BenchCouncil Transactions on Benchmarks, Standards and Evaluations 5 (2025) 100237

Contents lists available at ScienceDirect

BenchCouncil Transactions on Benchmarks, Standards and Evaluations

journal homepage: www.keaipublishing.com/en/journals/benchcouncil-transactions-onbenchmarks-standards-and-evaluations/

Full Length Article

Medical image fusion based on deep neural network via morphologically processed residuals

Supinder Kaur [a], Parminder Singh [b], Rajinder Vir [b], Arun Singh [b,*], Harpreet Kaur [b]

[a] RBIENT, India
[b] School of Computer Science & Engineering, Lovely Professional University, Phagwara, Punjab, India

ARTICLE INFO

Keywords:
Image fusion
Medical imaging
Image pyramid
Neural network
Residuals

ABSTRACT

Medical image fusion enhances the intrinsic statistical properties of original images by integrating complementary information from multiple imaging modalities, producing a fused representation that supports more accurate diagnosis and effective treatment planning than individual images alone. The principal challenge lies in combining the most informative features without discarding critical clinical details. Although various methods have been explored, it remains difficult to consistently preserve structural and functional features across modalities. To address this, we propose a deep neural network–based framework that incorporates morphologically processed residuals for competent fusion. The network is trained to directly map source images into weight maps thereby overcoming the limitations of traditional activity-level measurements and weight assignment algorithms, and enabling adaptive and reliable weighting of different modalities. The framework further employs image pyramids in a multi-scale design to align with human visual perception, and introduces a local similarity–based adaptive rule for decomposed coefficients to maintain consistency and fine detail preservation. An edge-preserving strategy combining linear low-pass filtering with nonlinear morphological operations is used to emphasize regions of high amplitude and preserve optimally sized structural boundaries. Residuals derived from the linear filter guide the morphological process ensuring significant regions are retained while reducing artifacts. Experimental results demonstrate that the proposed method effectively integrates complementary information from multimodal medical images while mitigating noise, blocking effects, and distortions, leading to fused images with improved clarity and clinical value. This work provides an advanced and reliable fusion approach that contributes substantially to the field of medical image analysis, offering clinicians enhanced visualization tools for decision-making in diagnosis and treatment planning.

1. Introduction

Image fusion technology produces a well-informed single fused image which is highly informative due to combination of multiple source images. In addition to combining image data, this procedure also entails applying one or more algorithms to specifically process the resultant image [1–3]. In medical imaging, fusion is the procedure to combine two or multiple images from several imaging technologies to retain the benefit of complimentary information to generate a more complete image. Due to the emergence of medical imaging technology which includes MRI (magnetic resonance imaging), CT (computed tomography), PET (positron emission tomography), SPECT (single photon emission computed tomography) and several other imaging modalities are widely used in clinical applications and treatment planning. There are many kinds of medical information available from each imaging method. As a useful reference for lesion localization, CT images for instance provide great spatial resolution and good bone imaging. However, soft tissue and fine details of invasive tumours are harder for CT to show. On the other hand, MRI is excellent at soft-tissue imaging which makes it perfect for figuring out how immense a lesion. While their lesser spatial resolution may restrict the ability to diagnose tumours, PET and SPECT can offer useful information about the body's metabolic activities. Depending on a single image type frequently does not yield the best visualization since each imaging modality has inherent limitations resulting from different imaging principles. Medical image fusion can therefore produce more accurate and comprehensive images

Peer review under the responsibility of The International Open Benchmark Council.
* Corresponding author.
E-mail address: arunmandiarun2001@gmail.com (A. Singh).

https://doi.org/10.1016/j.tbench.2025.100237
Received 26 April 2025; Received in revised form 8 September 2025; Accepted 24 September 2025
Available online 26 September 2025
2772-4859/© 2025 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

by integrating the complementary information and strengths of several imaging modalities, greatly assisting in diagnosis and treatment planning [4,5].

Recent decades have seen a rise in interest in medical image fusion research, especially in the domain of multiscale transform-based methods. To summarize, the foremost sequential procedures of multiscale transform-based fusion techniques comprise source image decomposition into several coefficients, each of which uses sophisticated operators to encapsulate distinct feature information. In addition, different pixel-level fusion criteria are utilized to combine the matching components and inverse transform is carried out for generating a final fusion result. The initial conventional methods were those that depends on wavelet transforms [6] and Laplacian pyramid transforms [7]. Existing methods acquired diverse image feature information via multiscale transformation, as various frequency components were preserved using identical or simplistic fusion algorithms during the amalgamation of sub-bands. However, a constraint of these approaches is that they analyse only a single image attribute, ignoring others which might significantly reduce the fusion effect [8]. To address this issue, several advanced methods were presented. Tian et al [9] used pixel or regional data to build fusion rules across frequency bands. The fractional wavelet transform approach by Xu et al [10] improves fusion coefficient description. Likewise, Lie et al. presented a sparse representation theory based on minimum spanning trees (MST) [11]. The above-mentioned systems use diverse rules across several frequency bands to get adequate results; however, their rules prevent them from fully integrating detail and contour information. Artifact difficulties are often inevitable due to down sampling and up sampling. To reduce the Gibbs phenomenon, stationary wavelet transform (SWT) image fusion approaches were presented [12], however poor rules hindered fusion of images. Additionally, a shift invariant approaches that include non-subsampled contourlet transform (NSCT) have gained attention. In terms of NSCT approach, Ganasala et al [13] used Laplacian operators and entropy on low and high coefficients. The NSCT based algorithms integrate low and high frequency coefficients via phase congruency, focused contrast, or Laplacian energy restrictions [14]. A popular shift-invariant image decomposition and reconstruction approach is such as nonsubsampled shearlet transform (NSST) uses a shear wave filter. According to Liu et al [15], the gradient factor improves coefficient optimization using the structure tensor and NSST. Singh et al [16] formed a cascaded model using ripplet transform and NSST to enhance direction information. The above methods reveal directional tissue features accurately and comprehensively. The huge difference in parameter efficiency among filters affects the stability of the fusion performance based on parameter selection [17]. This method use pulse coupled neural network (PCNN) which represents a simple subtle vision-based neural network. The global domain PCNN-based approach to image fusion can preserve complex features. However, activating neurons to increase efficiency short of training remains difficult.

Huang et al [18] stimulate each PCNN neuron with the image block's Laplacian energy. An adaptive PCNN method are used in high coefficient region of NSST for increasing PCNN efficiency and fusion quality [19]. Even though there are several PCNN-related fusion approaches, the optimisation of coefficient and threshold characterization are still being explored. Recently, convolutional neural network (CNN) related models are used to solve fusion of medical images problems [20]. The shortages of small sampling data as well as a difficulty for construction of deep neural networks have hampered their consistency. The multiscale domain recommends that CNN sampling or convolution can easily cause fusion phase information loss. Also, a fusion method on the basis of fuzzy radial basis operator neural network is introduced [21], which only activated input neurons using image domain pixel features, which may have limited neural network cognitive capacity and performance. Medical image fusion research emphases on shift-invariant multiscale transformations to create frequency sub-band-specific fusion rules. However, selecting the best fusion rules is difficult. An advanced SR-SCNN blending approach for image fusion [22] contains three phases. After entering complete source images into standard orthogonal matching pursuit, the super-resolution fusion result is generated utilising the max rule to enhance pixel localisation. Besides, each source image receives a special SCNN-based K-SVD dictionary learning approach. A technique improves image information extraction and fusion result sparsity due to its non-linear behaviour. Chao et al [23] develops a novel fusion approach utilising the DSWT and RBFNN. The method first use 2-level decomposition to split two images into seven segments with low- and high frequency sub-bands for DSWT processing. We used the upgraded RBFNN to replace incomplete parts in the same areas of the two images, taking into account the target's gradient and energy. An unsupervised sophisticated image fusion network by Xu et al [24]. This method uses artificial and profound restrictions to boost memory. Saliency and plenty are used to sustain subjective and intuitive qualities in the superficial limitation. The exclusive channels of a pre-trained encoder objectively describe distinguishing information in the deep-level constraint. A multiscale adaptive transformer (MATR) was used for fusion of medical images in unsupervised manner [25]. Instead of regular convolution, adaptive convolution modulates the convolutional layer on the basis of global complementary context. The global semantic extraction was enhanced through an adaptive Transformer, modelling long-range dependencies better. We use a multiscale network design to capture multimodal data at various sizes. Fu et al [26] present a fusion algorithm to fuse medial images. This network comprises fuser, feature extractor along with reconstructor. The feature extractor extracts multiscale features using three MSRPAN blocks. The reconstructor uses three convolutional layers to reconstruct fused features. It also provides the energy ration method for feature fusion. Goyal et al [27] use pixel-based fusion rules to fuse multimodal medical images using cross bilateral and edge-aware filtering. This method calculated final fusion rule weights in a novel way. Both source images are first filtered with a cross bilateral filter (CBF) that considers geometric proximity and neighbouring pixel grey levels. This method prevents edge smoothing by determining the kernel and filtering with one image and vice versa. Subtracting output of CBF from input images yields the detail images. The domain filter extracts smaller scale information from detailed images near large-scale features. A novel image fusion method to solve input image noise and poor contrast [28]. This enhancement algorithm uses CLAHE, BM3D, and the Chameleon Swarm algorithm. An adaptive parameter from the proposed image enhancement approach is utilised for decomposition of image into three augmented layers.

