolution of 256×256 . During the training set construction, we chose 100 pairs of functional and structural images from the mentioned five databases downloaded and cropped these photos into 10,230 patch pairs. The size of all structural patches and functional patches are set as 84×84 follows. This procedure not only guarantees the resilience of the training set but also ensures its variability. During the construction phase of a test set, 90 pairs of medical images from these databases are chosen as our test set.

Table 2 describes the datasets used for performing the simulation results of proposed methods with existing MMIF schemes. This work compares our methods with classical and recent MMIF methods, which are discrete cosine transform (DCT) [19], improved discrete wavelet transforms (DWT) [18], convolutional sparse representation (CSR) [24], Latent low-rank representation using VGG-19 (LatLRR-VGG) [10], and convolution neural network using hybrid optimization dynamic (CNN-HOD) [6]. All the experiments are performed using MATLAB 2021b on the 11th generation of core i5 2.40 GHz processor with 16 GB of RAM

4.1 Qualitative Evaluation (QE)

The qualitative evaluation is a useful way to compare all the methods by the human visual system (HVS). Various factors, including contrast, brightness, sharpness of edges, textures, and boundaries, assess the photo quality of each photo. Experiments are performed on ten pairs of MMIF photos, with a detailed discussion on the Quality Evaluation (QE) of each method.

Figure 6 shows the results for the first pair of source photos. It can be observed by HVS that the proposed methods in Fig. 6c retain clear edges and textures due to novel D-CNN compared to other SOTA methods. The proposed method also has better contrast due to the MI-TH approach, which uses different-size structuring elements to adjust the photo's contrast. Therefore, the proposed method retains better information regarding contrast, textures, edges, and precise tissues than all other methods. The DWT and DCT methods in Fig. 6g and h have more noise and blurred details than other methods, and it degrades the overall quality of a fused photo. The CSR method in Fig. 6f produces better results than DWT and DCT, but the edges are unclear, and it still has noise, so it cannot preserve the detailed edges and textures. The photo quality of CNN-HOD and LatLRR-VGG in Figs. 6d and e is more satisfactory than existing methods. However, the contrast of fused photos is not good, so it fails to distinguish the edges and soft tissue information. Moreover, these approaches cannot capture the significant textures due to noise introduced in the fused photo.

It can be analyzed in Fig. 7c that the proposed method produces outstanding quality photos than all other SOTA methods. The NL-AMF in the proposed method removes the noise by adaptively resizing the filter mask. Novel D-CNN preserves smooth edges and textures without loss of fusion effect, and the fused photo has good contrast due to the properties of the MI-TH method. The quality of CNN-HOD and LatLRR-VGG in Figs. 7d and e also produce better results, but the textures and edges are still not up to the mark as the proposed method. Moreover, it can be seen in Figs. 7d and e that some useful information is lost due to noise generated during the fusion process. The DCT method in Fig. 7h acquires the worst result than the other methods, and it is

Table 2 MMIF datasets for experiments

Dataset	Modalities	Body organ	Link for dataset
OASIS	MRI and PET	Brain	https://www.oasis-brains.org/
AANLIB	CT, MRI, PET, and SPECT	Brain	http://www.med.harvard.edu/AANLIB/home.html
ADNI	MRI, CT, and PET	Brain	http://adni.loni.usc.edu/
MIDAS	MRI, CT, SPECT, and PET	Brain, heart, head, liver, and bones	https://www.insight-journal.org/
IFD	CT and MRI	Brain and bones	http://www.imagefusion.org

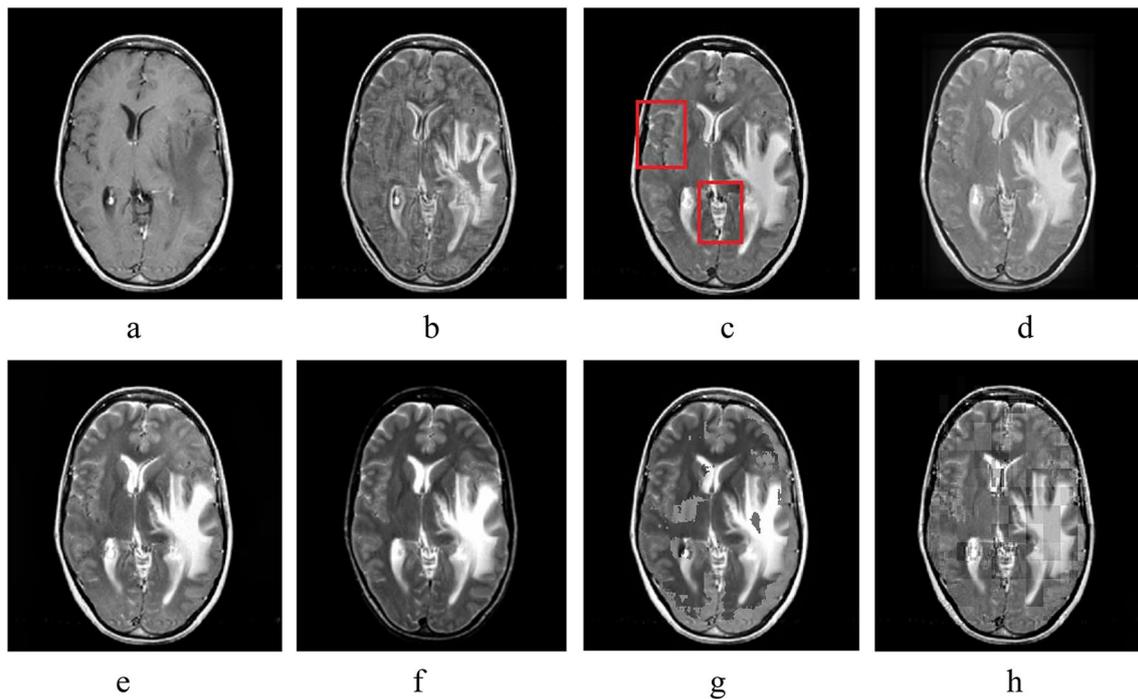


Fig. 6 Pair one: (a) CT image (b) MRI image (c) Proposed (d) CNN-HOD (e) LatLRR-VGG (f) CSR (g) DWT (h) DCT

highly affected by noise, which degrades the overall quality of the fused photo. The CSR and DWT in Figs. 7f and g preserve better results than DCT, but their performance is less than the other methods.

It can be depicted in Fig. 8c that the pixel consistency of the proposed method of the fused photo is the best, and the tissues are more precise than other SOTA methods. The quality of CNN-HOD and LatLRR-VGG also resembles, but these approaches still fail to distinguish between the textures and tissues in the photo. The DCT method in Fig. 8h cannot capture the edges and the textures, and the overall quality of the fused photo is worse than the existing methods. The CSR produces better textures and edges than CNN-HOD, LatLRR-VGG, and other existing methods.

The proposed method in Fig. 9c achieves its supremacy over SOTA methods by retaining all substantial detailed information from CT and MRI photos. The vivid contrast and noise-free fused photos are obtained due to the unique features of MI-TH and NL-AMF approaches. Moreover, due to the skipping and deconvolutional layers in novel D-CNN, the proposed method preserves significant patterns with smooth edges and boundaries. Therefore, the proposed method acquires the overall complement information from both source photos than all existing methods. The DCT method introduces more artifacts than the other methods that fail to produce good-quality photos. The DWT and CSR in Figs. 9f and g produce better information; however, there is still contrast and noise issue in

DWT. The CNN-HOD and LatLRR-VGG obtain better quality photos than existing methods. However, CNN-HOD retains more information from CT photos while LatLRR-VGG preserves more information from MRI photos, which makes these approaches unable to get sufficient detailed information over the proposed method.

It can be observed in Fig. 10c that the lesions, edges, and textures in the proposed method are more precise and vivid than all existing approaches due to the unique characteristics of novel D-CNN and NL-AD. Moreover, the contrast of the proposed method is more outstanding than that of other methods due to the different structuring elements of the MI-TH strategy that help detect all lesions in the fused photo. The noise is effectively eliminated due to the noise reduction approach of the proposed scheme. Due to the unique characteristics of each stage in the proposed scheme, the fused photo produces more significant information than any other existing methods. The fused photos obtained by DCT and DWT in Fig. 10g and h are affected due to the distortion, which affects the overall quality of the photo. The CNN-HOD approach in Fig. 10d produces better quality photos than CSR, DWT, and DCT, but this approach still cannot retain better contrast due to which it lacks in differentiating the bones and soft tissues in the body.

It can be seen in Fig. 11 that the proposed method acquires better contrast with very negligible noise than other existing methods due to the unique characteristics of NL-AMF and MI-TH. Besides, the proposed method has smooth edges,

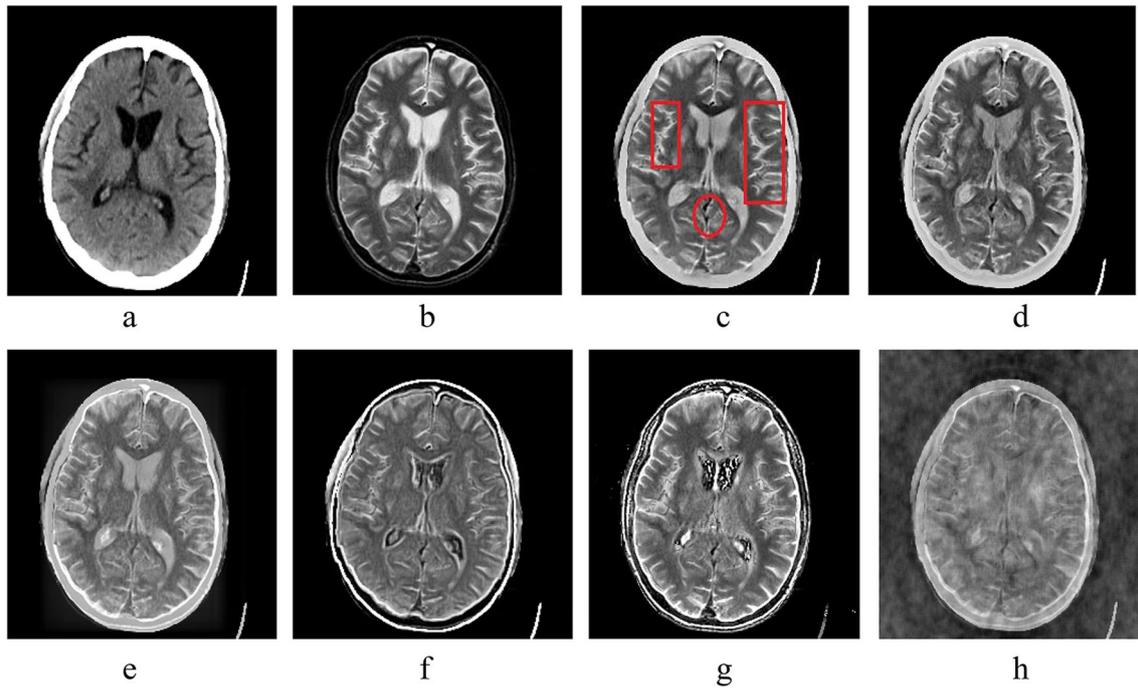


Fig. 7 Pair two of CT and MRI: (a) CT image (b) MRI image (c) Proposed (d) CNN-HOD (e) LatLRR-VGG (f) CSR (g) DWT (h) DCT

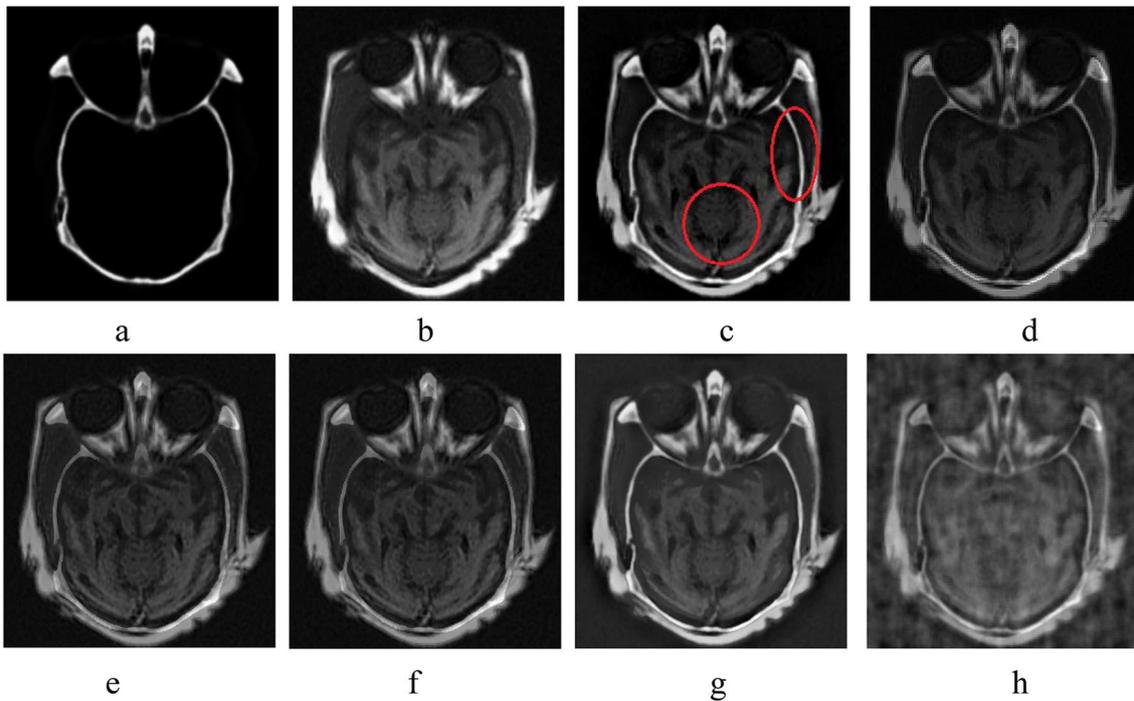


Fig. 8 Pair three of CT and MRI: (a) CT image (b) MRI image (c) Proposed (d) CNN-HOD (e) LatLRR-VGG (f) CSR (g) DWT (h) DCT

sharp textures, and significant energy information due to NL-AD, novel D-CNN, and dimension reduction approach. The CNN-HOD and LatLRR-VGG obtain better results than

the remaining existing methods, and both approaches produce similar results. The CNN-HOD produces better results than LatLRR-VGG in some cases, while LatLRR-VGG acquires