Zhang et al [29] proposes an end-to-end unsupervised learning fusion framework to address these problems. The MMIF implements feature-weighted directed learning for extracting complimentary information from the original images. The feature extraction framework evaluates feature differences at several levels, allowing the feature reconstruction framework for generating interactive weights and directly determine the fusion result. DFENet, a medical image fusion framework by Li et al [30] is a self-supervised blends CNN with vision transformer feature learning. The DFENet's encoder-decoder architecture allows it to train on large natural image datasets without special ground truth fusion images. This network has a feature fuser, encoder as well as decoder. The CNN as well as transformer modules helps for extraction of local and global image information in the encoder. The novel global semantic information aggregation module effectively combines the multiscale characteristics of transformer module by improving image quality and eliminating the need for up-sampling and concatenation. By combining a synthetic focus degree criterion and a specific kernel set, Lepcha et al [31] proposes an image fusion algorithm that performs well in noisy or low-contrast input images. Salient feature extraction initiates with a gaussian curvature filter (GCF) further improve image quality. The dual branch complementary feature injection fusion is proposed by Xie et al [32] using unsupervised CNN models and transformer methods. This method feeds the entire and segmented source images into an adaptive backbone network for capturing local

and global characteristics. As an auxiliary module, the method generates a multi-scale complementary feature extraction framework that emphasizes feature differences at each level for capturing apparent complementary details in source photographs. Song et al [33] present a medical imaging deep learning network model that merges Transformer architecture with an upgraded DenseNet module to overcome the above concerns. The method can be used on natural images. Transformer and dense concatenation reduce feature loss, improving feature extraction and reducing edge blurring.

This study presents a multiscale and pyramid-based approach to produce perceptually better fusion performance based on deep neural network (DNN) with morphological processing of residuals. In particular, each input image is split into Laplacian pyramid and the weight maps constructed from the neural network is split into Gaussian pyramid after morphological processing of residuals. The fusion process is taking place at each decomposition level. Besides, we employ a fusion process on the basis of local similarity for determining the fusion mode for the decomposed coefficients. A weighted-average fusion mode is used when there is a maximum level of similarity between the elements of the input images for preservation of important information. In this case, the weight maps are obtained from the neural network since they are more dependable than a measure based on coefficients. In contrast, a choose-max rule is implemented when there is little image content similarity because it retains the most prominent details from the original images. In this scenario, the outputs of neural network are less dependable, and the absolute values of the decomposed coefficients are utilised to measure pixel activity directly. The network is used to fuse medical images by expanding on these concepts. It is noteworthy that two fundamental methods are utilised in fusion of images i.e., a similarity-based fusion mode determination and the pyramid-based decomposition. In addition, an edge-preserving processing technique is incorporated which selects regions where edges should be retained by combining non-linear techniques with linear lowpass filtering. Based on the morphological processing of liner filter residuals, these regions are selected with the intention of finding important regions with edges that have the right size and high amplitude. Reconstruction operators and area opening are two morphological image processing techniques used to achieve in our method. In order to restore the edges original shape and the identified regions are combined with the results of lowpass filter. In addition, this method permits control over the contrast of they produced image, with four customizable factors influencing the processing result.

The reminder of the paper is organised as follows: In Section 2, we provide a systematic illustration of the proposed fusion method. Section 3 illustrate materials related to datasets, comparative methods, and evaluation measures. In Section 4, an experimental result is presented both visually and quantitatively along with detailed discussion. The final section illustrates the conclusion of the proposed study.

2. Proposed methodology

The network utilised in the proposed fusion approach is shown in Fig. 1. The branches of neural network are inhibited to having the same weights [34]. The convolutional layers as well as max poling layer are composed in each branch. In this network, we eliminate a fully connected layer from the network to construct a much smaller structure with the goal of reducing memory usage and improving computational efficiency. After concatenation, the 512 feature maps are directly connected with the two-dimensional vector. In single accuracy, the smaller mode utilises very less physical memory, which is significantly less than the model used in [34]. In the end, a two-way SoftMax layer receives this two-dimensional vector as source which then generates a probability distribution over two modules. These two modules, first patch 1 and second patch 0, correspondingly first 0 and second patch 1 which represents for two different kinds of normalised weight assessment results. Each probability modules indicates how likely it is that each weight will be assigned. Since the total of the two output probabilities in this case is 1, a possibility of individual module characterizes the weight allocated to each input patch. As described in [34], the network is trained using higher quality image patches and the corresponding blurred fluctuations. Based on the study, the spatial, dimension of the input patch are set to 16×16 during training phase. Random sampling and multiscale gaussian filtering are used in the formation of training. The SoftMax loss function is used as an optimisation goal, which can be reduce utilising the stochastic gradient descent method.

The widely used deep learning structure Caffe is utilised for the training process [35]. Refer [34] for information on network training for the proposed study for proper illustrations. Meanwhile the network includes a fully connected layers whose dimensions are static for both source and the output data since the source of the network needs to have the static size in order to guarantee the source data for the fully connected layers never changes. In case of image fusion, the source image with different sizes can be accommodated by image segmentation into overlapped patches and feeding each pair of patches into the network; however, this procedure significantly increases the number of redundant computations. In order to tackle this issue, the proposed method initially convert fully connected layers into the corresponding conv layer that consists of $8 \times 8 \times 512$ kernel [36]. After conversion, the source images of any size can be processed by the network collectively for generation of dense prediction map. Furthermore, each prediction map which is a two-dimensional vector incorporates the relative clarity details of the source patch pairs at the correlative location. Since that each forecast consists of two dimensions that have been adjusted to a sum of one, the output can reduce to initial weight and the subsequent source. In the end, we allocate the value as the weights for every pixel inside the patch position and computed a mean of the overlapping pixels to produce a weight map that resembles the dimensions of the original images. Fig. 1 demonstrates the schematic representation of proposed fusion algorithm. The algorithm could be described in four stages.

2.1. DNN based weight map generation

Insert two input images A and B into the corresponding neural network branches [37]. Detailed process is described in above section for weight map generation W.