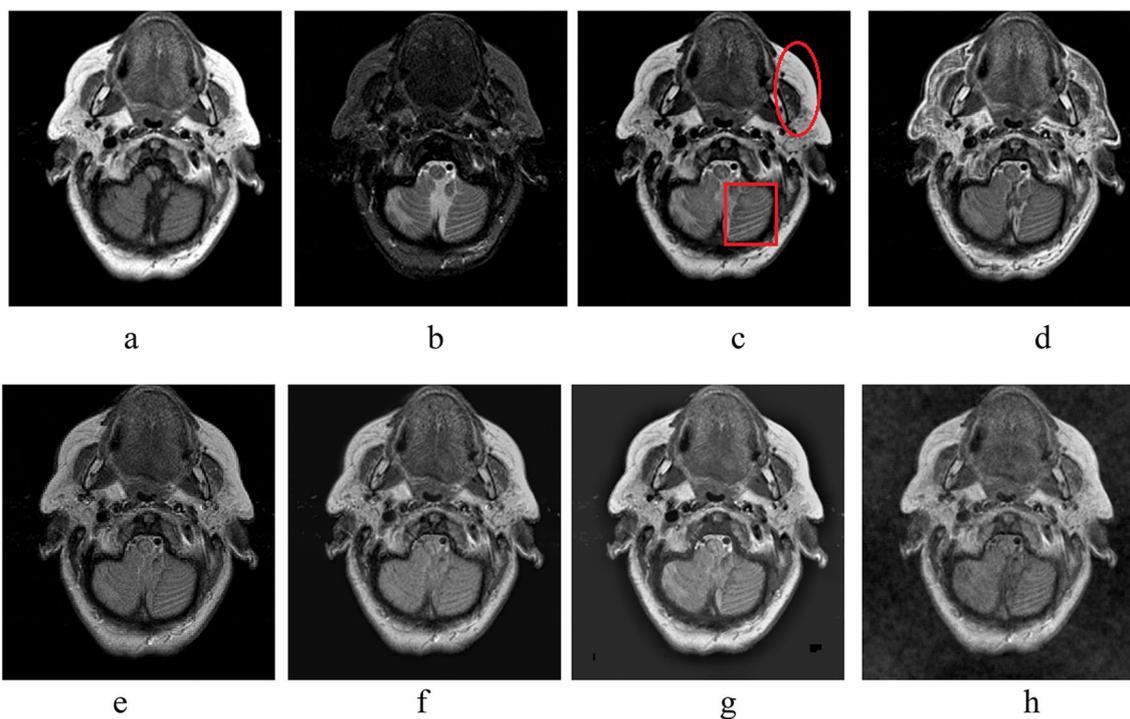


Fig. 9 Pair four of CT and MRI: (a) CT image (b) MRI image (c) Proposed (d) CNN-HOD (e) LatLRR-VGG (f) CSR (g) DWT (h) DCT

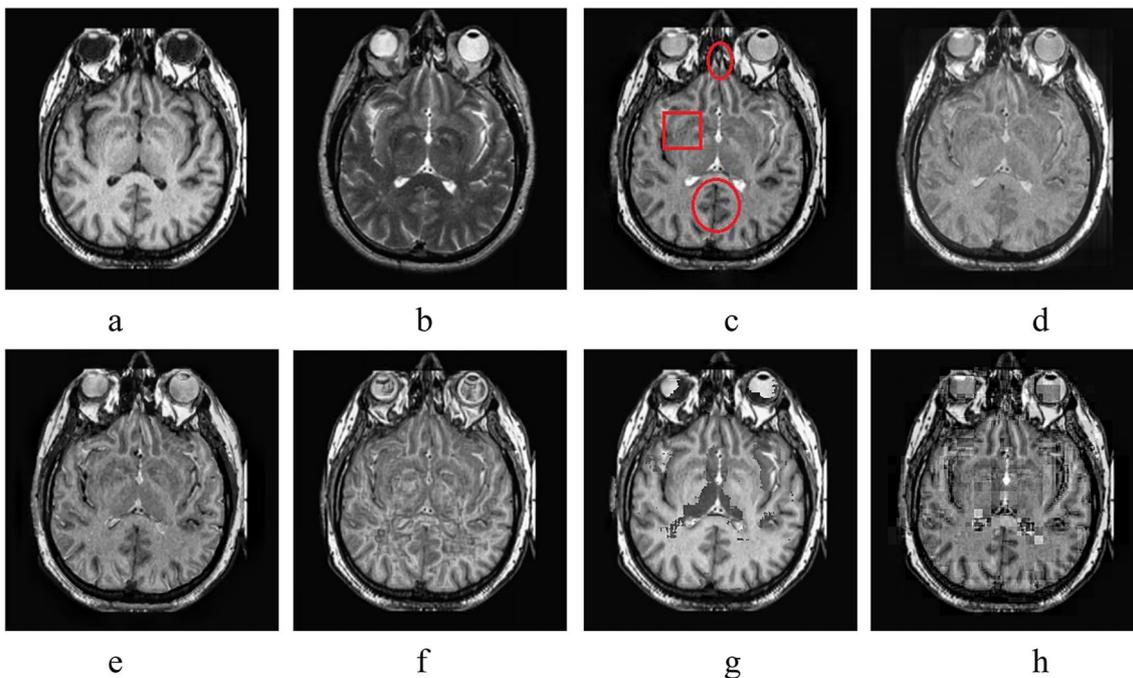


Fig. 10 Pair five of CT and MRI (a) CT image (b) MRI image (c) Proposed (d) CNN-HOD (e) LatLRR-VGG (f) CSR (g) DWT (h) DCT

better results than CNN-HOD in others, making it difficult to give a second rank from these approaches through HVS. The DCT falls in last in this race as it fails to capture good-quality photos because it is highly affected by noise and artifacts. The

fused photos generated by DWT are also distorted due to artifacts, and improper fusion strategy results in an image with less information. The CSR generates better results than DWT and DCT, but the quality of this approach is still not up to the mark.

4.1.1 Case study for qualitative evaluation by experts

The proposed method achieves superior results compared to existing methods through the Human Visual System (HVS). This superiority is further supported by a case study wherein ten field experts were engaged to rank each method. These specialists validated the qualitative evaluation by several factors: photo contrast, artifacts in the generated photo, information loss, energy information in the photo, and detailed features, including edges, textures, patterns, and boundaries in the fused photo. The experts were requested to rank each method on a scale from one to five, in which a higher value corresponds to a better-quality photo.

The ten pairs of source photos are used in this work, and each expert was asked to score for the proposed method, CNN-HOD, LatLRR-VGG, CSR, DWT, and DCT. The anonymous results were provided to each expert in random order for a fair trial to avoid bias in the qualitative evaluation. In addition, all the simulation results were provided to all experts on the same computer screen, light and were seated in the same room for fair observation. Each field expert ranked the simulation results of all methods for ten pairs of source photos. Table 3 presents the average score assigned by each expert for all methods for each pair of source photos.

Figure 12 presents the score points for each method given by field experts. The proposed method achieves a higher score than existing methods except for pair-6 and pair-7. Hence, it can be validated that the proposed method shows its supremacy in terms of not only qualitative evaluation but also field experts through user case studies.

4.1.2 Stability analysis

The stability of the proposed model can be obtained in terms of how well the model is generalized to new, unseen data and how it can effectively train the data. In order to accomplish this task, this work uses Wilcoxon signed rank to present the superior stability of the proposed method over existing methods. The Wilcoxon signed-rank test detects a significant difference between the source images and the fused photo. We have generated two types of image pairs for testing purposes: the CT image versus the fused photo and the MRI image versus the fused photo. This helps to assess the effect of different methods on image fusion. The smaller the p-value, the more chance it is to reject the null hypothesis, indicating that the fused photo differs from the source photos and thereby carries substantial information about the CT/MRI images.

It can be depicted from Table 4 that the p-value of CT-fused acquired by the proposed method is the lowest, indicating that the proposed method can fuse the input images more effectively.

4.2 Quantitative evaluation

The qualitative evaluation is carried out to prove the validation of any method that is acquired by HVS through numerous factors, but this evaluation is not enough for justification of superiority for any method. There is also a need to prove the supremacy of any method by using mathematical formulas, known as quantitative evaluation. Numerous authors have used different quantitative assessment parameters because there is no universal standard for them, but this work uses those assessment parameters that are commonly used by existing methods that we have compared. A short view for each parameter is presented, which is used in this paper for comparison.