2.2. Morphological processed residuals

This section provides a procedure of the morphological processing of residuals [38,39]. It is related to residual of the gaussian filter:

$$I = W * \mathscr{L} \tag{1}$$

where W stands for a weight map from neural network, * for a convolutional function, and L denotes mask of the gaussian filter (or any other lowpass filter). Besides, the residual of the liner filtering is defined as follows:

$$Res(W) = W - I \tag{2}$$

Operators defined on weight map with positive values are applied to the residual in order to do further processing. As a result, the Res (W), which contains both positive and negative values which can be divided into two parts: positive and negative values.

$$\begin{aligned} I_{res+} &= 0.5(Res(W) + |Res(W)|) \\ I_{res-} &= 0.5(Res(W) - |Res(W)|) \end{aligned} \tag{3}$$

The following apparent relationship is satisfied by the proportion of the residual:

$$Res(W) = I_{res+} + I_{res-} \tag{4}$$

Both fractions of the residual (I_{res+}, I_{res-}) are processed based on the residual's amplitude in order to remove negligible fluctuations while keeping significant ones. The process relies on the morphological

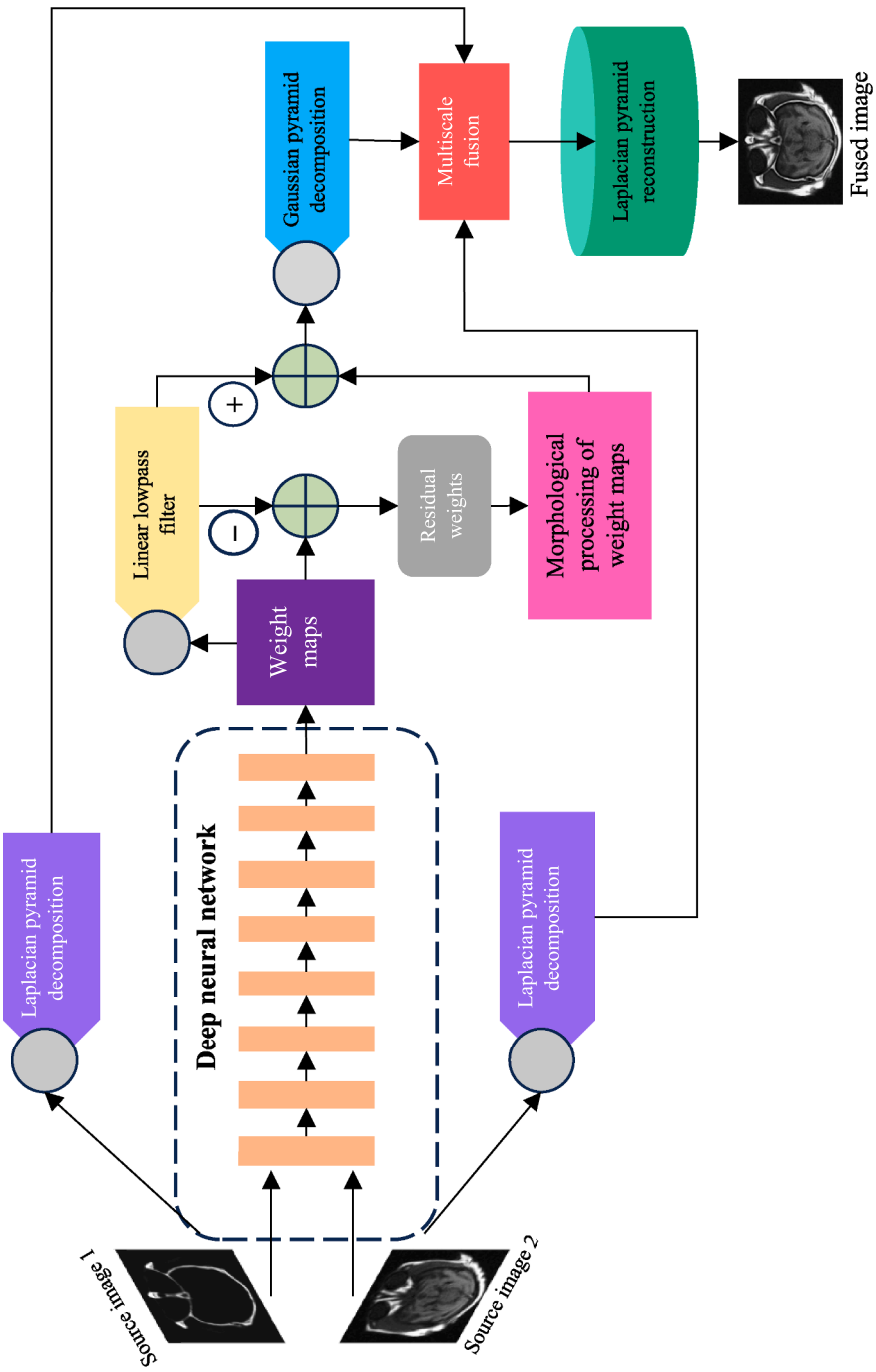

Fig. 1. Methodological flowchart of the proposed method.

function M that maintains the original structures of the residuum while selecting important regions based on the reconstruction. In order to retain relevant borders and blurriness in irrelevant areas, the weight map is finally updated to include large residual regions.

$$I_{out} = I + \mathscr{M}(I_{res+}) - \mathscr{M}(I_{res-}) \tag{5}$$

For both residuals, the operator M is specified as

$$\mathscr{M}(I) = R_I\left(\min\left(I, S_t(I)|_{\{\min\{I\},\ \max\ \{I\}\}}\right)\right) \tag{6}$$

where "|" indicates a mapping operator that generates grey level images by substituting two specified values for the original binary values of 1 s and 0 s, and R_I (A) refers to morphological restoration of the grey level mask I from the marker A. Finally, for two weight maps, '*min*' stands for the pointwise minimal function. Besides, S denotes a mask including significant components, and the concept of M is primarily concerned with contrast preservation (Eq. (6)). A residual amplitude serves as the basis of the substitute of the regions:

$$S_t(I) = \ (I\ \geq t) \tag{7}$$

where S is a selection function that extracts I segments whose amplitude is greater than the specified threshold t. An illustration of image filtering using the previously discussed method was presented in [38]. The test images are filtered using different t values, and the results are shown. Furthermore, a mask is defined as follows: grey denotes regions where both masks equal zero, while white indicates the positive mask $S_t(I_{res(+)})$ and black denotes the negative mask $S_t(I_{res(-)})$. By raising the threshold, fewer components are detected and subsequently restored and used to get the original image data. Ultimately, the quantity of parts displaying their initial clarity is reconstructed. The first function of this method is a mask that is created through thresholding and defines the limits of pertinent elements. A measure of meaningfulness is the amplitude of residuals. The importance of a section of a weight map is defined by more than just amplitude, though. A component of an image with a larger residual amplitude that is unimportant for visual comprehension can be readily imagined. A source image that contains salt and pepper, for example, produces a number of small, high-amplitude image elements, which alters the undesirable details by addition of high amplitude elements. It will be recognized as important but unpleasant places as a result. A further step has been added to our plan to address the previously noted problems. To filter the mask, an area opening filter [40] is utilised. This filter removes all related elements in the image that are less than a given threshold size (i.e. size coefficient). Theus, the expansion of Eq. (7) is described as follows:

$$S_{t,s}(I) = \ (I\ \geq t)\ o\ (s) \tag{8}$$

The area opening which eliminates coefficients smaller than the size represented indicated by s is recognized by the notation $o(s)$. The impact of utilising the size coefficients s and t is illustrated in [38]. It is clear that the number of selected regions reduces as the factors s and t increases. As a result, it allows objects with a small number of pixels to be rejected from the residual even if their amplitude is large. In the end, it makes it easier to maintain these regions in a hazy form in the final weight maps. Addition of (or removal, relies on filter mask coefficients) highpass filtering from the image itself is a conventional process to enhance visual contrasts. This relates to the high pass filtering characteristic that makes it easier to identify local differences in the image pixel values. Alternative technique to obtain high pass filtering outputs is to distinguish between lowpass filter and the image itself. The morphological processing of residuals concerns high frequency components of images which include areas with amplitudes higher than threshold t. Eq. (5) is changed by adding the contrast control coefficient (c) to regulate contrast of the final fused result, which defines in the following formula:

$$W_{out} = I + \ (\mathscr{M}(I_{res+}) - \mathscr{M}(I_{res-})).c. \tag{9}$$

Depending on the value of c, the contrast is either preserved ($c = 1$); enhanced ($c > 1$) or decreased ($0 < c < 1$). Examples indicating how the contrast control coefficient (c) influences the processing results is presented in [38]. Besides, contrast enhancement results are shown without the processing of morphological residuals. The influence of morphological processing on the amount of subtle image details visible in the output is clear in [38]. This processing technique makes it easier to ignore less important information by improving the contrast of relevant weight map regions in images.

2.3. Laplacian and gaussian pyramid decomposition

All the source images are decomposing into a Laplacian pyramid [41]. The pyramids of source images A and B are represented by $S\{A\}^l$ and $S\{B\}^l$, where l is the lth-decomposition level. After morphological processing the residuals of the weight map W, the weight map is further decomposing into a Gaussian pyramid $R\{W\}^l$. For each pyramid, the overall decomposition level is set at the maximum probable values $\lfloor \log_2 \min(H, W) \rfloor$ in which $H \times W$ is the spatial dimensions of the source image and $\lfloor \cdot \cdot \rfloor$ indicates the flooring function.

2.4. Fusion of coefficients

A local energy map (i.e. the sum of the squares of coefficients inside a small window) [42] of $S\{A\}^l$ and $S\{B\}^l$, respectively is computed for each decomposition level l [37].

$$\begin{aligned} K_A^l(u,v) &= \sum_m \sum_n S\{A\}^l(u+p,\ v+q)^2, \\ K_B^l(u,v) &= \sum_m \sum_n S\{B\}^l(u+p,\ v+q)^2, \end{aligned} \tag{10}$$

The similarity measure utilised to identify the fusion mode is computed as

$$Z^l(u,v) = \frac{2\sum_p \sum_q S\{A\}^l(u+p,\ v+q)\ S\{B\}^l(u+p, v+q),}{K_A^l(u,v) + K_B^l(u,v)} \tag{11}$$

This measure has a range of $[-1,\ 1]$, where values nearer 1 indicate greater similarity. To decide which fusion mode should be used, a threshold t is set. A weighted average fusion mode using weight map W is used if $Z^l(u,v) \geq t$.

$$S\{F\}^l(u,v) = \ R\{W\}^l(u,v)\ .\ S\{A\}^l(u,v) + \ \left(1 - R\{W\}^l(u,v)\right)\ .\ S\{B\}^l(u,v) \tag{12}$$

Eq. (10), which compares the local energy, is utilized to determine the fusion mode if $Z^l(u,v) < \ t$:

$$S\{F\}^l(u,v) = \begin{cases} S\{A\}^l(u,v), \ \text{if}\ K_A^l(u,v)\ \geq\ K_B^l(u,v) \\ S\{B\}^l(u,v), \ \text{if}\ K_A^l(u,v) < \ K_B^l(u,v) \end{cases} \tag{13}$$

The Eq. (14) summarizes the fusion strategy in its entirety. The final fusion result is obtained using Laplacian pyramid reconstruction $S\{F\}^l$.

$$S\{F\}^l(u,v) = \begin{cases} R\{W\}^l(u,v)\ .\ L\{A\}^l(u,v) + \ \left(1 - R\{W\}^l(u,v)\right)\ .\ S\{B\}^l(u,v), \ \text{if}\ Z^l(u,v) \geq t \\ S\{A\}^l(u,v), \ \text{if}\ Z^l(u,v) < t \ \text{and if}\ K_A^l(u,v)\ \geq\ K_B^l(u,v) \\ S\{B\}^l(u,v), \ \text{if}\ Z^l(u,v) < t \ \text{and if}\ K_A^l(u,v) < \ K_B^l(u,v) \end{cases} \tag{14}$$

The adoption of Laplacian and Gaussian pyramid decomposition for multi-scale representation as well as the use of local energy and similarity measures for coefficient analysis are adapted from prior image fusion research. These elements are well-recognized in the literature and have been shown to effectively capture structural details and assess saliency. The novelty of the present work lies in three key aspects. First, we introduce morphological processing of residuals derived from the DNN-generated weight map which allows selective preservation of clinically relevant edges while suppressing noise, with an additional tunable contrast control mechanism for enhancing visibility. Second, we propose a hybrid adaptive fusion rule that dynamically switches between similarity-based weighted fusion and coefficient-based choose-max fusion, thereby integrating the advantages of learned features with robust model-driven decisions. Third, we implement the conversion of fully connected layers to convolutional layers in the training network, enabling efficient generation of dense per-pixel weight maps and avoiding redundant patch-wise processing. Together, these innovations distinguish our approach from conventional pyramid-based or morphology-based fusion methods and constitute the principal contributions of this study.

Algorithm: Medical Image Fusion (DNN + Morphological Residuals)

Input: Two source images A, B
Output: Fused image F

1. Weight Map Generation:
   W = Net.forward(A, B) # DNN produces dense weights
2. Morphological Residual Processing:
   I = conv (W, Gaussian)
   Res = $W - I$
   Res_{pos}, Res_{neg} = split (Res)
   M_{pos} = MorphReconstruct (Res_{pos}, t_{amp}, S_{area})
   M_{neg} = MorphReconstruct (Res_{neg}, t_{amp}, S_{area})
   $W_{ref} = I + c * (\text{M}_{pos} - \text{M}_{neg})$ # contrast control
3. Pyramid Decomposition:
   S_A= LaplacianPyramid(A)
   S_B= LaplacianPyramid(B)
   R_W = GaussianPyramid (W_{ref})
4. Adaptive Fusion (per level l):
   Compute local energy K_A, K_B and similarity Z
   If $Z \geq \tau_{sim}$:
   S_F [l] = R_W [l]* S_A [l] + (1−R_W [l]) * S_B [l]
   Else:
   S_F [l] = max (S_A [l], S_B [l]) by local energy
5. Reconstruction:
   F = LaplacianReconstruct (S_F)
   Return F

3. Materials and data analysis

We used three different types of image dataset pairs (CT-MRI, MRI-SPECT and MRI-PET) to scientifically validate the proposed methodology. Every image is taken from the medical imaging resource The Whole Brain Atlas [43]. Every image is 256 × 256 in size. For each dataset type, we chose 60 image pairings for training and 25 pairs are utilised for testing, i.e. the proposed neural network is trained using 60 pairs of CT-MRI data pairs while using CT-MRI dataset for experiments. We examined 25 data pairs for every type of dataset after training. Nine standard mainstream approaches were used for enhancing the comparison such as enhanced medical image fusion network (EMFusion) [24], discrete stationary wavelet transform and the enhanced radial basis function neural network (DSWT-RBFNN) [23], multiscale adaptive transformer (MATR) [25], feature difference guided network (FDGNet) [29], and dual branch feature enhanced network (DFENet) [30], semantic-preserving fusion (SMFusion) [44], progressive parallel strategy based on deep learning (PPMF-Net) [45], a mamba-based dual--phase model (Mambadfuse) [46]. To improve scientific validation, we assessed each algorithm using five standard quantitative image fusion assessment metrics: Tsallis entropy (TN) [47] that computes the information incorporated in fusion image; mutual information (MI) [4], computes information shared across a fused image and the input images; QAB/F [4], which measures edge similarity based on gradient congruency; spatial frequency (SF) [5], which computes the level of details and textures in the fused image; and the edge preservation index (EPI) [5]. A higher value for each of the above-mentioned criteria denotes a superior fusion performance.