4.2.1 Quality transfer ratio from source photo to a generated photo (Q_G^{AB})

The Q_G^{AB} is utilized to observe the transfer of detailed information such as edges, textures, and boundaries from both source photos to a final generated photo [40]. The higher value of Q_G^{AB} corresponds that more detailed information is transferred from source photos to a generated photo.

4.2.2 Feature mutual information (FMI)

The FMI is computed to observe the transformation of photo features, including contrast, textures, and edges from individual source photos to a final generated photo [4]. There will be more transformation of information from source photos to generated photos if the value of FMI is large.

4.2.3 Structural similarity index (SSIM)

The SSIM indicates the structural similarity between the source photos and the generated photo [4]. The larger value of SSIM relates that a large amount of similar information is extracted from the source photos to fused photos.

4.2.4 Average pixel intensity (API)

The API quality parameter is computed to check the contrast of the fused photo. The larger value of API indicates that the fused photo has better contrast, while the smaller value of API indicates that the generated photo has a dull contrast [4].

4.2.5 Noise and artifacts ratio from source photos to a generated photo (N_G^{AB})

The N_G^{AB} is computed to find the amount of noise and artifacts in the generated photo. The higher value of N_G^{AB} indicates that a large amount of noise or artifacts is introduced

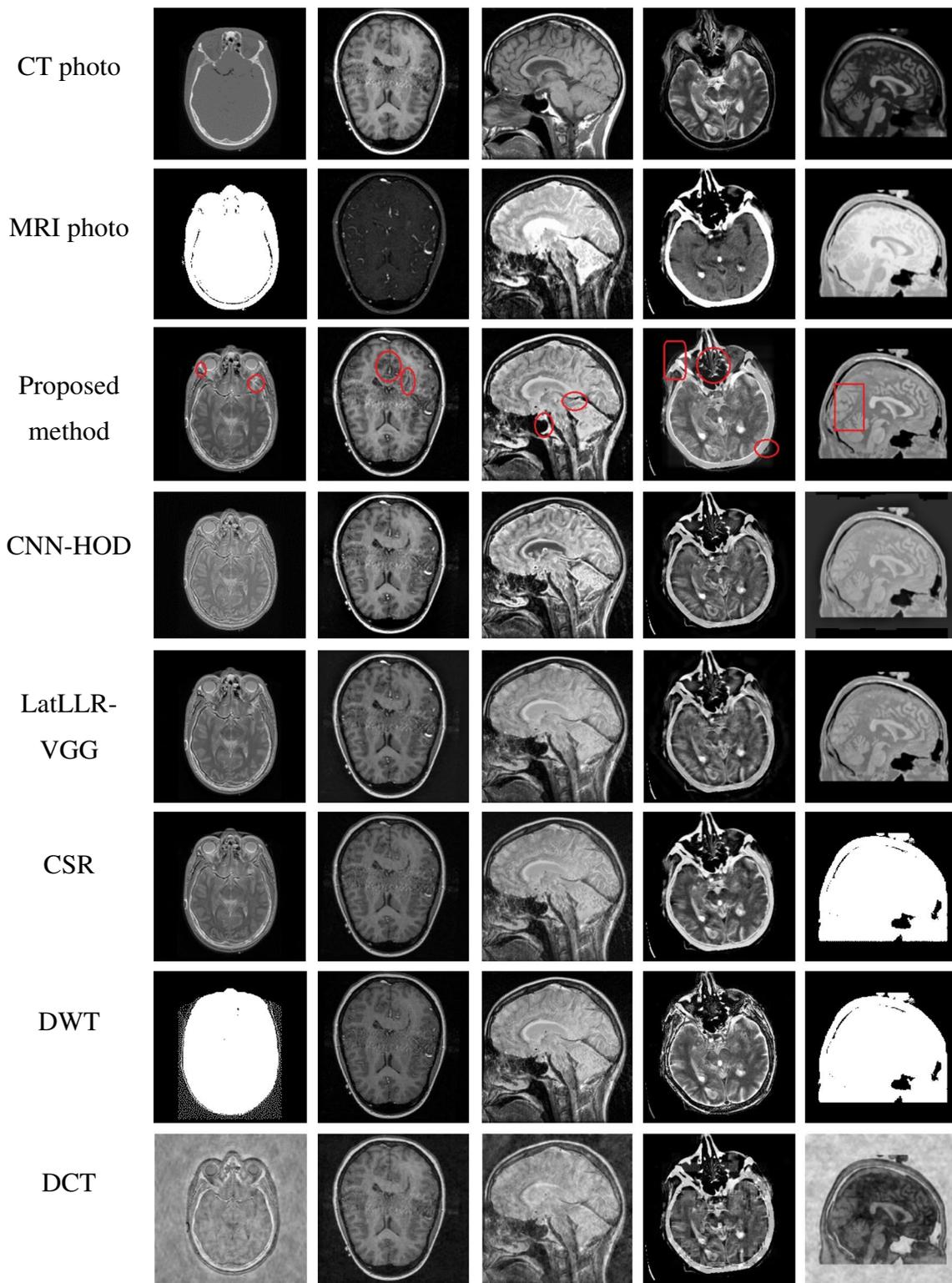
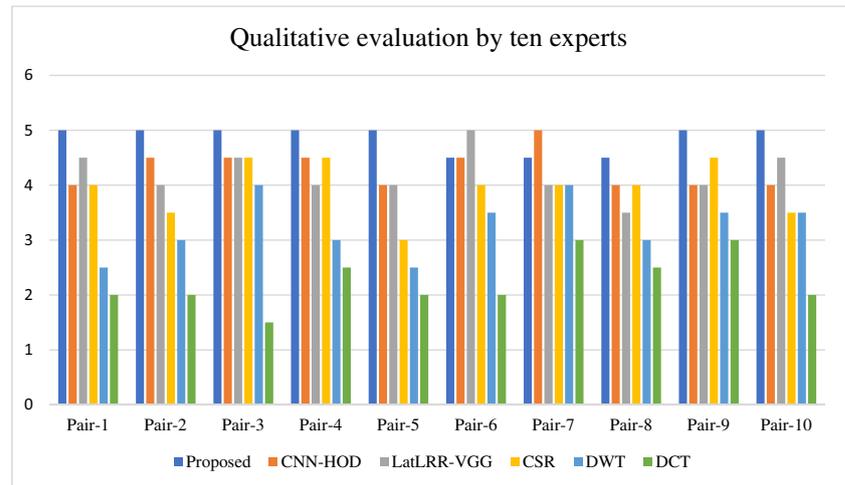


Fig. 11 CT photos, MRI photos, proposed method, and other fusion methods. From the top first row to the bottom 8th row shows CT photo, MRI photo, proposed method, CNN-HOD, LatLRR-VGG, CSR, DWT, and DCT fusion methods

Table 3 Average score gained by each method from all experts

Pair of CT and MRI	Proposed	CNN-HOD [6]	LatLRR-VGG[10]	CSR [24]	DWT [18]	DCT [19]
1	5	4	4.5	4	2.5	2
2	5	4.5	4	3.5	3	2
3	5	4.5	4.5	4.5	4	1.5
4	5	4.5	4	4.5	3	2.5
5	5	4	4	3	2.5	2
6	4.5	4.5	5	4	3.5	2
7	4.5	5	4	4	4	3
8	4.5	4	3.5	4	3	2.5
9	5	4	4	4.5	3.5	3
10	5	4	4.5	3.5	3.5	2

Fig. 12 The average score obtained by experts for each method**Table 4** Stability analysis of the proposed method with existing methods

Methods	Wilcoxon Signed-Rank Test	
	CT-Fused	MRI-Fused
DCT	3.75e-18	3.71e-18
DWT	3.72e-19	5.68e-19
CSR	5.67e-19	3.61e-19
LatLRR	3.71e-19	5.32e-19
CNN-HOD	3.62e-19	3.63e-19
Proposed	3.41e-19	3.39e-19

in the fused photo, while a small value results in a photo with less noise [40].