4. Results and discussions

The fused images were generated for every pair of datasets that were considered for fusion as shown in Fig. (2). Figs. (3–9) show the fusion performances of seven dataset pairs using nine different methods. Fig. 3 (c-k) denotes the fusion results generated from the EMFusion, DSWT-RBFNN, MATR, FDGNet, DFENet, Mambadfuse, PPMF-Net SMFusion, and Proposed, respectively. Fig. 3 (a-b) represents two original images considered for fusion. We started by observing at the CT-MRI dataset where Figs. (3–5) show the fusion results of CT-MRI (Dataset 1–3). The CT and MRI (Dataset 1) incorporate of bronchogenic carcinoma. Fig. 3 (c–d) shows the poor fusion effects from the EMFusion and DSWT-RBFNN methods and the brightness of images are greatly distorted as well as the CT contour details are not adequately merged into the fusion performance and the details of soft tissue and bone are not easily noticeable especially with regard to the lesion. The overall fusion outcome remained reasonable even though A MATR-based approach enhanced the information quality representation when comparison to images generated through other two approaches. The features of the lesion are distinctively not visible in MATR method (see Fig. 3(e)). However, as shown in Fig. 3 (f-k), the fusion results of the six methods (i.e. FDGNet, DFENet, Mambadfuse, PPMF-Net, SMFusion and Proposed methods are noticeably better. Significant improvements in structural fidelity and image distortion prevention are made while maintaining the integrity of

A space-occupying sarcoma's pathological changes are represented in the CT-MRI Dataset (2–3). The MRI images represent the edge of the tumour in greater information, whereas the CT images represent the general form of the tumour (see Fig. 4–5). Therefore, it is vital to fuse both the images to accomplished greater detailed images. The proposed algorithm successfully integrates the benefits of CT-MRI images (shown in blue and green circles in Fig. 4 and 5(k) by producing fusion results that are expressively superior than comparative methods. These fused images are illustrious by enhanced image contrast and a complementary broad representation of the edges of the tumour structure. As shown in Figs. (6–7), the fusion results from MRI – SPECT (Dataset 1–2) represents how well the integration of soft tissue and metabolic activity occurs. The intrinsic features and rich feature information of the images are not well represented by the MATR and DFENet based algorithms. In particular, the energy information shows increasing distortion compared to the original SPECT images (Figs. 6 and 7 (c–d)). However, the outputs of the original images generated by the other four algorithms (Mambadfuse, PPMF-Net, SMFusion, and the proposed) presents superior in case of maintaining the energy detail of SPECT images. Besides, when compared of these four approaches, the fusion efficiency of our algorithm clearly outperforms the others by producing sensitive details that are clearer and more comprehensive information features (Red circles in Figs. 9 and 10 (e-*i*) specify the performance of specific details). Figs. (8–9) display the fused images from the MRI-PET Dataset (1–2). Subcortical blood vessel changes are better explained by MRI. A better depiction of metabolism is offered by Positron Emission Tomography (PET). The properties of PECT images are not adequately presented in Figs. 8 and 9 (c–d), which were derived from the MATR and DFENet approaches. There is less contrast as a result of the overall visual brightness being lowered. The fusion consequences are therefore less than ideal. The fusion results of Figs. 8 and 9(h) attained from the proposed algorithm are higher (as shown by red circles) regardless of the energy restoration of the functional images as well as the soft tissues, blood arteries, and

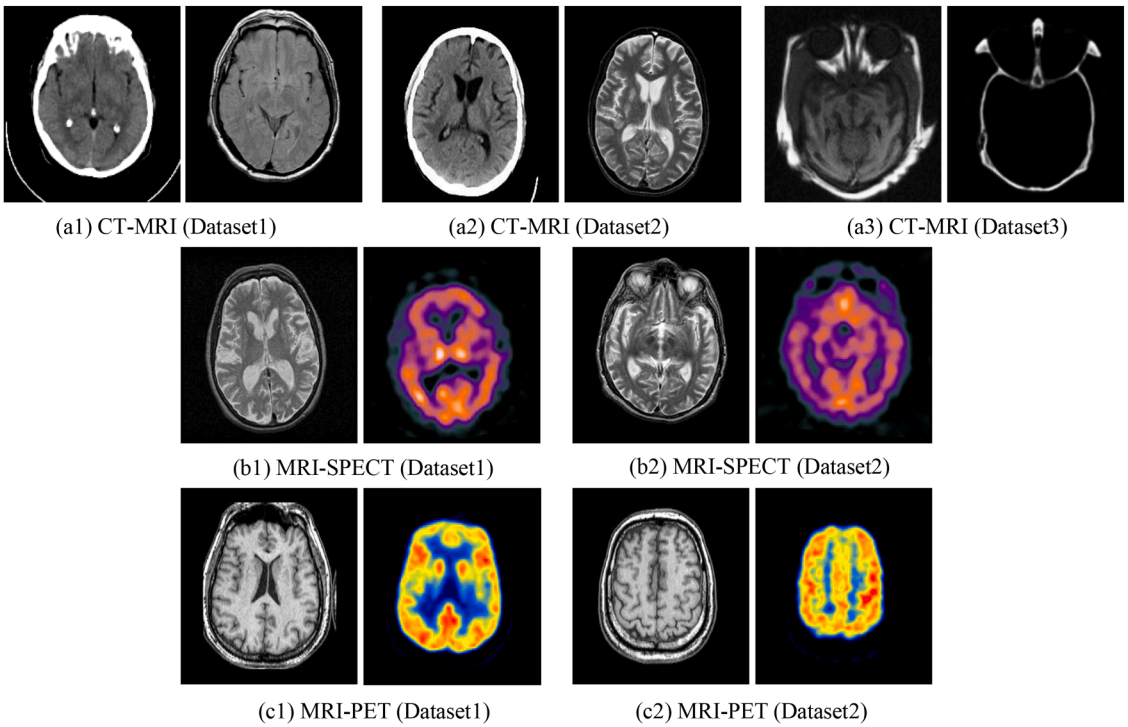

Fig. 2. Experimental pair of source images: (a) CT-MRI; (b) MRI-SPECT and (c) MRI-PET.

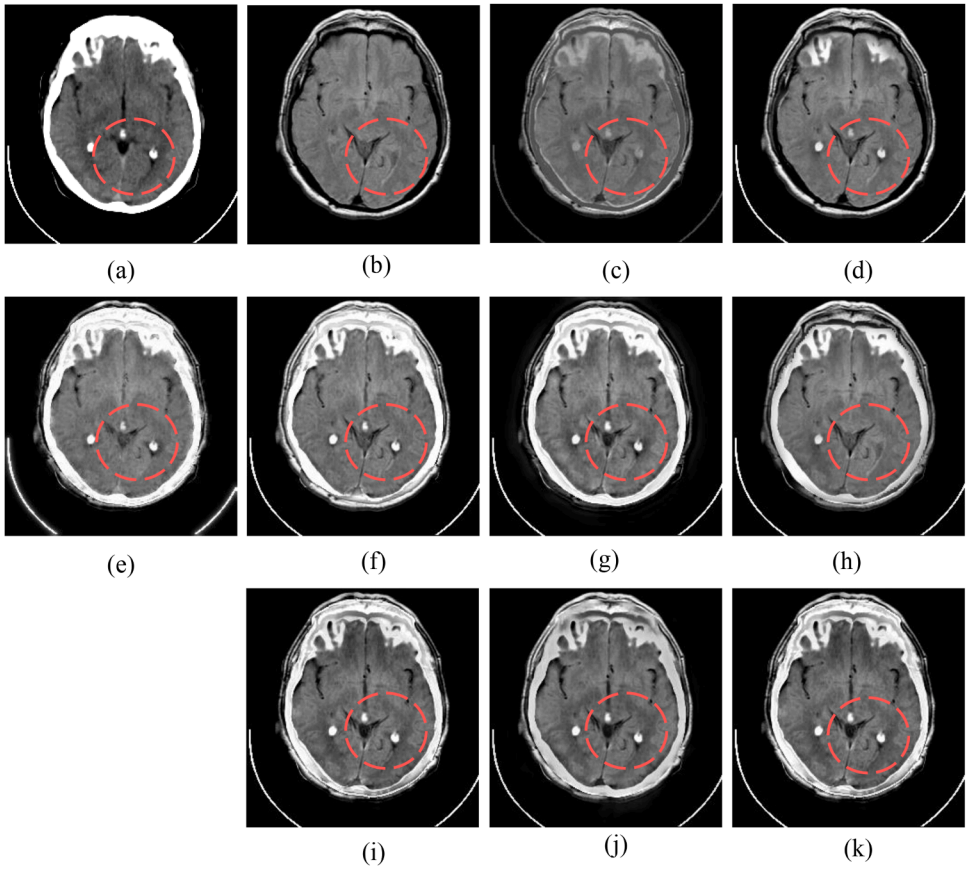

Fig. 3. Fusion performance of CT-MRI Dataset 1 based on nine algorithms: (a) CT; (b) MRI; (c) EMFusion; (d) DSWT-RBFNN; (e) MATR; (f) FDGNet; (g) DFENet; (h) Mambadfuse; (i) PPMF-Net; (j) SMFusion and (k) Proposed, respectively.

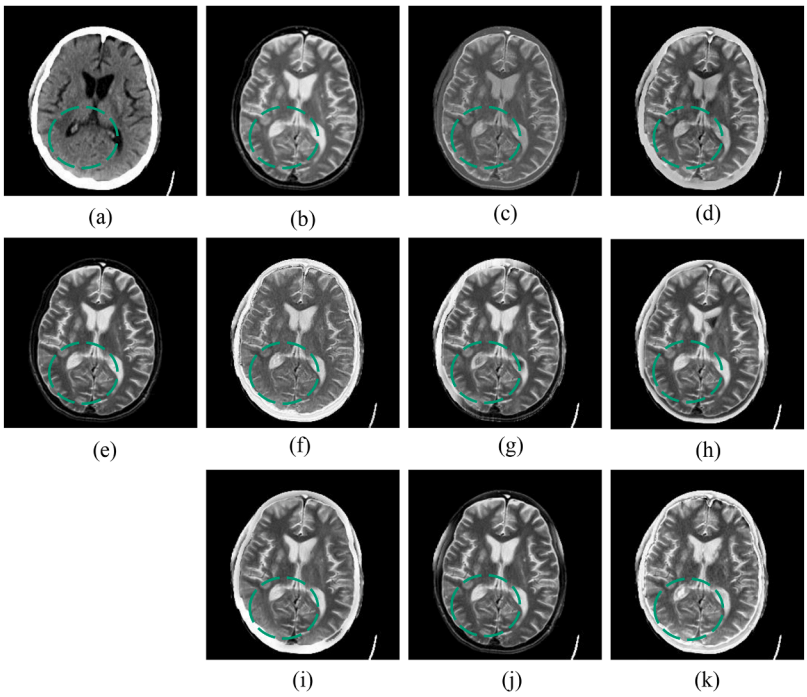

Fig. 4. Fusion performance of CT-MRI Dataset 2 based on nine algorithms: (a) CT; (b) MRI; (c) EMFusion; (d) DSWT-RBFNN; (e) MATR; (f) FDGNet; (g) DFENet; (h) Mambadfuse; (i) PPMF-Net; (j) SMFusion and (k) Proposed, respectively.

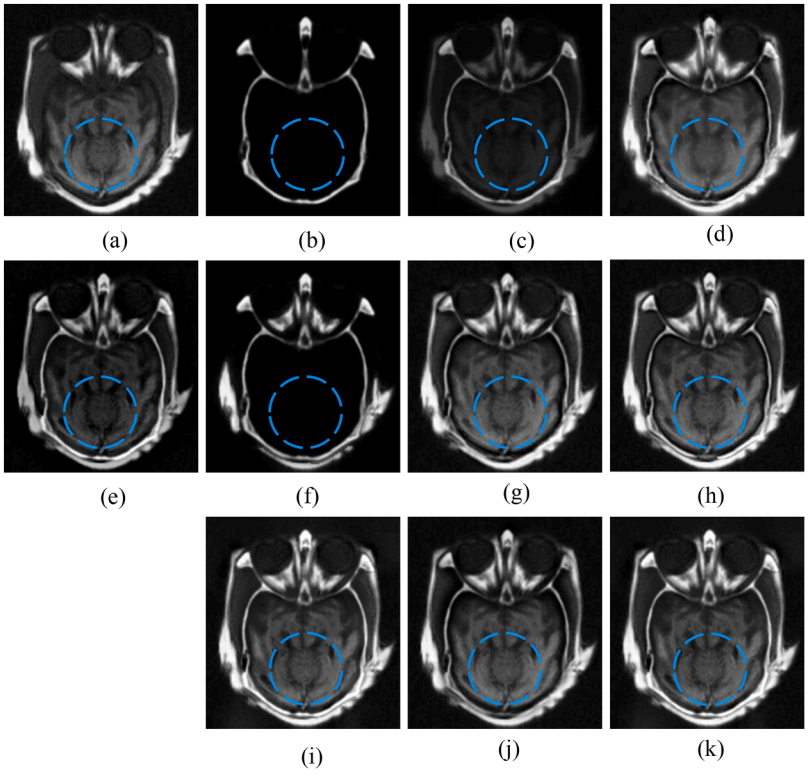

Fig. 5. Fusion performance of CT-MRI Dataset 3 based on nine algorithms: (a) CT; (b) MRI; (c) EMFusion; (d) DSWT-RBFNN; (e) MATR; (f) FDGNet; (g) DFENet; (h) Mambadfuse; (i) PPMF-Net; (j) SMFusion and (k) Proposed, respectively.

other information represented in the MRI images.

The evaluation metrics EPI, MI, SF, TE, and QAB/F were and MRI-PET test data pairs. The mean fusion performance of five metrics across all datasets is shown in Tables (1–3). The objective comprehensive performances obtained from the Mambadfuse, PPMF-Net, SMFusion methods demonstrated good results in terms of visual appraisal, border fusion effect evaluation, and detailed information evaluation. The proposed algorithm outperformed other methods in nearly all of the five metrics such as SF, TE, and QAB/F across all datasets according to a comparison of their performances. The proposed approach also presents

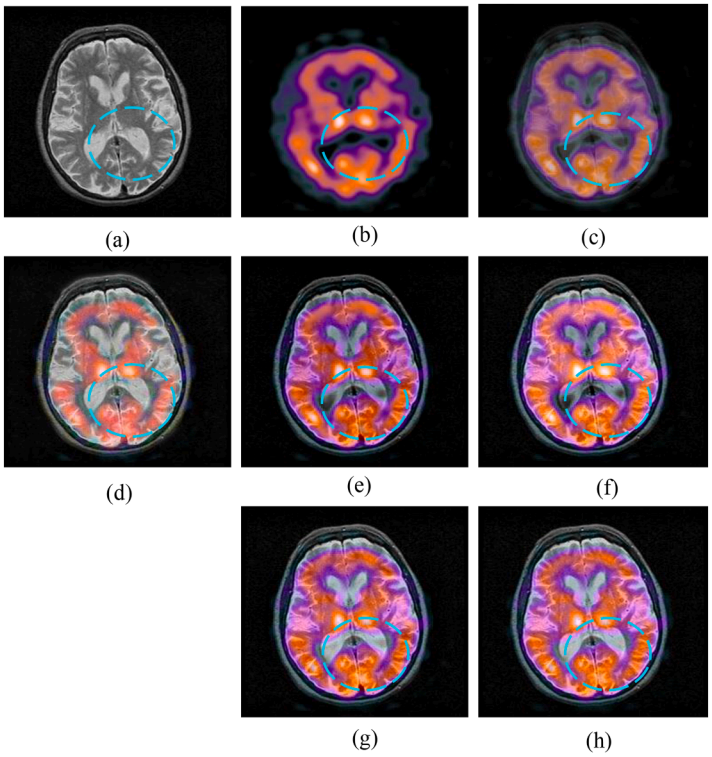

Fig. 6. Fusion performance of MRI-SPECT Dataset 1 based on six algorithms: (a) MRI; (b) SPECT; (c) MATR; (d) DFENet; (e) Mambadfuset; (f) PPMF-Net (g) SMFusion and (h) Proposed, respectively.