The simulation results for ten pairs of CT and MRI source photos are carried out in this paper. The average values for Q_G^{AB} , FMI, SSIM, API and N_G^{AB} are presented in Table 5. As mentioned above, higher values of Q_G^{AB} , FMI, SSIM, and API relate to better quality for a generated photo, and higher values are written in bold letters. In contrast, the smaller value of N_G^{AB}

corresponds to a photo with less noise, and smaller values are written in bold letters.

In addition, this work has also plotted the graphs for all parameters, which are presented in Figs. 13, 14, 15, 16 and 17. It can be seen from Table 5 that the average values of quantitative parameters are higher for the proposed method over existing methods. It can be observed in Fig. 13 that the proposed method achieves higher values of Q_G^{AB} in all pairs of source photos except for the sixth pair. The proposed method also attains higher values for the FMI quality parameter for all pairs of photos except for the fifth and sixth pair of photos, which can be seen in Fig. 14. The proposed method also obtains larger values for all pairs of source photos for the FMI parameter except for the fifth and sixth pair of source photos, where CNN-HOD and LatLRR-VGG have slightly higher values, as shown in Fig. 15.

Similarly, it can be observed in Fig. 16 that the proposed method has larger values for all fused photos of API parameter except for the seventh pair of source photos where CNN-HOD has a little higher values. Hence, it reveals that the fused

Table 5 The average values of quantitative parameters for all methods

Quality parameter	Proposed	CNN-HOD	LatLRR-VGG	CSR	DWT	DCT
Q_G^{AB}	0.8986	0.8527	0.8503	0.82945	0.8038	0.7649
FMI	0.4783	0.4497	0.4382	0.4193	0.3893	0.3520
SSIM	0.9056	0.8625	0.8502	0.8265	0.7928	0.7639
API	81.0487	78.0291	78.9432	74.8907	72.6984	68.8932
N_G^{AB}	0.1275	0.1730	0.1869	0.2193	0.2498	0.2893

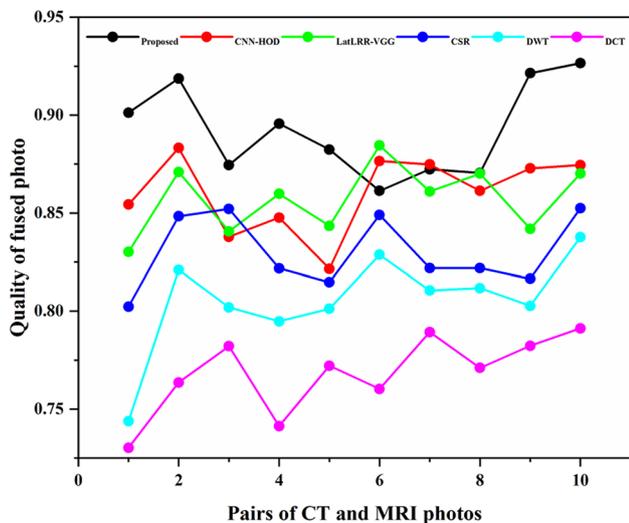


Fig. 13 The Q_G^{AB} plotting for ten pairs of source photos

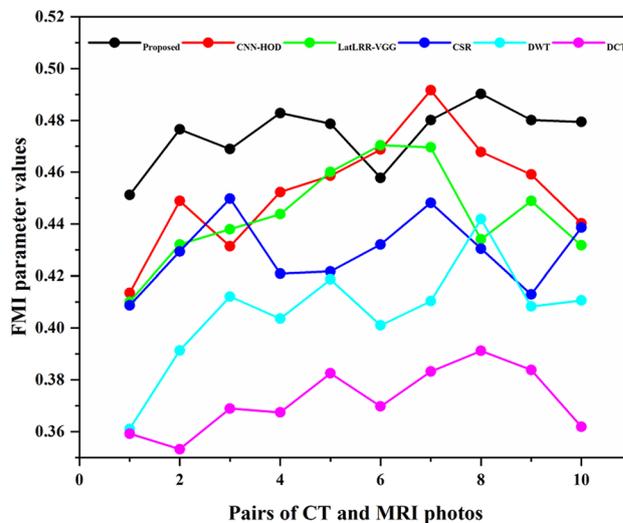


Fig. 14 The FMI plotting for ten pairs of source photos

photo of the proposed method attains better contrast, sufficient energy information, and enriched detailed features such as smooth edges, textures, and boundaries than other methods. The proposed method also handles the collision more effectively than existing methods. This paper uses the quality transfer ratio from source photos to output photo (Q_G^{AB}), feature mutual information (FMI), average pixel intensity (API) and structural similarity index (SSIM) for computing the extraction of features. The proposed method gains higher values in these four parameters than existing methods, which shows that the proposed method captures more salient features by avoiding collision. This is due to effective pre-fusion refinement approaches such as non-linear average median filtering (NL-AMF) and multiscale improved top-hat (MI-TH) approach that are particularly implemented to circumvent issues such as collision, noise, and artifacts, enabling the improvement of overall visual quality and clarity of fused photos. These preprocessing steps contribute to the mitigation of collision-related distortions, ensuring a more accurate and reliable representation of the original scene.

Likewise, it can be analyzed in Fig. 17 that there is also very negligible noise in the proposed method compared to all other methods except for the sixth pair of source photos, where LatLRR-VGG has a little less noise.

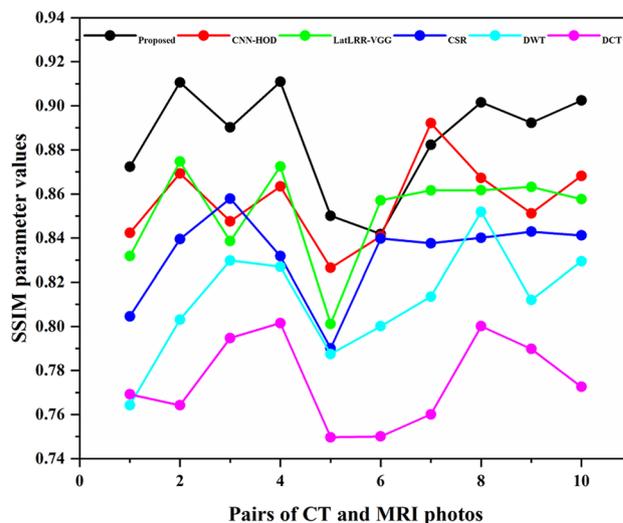


Fig. 15 The SSIM plotting for ten pairs of source photos

Therefore, it is concluded that the proposed method not only preserves better contrast, complete energy information, and substantial detailed features, but it also has

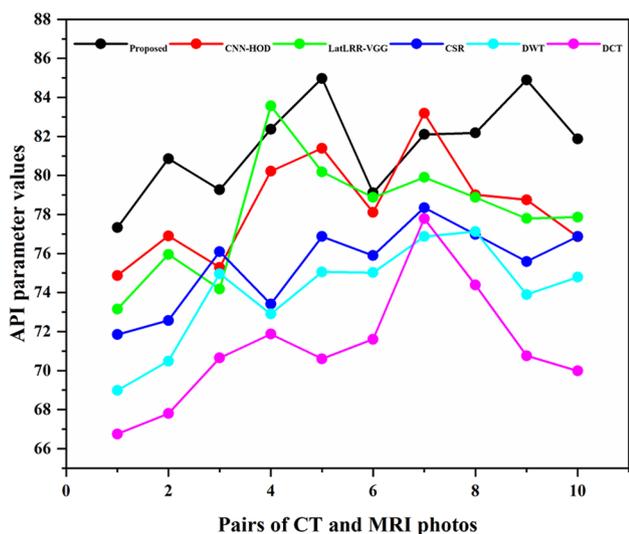


Fig. 16 The API plotting for ten pairs of source photos

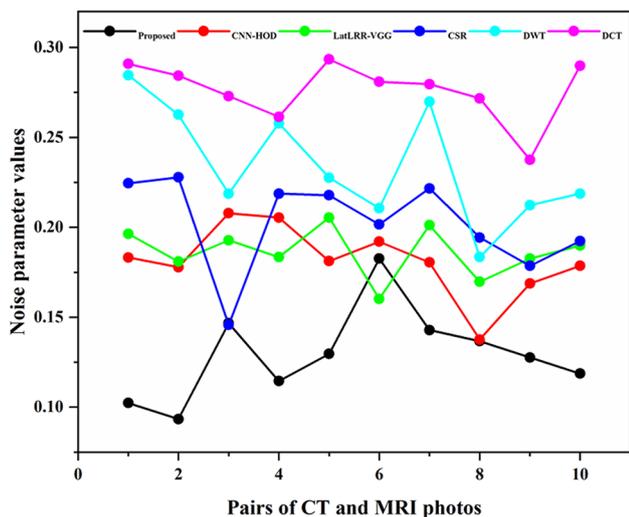


Fig. 17 The N_G^{AB} plotting for ten pairs of source photos

very negligible noise in comparison to all other existing methods.

4.3 Computation time

Considering the computation in the field of multimodal medical image fusion is important, specifically in the context of practical implementation of real-world applications. The

average computation time for the proposed method and existing MMIF schemes is presented in Table 6. All the simulation results are obtained for computation time (t) in seconds using MATLAB 2021b on the 11th generation of core i5 2.40 GHz processor with RAM of 16 GB.

It can be depicted in Table 6 that the average computation time for LatLRR-VGG and CSR is higher than other methods, leading to the poorest timeliness. The computation time for DCT and DWT is the fastest; however, their performance on subjective and objective evaluation is worse. The computation time for the proposed method is longer than DCT and DWT, but better visual effects can sacrifice a small amount of time. The computation time for the proposed method is shorter than CNN-HOD, CSR, and LatLRR-VGG, and it also generates better visual effects on subjective and objective evaluation. There is a tradeoff between the computation time and the better visual effects. The proposed method consumes less time than recent state-of-the-art methods while also achieving better visual effects on subjective and objective evaluation. Nonetheless, there is still a need to improve the consumption time, which is needed for real-time applications.

5 Discussion

Designing an MMIF algorithm that extracts all substantial features with uniform contrast by removing noise from multiple source photos into a single fused photo is one of the hot research area in the field of medical applications. This paper is one of the attempts in the field of a medical research area that uses a novel fusion approach. Based on the implemented MMIF approach in this work, the generated fused photo does not only remove the noise and adjust the contrast but also retains sufficient energy information with enriched detailed features such as textures, edges, and boundaries, which can be applied in numerous medical applications for diagnosing the disease.

Various hybrid MMIF approaches have been designed to amalgamate the specific features from different methods to enhance the quality of the generated photo. The proposed novel hybrid method consists of several sequential stages, and each stage is designed for the specific task of improving the quality of the fused photo. Noise is the pivotal issue that degrades the overall performance of any fusion method, and the authors haven't addressed this issue. The proposed work designs an NL-AMF preprocessing noise reduction approach that adaptively resizes the mask filter in accordance with the noise level. This

Table 6 The average running time for the proposed method and other methods (Unit of time is seconds)

Method	DCT	DWT	CSR	LatLRR-VGG	CNN-HOD	Proposed
Time	10.6	9.4	93	96	15.4	13.1

approach can eliminate the noise and reduce the computation time due to the unique feature of a statistical histogram that makes the search process faster for finding median values. Another unique stage in this paper is to design an MI-TH preprocessing method for adjusting the contrast of source photos, which is neglected by researchers in the field of MMIF. Though very few scholars have considered adjusting contrast, their approach still fails to produce the desired photo contrast. The MI-TH approach at the preprocessing stage uses different-size multiscale structuring elements for accurately retaining the ROI, resulting in a photo with adjustable contrast and enriched features.

It is well-known that any source photo contains energy information (base part) and detailed features such as edges, textures, and boundaries (detailed part). Therefore, it is one of the fundamental steps of any MMIF to split the source photos into the base and detailed parts before applying the fusion strategy because both (the base and detailed parts) contain different information, and one fusion strategy cannot retain substantial information. This paper employs an NL-AD using a forward time space central and partial differential equation approach, which cannot only split the source photos into a base part and detailed parts but also smoothen the homogenous regions with its intra-region smoothing process.

Numerous authors have designed appropriate fusion strategies for a fusion of base and detailed parts. Nonetheless, their fusion methods either lack in preserving sufficient detailed features or artifacts are introduced due to improper fusion strategies. To the best of our knowledge, this paper is the first attempt to design a novel D-CNN method that employs skipping convolution and deconvolution layers using direct mapping and the focus map to shield the enriched detailed features without losing colors and other crucial details. Moreover, the dimension reduction approach is utilized for a fusion of base layers, which attain not only vital energy information but also discards unnecessary information and addresses overfitting issues. At last, the ultimate generated fused photo is retrieved through the superposition of the base and detailed parts. This fused photo preserves all complementary energy information and detailed enriched textures with desired contrast by avoiding uneven illumination and mitigating noise and artifacts. These innovative stages entitle this proposed study to acquire outstanding results compared to existing methods on qualitative and quantitative evaluation for examined normal scenes and complex scenes.

6 Limitations and future recommendations

Although the proposed method justifies its superiority over existing methods by producing high-quality fused photos, it still has some limitations that need to be addressed

for future work. The consumption time of the proposed method is less, but it is still not promising for numerous medical applications. In future work, one of the main tasks would be shortening the consumption time while maintaining the quality of fused photos, which can be implemented in numerous industrial applications that require real-time performance. Another major challenge can be implementing the model from research theory to real-world deployments, which can encounter practical challenges related to integration with existing systems and completeness. To cope with that issue, collaboration with medical hospitals can facilitate the identification of deployment issues. Our next focus would be to implement medical image fusion for specific applications and deploy it in medical hospitals. Another limitation of this work pertains to the unified pixel size utilized in the image fusion process. The reliance on a fixed pixel size poses a constraint. This limitation implies that the model's performance may be compromised when applied to images captured with different cameras or specifications. One of our future potential research directions is to investigate adaptive fusion methods that can dynamically adjust to different pixel sizes encountered in diverse imaging setups. This could involve the development of algorithms capable of automatically detecting and adapting to variations in pixel sizes, thereby ensuring robust performance across a range of imaging devices and specifications. In addition, exploring techniques such as image resampling or interpolation to standardize pixel sizes before fusion could be beneficial for the mentioned issue.

Another limitation of this work is that it cannot effectively treat the occlusion, partial visibility that occurs in the real-time application. Although this work is based on static images, the fused images are not affected by occlusion and partial visibility. However, real-time images are highly affected by this issue. Future research on real-time images to treat occlusion and partial visibility could be based on the implementation of CNN architecture, which uses semantic segmentation and multi-resolution fusion approaches. These approaches can be potential solutions to treat the occlusion, partial visibility, and overlapping regions.