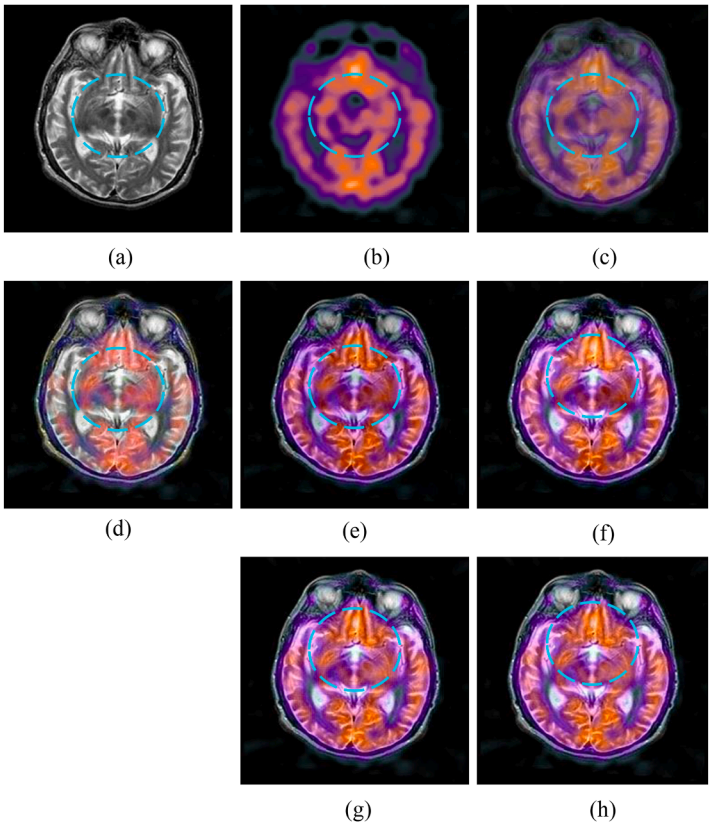

Fig. 7. Fusion performance of MRI and SPECT Dataset 2 based on six algorithms: (a) MRI; (b) SPECT; (c) MATR; (d) DFENet; (e) Mambadfuset; (f) PPMF-Net (g) SMFusion and (h) Proposed, respectively.

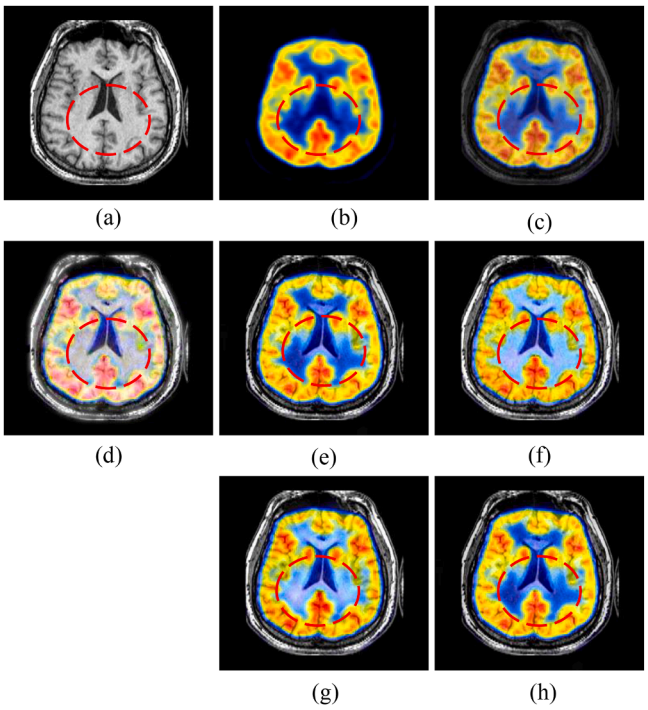

Fig. 8. Fusion performance of MRI-PET Dataset 1 based on six algorithms: (a) MRI; (b) PET; (c) MATR; (d) DFENet; (e) Mambadfuset; (f) PPMF-Net (g) SMFusion and (h) Proposed, respectively.

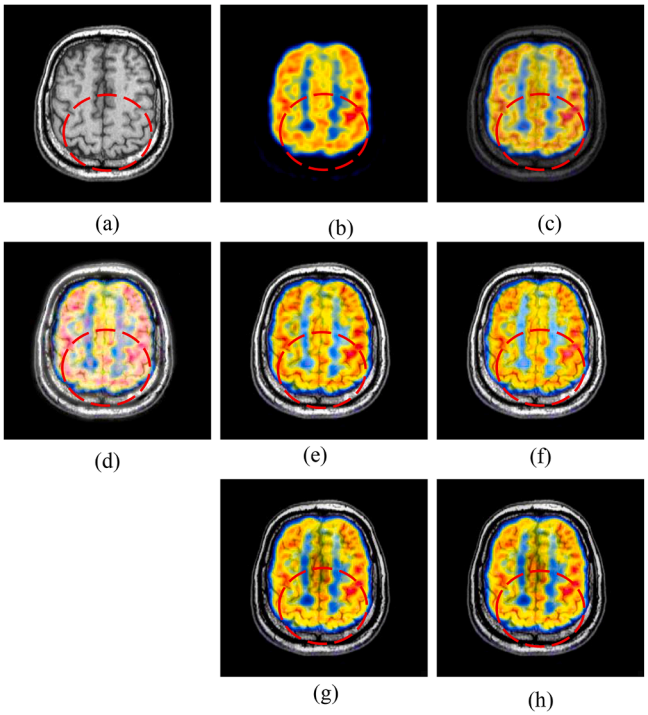

Fig. 9. Fusion performance of MRI and PET Dataset 2 based on six algorithms: (a) MRI; (b) PET; (c) MATR; (d) DFENet; (e) Mambadfuset; (f) PPMF-Net (g) SMFusion and (h) Proposed, respectively.

outstanding performance for EPI and MI for all datasets. The effectiveness of our approach was marginally better to that of PPMF-Net and SMFusion -based methods in all datasets. Subsequently, from a general performance standpoint, the proposed strategy demonstrated superior results in objective analysis which indicates exceptional robustness in fused images. Fig. 10 (a-e) represent an experimental result of different values of Threshold (t) on metrics EPI, MI, QAB/F, SF, TE for CT-MRI images. One can see that when $t = 0.05$, $t = 0.08$, the performance is the better in most scenarios as shown Fig. 10.

The fusion performances of the proposed approach are shown by the final outcomes. The outcomes need to be confirmed by more research that includes a thorough comparison of different imaging types and decomposition levels. The benefits of the DNN-based method occur from its rich deeper architecture and considerable training set. Thus, we maintain the sufficient training data is the key reasons for the improvement of the DNN-based approach. To make the comparison more precise, we trained the DNN based method using the same number of data points as the proposed method. By modifying the neural network design to meet specific experimental needs and choosing an efficient training strategy, we observed that the proposed method permitted the model to exhibit intelligence and self-learning capabilities. The determination of the properties of the input networks is a vital stage in the development of the network design. By combining the pixel and regional block levels of medical image fusion, we were able to achieve a complimentary effect by combining the specific pixel regions and pertinent characteristic properties of the one input image with those of another source image in the same locations. As well as the correlation across pixels and neighbouring pixel values within the regions, the pixel value and the regional energy indicate differences in the properties of pixel characteristics. The selection of these properties from the many subbands gathered by decomposition level is intended to enhance the potentiality of the input layers of network to correctly recognises the signal without causing self-confusion. We used several dimensions in a comparison analysis of 25 pairs of different set of datasets (CT-MRI, MRI-SPECT, MRI-PET) to find the optimal regions. Figures (3–9) shows how a dataset performs visually across different categories. In addition, Tables (1–3) analysed the quantitative performance of the different measures assessed across 25 data pairs shows that our approach produces superior and optimal fusion results. The proposed approach is able to achieved better performance in comparison to recent Mambadfuse, PPMF-Net and SMFusion methods. Medical image fusion the process utilised to complementarily fuse images composed from different medical image technologies. By integrating anatomical as well as functional information, the fusion result presents detailed information from different singe mode imaging modalities without sacrificing any of the energy and information from several images. In the domain of medical image fusion, multiscale transform-based methods are progressively gaining attraction. Also, the efficiency of the fusion procedure is directly impacted by the generation of efficient fusion rules based on the subbands of different coefficients. Despite the simplicity of the fusion rules that are produced by these two approaches, the refinements and features of the original images are not sufficiently recognized and recovered which causes image distortion or degradation. Due to shift variance, the EMFusion method is severely hampered which may result in aliasing and information loss. However, the DSWT-RBFNN based algorithms presented better fusion performance; however, the problem of selecting the best parameters for algorithm activation still exists. There is compromise regarding the utilisation of shift invariant multiscale transforms that is based on processing at the pixel level. Nevertheless, the problems with selection of the appropriate thresholds and parameters or generating suitable fusion rules suffer. We developed fusion rules in this study using neural networks and presented a hybrid approach.