7 Conclusion

The novel hybrid MMIF algorithm is designed in this paper to integrate the merits of each stage for improving the quality of generated photos, which can be used in numerous medical applications for diagnosing. The NL-AMF approach is employed at a preprocessing stage, which mitigates the noise and minimizes the computation time due to the unique attribute of the statistical histogram, which swifts the searching process for finding median values.

The photo is then processed by the MI-TH method, which has distinctive characteristics of using different-sized structuring elements to preserve the desired ROI, resulting in an image with better contrast and detailed features. The noise-free and improved contrast photos are split into base and detailed parts by NL-AD using the forward time space central method that smoothens the homogenous regions with its intra-region smoothing feature. The detailed parts are fused by a novel D-CNN, which is the first attempt to design skipping convolution and deconvolution layers to shield the detailed features without loss of colors and other crucial details using direct mapping and focus map. Conversely, the base parts are fused by a dimension reduction fusion strategy, which is proven to be the best method for preserving the energy information, and it also addresses the overfitting issues while discarding unnecessary data. Finally, a fused photo is generated by the superposition of base and detailed parts, which contains significant energy information, substantial detailed features, and visually better contrast with negligible noise. The comparison of the proposed method and SOTA methods is conducted by qualitative and quantitative evaluation of ten pairs of source photos. The qualitative evaluation is carried out by ten field experts based on their HVS experience, and the proposed method achieved a higher score than all other existing methods. Similarly, the qualitative evaluation (Q_G^{AB} , FMI, SSIM, API, and N_G^{AB}) is acquired by mathematical computation, and the proposed method surpasses existing methods, which further justifies the supremacy of the proposed method.

Declarations

Conflict of interest The authors declare no conflict of interest in this research work.

References

- Karim S, Tong G, Li J, Qadir A, Farooq U, Yiting Y (2023) Current advances and future perspectives of image fusion: a comprehensive review. *Information Fusion* 90:185–217. <https://doi.org/10.1016/j.inffus.2022.09.019>
- Aamir M, Rahman Z, Dayo ZA, Abro WA, Irfan Uddin M, Khan I, Imran AS, Ali Z, Ishfaq M, Guan Y, Zhihua H (2021) A deep learning approach for brain tumor classification using MRI images. *Comput Electr Eng* 101:108105. <https://doi.org/10.1016/j.compeleceng.2022.108105>
- Hermessi H, Mourali O, Zagrouba EJSP (2021) Multimodal medical image fusion review: theoretical background and recent advances. *Signal Process* 183:108036. <https://doi.org/10.1016/j.sigpro.2021.108036>
- Bhutto JA, Tian L, Qiliang D, Sun Z, Lubin Y, Soomro TA (2022) An improved infrared and visible image fusion using an adaptive contrast enhancement method and deep learning network with transfer learning. *Remote Sens* 14(4):939. <https://doi.org/10.3390/rs14040939>
- Azam MA, Khan KB, Salahuddin S, Rehman E, Khan SA, Khan MA, Kadry S, Gandomi AH (2022) A review on multimodal medical image fusion: compendious analysis of medical modalities, multimodal databases, fusion techniques and quality metrics. *Comput Biol Med* 144:105253. <https://doi.org/10.1016/j.combiomed.2022.105253>
- Almasri MM, Alajlan AM (2022) Artificial intelligence-based multimodal medical image fusion using hybrid S^2 optimal CNN. *Electronics* 11(14):2124. <https://doi.org/10.3390/electronic11142124>
- Guan Y, Aamir M, Rahman Z, Ali A, Abro WA, Dayo ZA, Bhutta MS, Zhihua H (2021) A framework for efficient brain tumor classification using MRI images[J]. *Math Biosci Eng* 18(5):5790–5815. <https://www.aimspress.com/article/doi/10.3934/mbe.2021292>
- Li Y, Zhao J, Lv Z, Li J (2021) Medical image fusion method by deep learning. *Int J Cogn Comput Eng* 2:21–29. <https://doi.org/10.1016/j.ijcce.2020.12.004>
- Kong W, Li C, Lei Y (2022) Multimodal medical image fusion using convolutional neural network and extreme learning machine. *Front Neurobot* 16:1050981. <https://doi.org/10.3389/fnbot.2022.1050981>
- Lou X-C, Feng XJC, Juhola M (2021) Medicine, multimodal medical image fusion based on multiple latent low-rank representation. *Comput Math Methods Med*:1–16. <https://doi.org/10.1155/2021/1544955>
- Wang X, Hua Z, Li J (2023) Multi-focus image fusion framework based on transformer and feedback mechanism. *Ain Shams Eng J* 14(5):101978. <https://doi.org/10.1016/j.asej.2022.101978>
- Zhou T, Li Q, Huiling L, Cheng Q, Zhang X (2023) GAN review: models and medical image fusion applications. *Inform Fusion* 91:134–148. <https://doi.org/10.1016/j.inffus.2022.10.017>
- Soomro TA, Khan TM, Khan MAU, Gao J, Paul M, Zheng L (2018) Impact of ICA-based image enhancement technique on retinal blood vessels segmentation. *IEEE Access* 6:3524–3538. <https://doi.org/10.1109/ACCESS.2018.2794463>
- Aamir M, Yi-Fei P, Rahman Z, Tahir M, Naeem H, Dai Q (2018) A framework for automatic building detection from low-contrast satellite images. *Symmetry* 11(1):3. <https://doi.org/10.3390/sym11010003>
- Bhutto JA, Lianfang T, Qiliang D, Soomro TA, Lubin Y, Tahir MF (2020) An enhanced image fusion algorithm by combined histogram equalization and fast gray level grouping using multi-scale decomposition and gray-PCA. *IEEE Access* 8:157005–157021. <https://doi.org/10.1109/ACCESS.2020.3018264>
- Ma W, Wang K, Li J, Yang SX, Li J, Song L, Li Q (2023) Infrared and visible image fusion technology and application: a review. *Sensors* 23(2):599. <https://doi.org/10.3390/s23020599>
- Choudhary G, Sethi D (2022) From conventional approach to machine learning and deep learning approach: an experimental and comprehensive review of image fusion techniques. *Arch Comput Methods Eng* 30:1267–1304 <https://link.springer.com/article/10.1007/s11831-022-09833-5>
- Tawade L, Aboobacker AB, Ghante F (2014) Image fusion based on wavelet transforms. *Int J Bio-Sci Bio-Technol* 6(3):149–162. <https://doi.org/10.14257/ijbsbt.2014.6.3.18>
- Shreyamsha Kumar BK (2013) (2012), multifocus and multi-spectral image fusion based on pixel significance using discrete cosine harmonic wavelet transform, in signal. *Image Video Process* 7:1125–1143 <https://link.springer.com/article/10.1007/s11760-012-0361-x>
- Gao G, Xu L, Dongzhu F (2013) Multi-focus image fusion based on non-subsampled shearlet transform. *IET Image Process* 7(6):543–639. <https://doi.org/10.1049/iet-ipr.2012.0558>

21. Tawfik N, Elnemr HA, Fakhr M, Dessouky MI, Abd El-Samie FE (2022) Multimodal medical image fusion using stacked auto-encoder in NSCT domain. *J Digit Imaging* 35:1308–1325 <https://link.springer.com/article/10.1007/s10278-021-00554-y>
22. Nagaraja Kumar N, Jayachandra Prasad T, Satya Prasad K (2022) An intelligent multimodal medical image fusion model based on improved fast discrete Curvelet transform and Type-2 fuzzy entropy. *Int J Fuzzy Syst* 25:96–117 <https://link.springer.com/article/10.1007/s40815-022-01379-9>
23. Shreyamsha Kumar BK (2013) Image fusion based on pixel significance using cross bilateral filter. *SIViP* 9:1193–1204 <https://link.springer.com/article/10.1007/s11760-013-0556-9>
24. Liu Y, Chen X, Ward RK, Jane Wang Z (2016) Image fusion with convolutional sparse representation. *IEEE Signal Process Lett* 23(12):1882–1886. <https://doi.org/10.1109/LSP.2016.2618776>
25. Li X, Zhang X, Ding M (2019) A sum-modified-Laplacian and sparse representation based multimodal medical image fusion in Laplacian pyramid domain. *Med Biol Eng Comput* 57:2265–2275 <https://link.springer.com/article/10.1007/s11517-019-02023-9>
26. Rajalingam B, Fadi Al-Turjman R, Santhoshkumar MR (2020) Intelligent multimodal medical image fusion with deep guided filtering. *Multimed Syst* 28:1449–1463. <https://link.springer.com/article/10.1007/s00530-020-00706-0>
27. Wang L, Dou J, Qin P, Lin S, Gao Y, Wang R, Zhang J (2021) Multimodal medical image fusion based on nonsubsampling shearlet transform and convolutional sparse representation. *Multimed Tools Appl* 80:36401–36421. <https://link.springer.com/article/10.1007/s11042-021-11379-w>
28. Guo P, Xie G, Li R, Hui H (2022) Multimodal medical image fusion with convolution sparse representation and mutual information correlation in NSST domain. *Complex Intell Syst* 9:317–328 <https://link.springer.com/article/10.1007/s40747-022-00792-9>
29. Li H, Wu X-J (2018) Infrared and visible image fusion using latent low-rank representation. *Comput Vis Pattern Recognit* 5:6. <https://doi.org/10.48550/arXiv.1804.08992>
30. Tawfik N, Elnemr HA, Fakhr M, Dessouky MI, Abd El-Samie FE (2021) Hybrid pixel-feature fusion system for multimodal medical images. *J Ambient Intell Humaniz Comput* 12:6001–6018. <https://link.springer.com/article/10.1007/s12652-020-02154-0>
31. Venkatesan B, Ragupathy US (2022) Integrated fusion framework using hybrid domain and deep neural network for multimodal medical images. *Multimed Syst Sign Process* 33:819–834 <https://link.springer.com/article/10.1007/s11045-021-00813-9>
32. Soomro TA, Afifi AJ, Gao J, Hellwich O, Zheng L, Paul M (2019) Strided fully convolutional neural network for boosting the sensitivity of retinal blood vessels segmentation. *Expert Syst Appl* 134(15):36–52. <https://doi.org/10.1016/j.eswa.2019.05.029>
33. Aamir M, Rahman Z, Abro WA, Tahir M, Ahmed SM (2019) An optimized architecture of image classification using convolutional NeuralNetwork. *Int J Image Graph Signal Process* 10:30–39 <https://www.mecs-press.org/ijigsp/ijigsp-v11-n10/IJIGSP-V11-N10-5.pdf>
34. Soomro TA, Afifi AJ, Zheng L, Soomro S, Gao J, Hellwich O, Paul M (2019) Deep learning models for retinal blood vessels segmentation: a review. *IEEE Access* 7:71696–71717. <https://doi.org/10.1109/ACCESS.2019.2920616>
35. Zhang L, Li H, Zhu R, Ping D (2022) An infrared and visible image fusion algorithm based on ResNet-152. *Multimed Tools Appl* 81:9277–9287. <https://link.springer.com/article/10.1007/s11042-021-11549-w>
36. Feng Y, Houqing L, Bai J, Cao L, Yin H (2020) Fully convolutional network-based infrared and visible image fusion. *Multimed Tools Appl* 79:15001–15014 <https://link.springer.com/article/10.1007/s11042-019-08579-w>
37. Xu L, Jimmy SJ, Ren CL, Jia J (2014) Deep convolutional neural network for image deconvolution. In: *Advances in neural information processing systems* 27 (NIPS 2014). <https://proceedings.neurips.cc/paper/2014/hash/1c1d4df596d01da60385f0bb17a4a9e0-Abstract.html>
38. Krokos V, Xuan VB, Bordas SPA, Young P, Kerfriden P (2021) A Bayesian multiscale CNN framework to predict local stress fields in structures with microscale features. *Comput Mech* 69:733–766 <https://link.springer.com/article/10.1007/s00466-021-02112-3>
39. Perona P, Malik J (1990) Scale-space and edge detection using anisotropic diffusion. *IEEE Trans Pattern Anal Mach Intell* 12(7):629–639. <https://doi.org/10.1109/34.56205>
40. Bhutto JA, Tian L, Qiliang D, Sun Z, Lubin Y, Tahir MF (2022) CT and MRI medical image fusion using noise-removal and contrast enhancement scheme with convolutional neural network. *Entropy* 24(3):393. <https://doi.org/10.3390/e24030393>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.



Jameel Ahmed Bhutto received his Bachelor of Engineering in Telecommunication Engineering from Mehran University of Engineering and Technology, in 2010, and Master of Engineering degree in Communication Systems and Networking, from Mehran University of Engineering and Technology, in 2016. He has completed his PhD degree from South China University of Technology, China. He is working as an associate professor at the Department of Computer Science, Huanggang Normal University, China. His major research interests include pattern recognition, deep learning, computer vision, image fusion, image enhancement and digital image processing.



Jiang Guosong Male, Born in June 1976 in Ezhou, Hubei Province, Professor, Ph. D. in computer system structure of Huazhong University of Science and Technology, postdoctoral of Communication engineering of Communication University of China, vice dean of computer science of Huanggang Normal University. Master thesis won excellent master thesis in Hubei province; He has published more than 10 Chinese core papers, 9 EI retrieval papers, and chaired or participated in many projects

above the university level.



Ziaur Rahman has received MS degree in software engineering in 2017 from Chongqing University, Chongqing, China. He has received his PhD degree from Sichuan University, Chengdu, China. Currently, he is working as an Associate Professor in Huanggang Normal University, China. His research includes are image processing, deep learning and fractional calculus.



Muhammad Ishfaq male born in 1990. He obtained a doctorate degree from Northeast Agricultural University. The main research directions are computer pharmacology, information medicine theory and artificial intelligence. Currently, he is working as an Associate Professor in Huanggang Normal University, China.



Zhengzheng Sun was born in Jiaozuo, Henan, China, in 1996. He received the B.S. degree in automation from Fuzhou University, Fujian, China, in 2018. He received the Ph.D. degree in pattern recognition and intelligence system with the South China University of Technology, Guangzhou, China, in 2023. He is currently working at GRGBanking Equipment Co., Ltd., as a senior algorithm engineer. His research interests include face detection, face recognition, artificial intelligence, computer vision, LLM and AIGC.



Toufique Ahmed Soomro obtained a Ph.D. degree in AI and image processing from the School of Computing, Mathematics and Engineering, Charles Sturt University, Australia, in 2018. Currently, he is an associate professor and head of the Department of Electronic Engineering at QUEST Campus Larkano, Pakistan. His research interests include most aspects of image enhancement methods, segmentation methods, classification methods, and image analysis for medical images. He has 65 research publications in the field of AI for medical imaging with a commutative impact factor of 137.98 and an h-index of 19. He has also worked as a collaborator on various AI projects for biomedical imaging with the Ministry of Education of the Kingdom of Saudi Arabia and Najran University of the Kingdom of Saudi Arabia. He was honored to be invited as a young research professor at the University of Technology, Guangdong, China, in 2019-2020.