In terms of computational efficiency, the proposed method is designed to minimize overhead by converting the fully connected layers of the network into convolutional layers, which allows dense prediction without redundant patch-wise processing and thereby reduces memory usage and accelerates inference. The additional morphological residual processing step involves only lightweight morphological operators and contributes negligible extra cost. As a result, the inference speed of the proposed approach is competitive with or faster than recent transformer- and dictionary-based fusion methods, while moderately higher

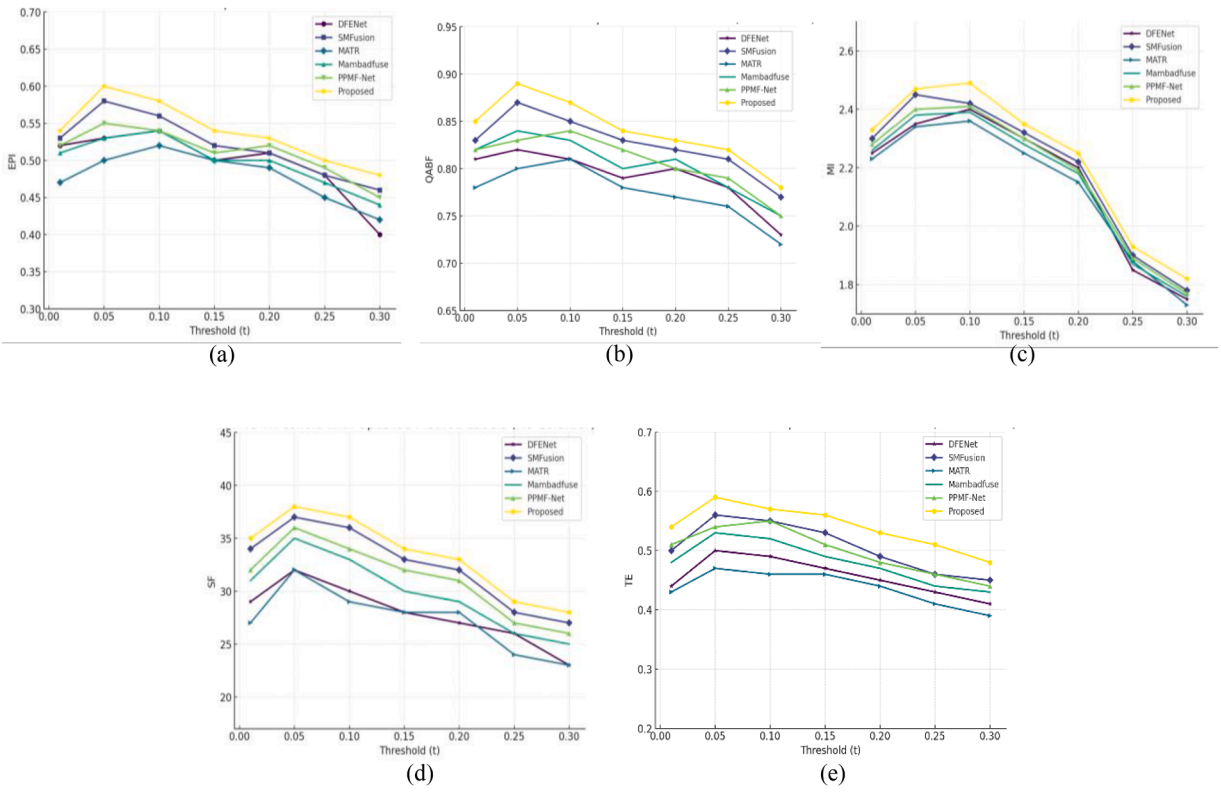

Fig. 10. Experimental performance on different values of Threshold (t) on (a) EPI, (b) QABF, (c) MI, (d) SF, and (e) MRI-PET dataset.

than classical multiscale transforms due to the neural component. Overall, the method achieves a favorable balance between computational cost and fusion performance, ensuring practicality for clinical use.

5. Ablation study

To better understand the contribution of the two core components of the proposed framework such as morphological residual processing (MRP) and the multi-scale pyramid (MSP) design while we carried out an ablation study by selectively removing them from the pipeline and observing the impact on fusion quality. When the morphological residual processing step was excluded, the fused results exhibited reduced sharpness in lesion boundaries and loss of subtle anatomical structures. This that MRP plays a critical role in preserving edges, enhancing local contrast, and suppressing irrelevant noise fluctuations. Its inclusion ensures that diagnostically important regions, such as tumor margins and vascular details, remain well defined in the final fused images. In contrast, removing the multi-scale pyramid design led to a noticeable reduction in the richness of detail and information integration. Single-scale fusion was insufficient to capture both global structural information and fine textural patterns, resulting in fused images with compromised clarity and reduced perceptual quality. The MSP design thus contributes to multi-resolution feature extraction, enabling effective blending of coarse anatomical context with fine localized details. Finally, combining both MRP and MSP provided the most balanced results, delivering fused images with high structural fidelity, strong edge preservation, and comprehensive detail integration. This demonstrates that the two components are complementary and mutually reinforcing, and together they form the backbone of the robustness and clinical relevance of the proposed method.

A limitation of this study is the relatively small training set (60 pairs per modality), which could restrict feature diversity and raise concerns about generalization. To address this, we adopted patch-based training with overlapping 16 × 16 patches and integrated morphological residual processing, reducing dependence on purely data-driven learning.

Table 1
Quantitative evaluation of CT-MRI fusion images.

m	EPI	MI	$Q^{AB/F}$	SF	TE
EMFusion	0.3562	1.4425	0.5672	26.8715	0.3828
DSWT-RBFNN	0.3892	1.5626	0.6344	28.8162	0.4081
MATR	0.4252	1.6751	0.6726	32.8816	0.4312
FDGNet	0.4362	1.7718	0.7125	31.8373	0.4971
DFENet	0.4692	1.8826	0.7552	33.0081	0.4816
Mambadfuse	0.5352	1.9727	0.8077	34.8811	0.5244
PPMF-Net	0.5072	1.9923	0.8252	36.8817	0.5517
SMFusion	0.4987	2.0028	0.8352	35.9817	0.5431
Proposed	0.5862	2.4596	0.8445	37.8622	0.5775

Table 2
Quantitative evaluation of MRI-SPECT fusion images.

	EPI	MI	$Q^{AB/F}$	SF	TE
MATR	0.4627	2.5381	0.7625	17.9928	0.2344
DFENet	0.5177	2.6715	0.7982	19.9172	0.2554
Mambadfuse	0.4926	3.0018	0.8026	21.4413	0.2617
PPMF-Net	0.5673	3.2891	0.8055	20.1872	0.2803
SMFusion	0.5424	3.4324	0.8141	21.9272	0.2781
Proposed	0.5813	3.6772	0.8261	23.8229	0.2897

Table 3
Quantitative evaluation of MRI-PET fusion images.

	EPI	MI	$Q^{AB/F}$	SF	TE
MATR	0.4625	1.9972	72.9937	24.9198	0.3927
DFENet	0.4972	2.1888	74.8838	25.8827	0.4366
Mambadfuse	0.5193	2.3233	77.8272	27.1993	0.4628
PPMF-Net	0.5623	2.5728	80.9727	28.1433	0.4902
SMFusion	0.5565	2.5552	81.8827	28.3837	0.5321
Proposed	0.5896	2.6954	83.8272	30.5286	0.5451

These strategies, along with consistent improvements across CT-MRI, MRI-SPECT, and MRI-PET datasets (Tables 1–3), indicate that the model generalizes well despite limited data. Nonetheless, future work will expand training using larger public datasets, apply augmentation and transfer learning, and validate on multi-center clinical images to further strengthen robustness and applicability.

6. Conclusion

This paper introduces a robust medical image fusion framework that leverages a deep neural network in combination with morphological processing of residuals. The core of the approach lies in the generation of adaptive weight maps through a deep neural network, which effectively captures and integrates pixel-level activity information from the source images. To align the fusion process with human visual perception, image pyramids are employed to achieve multiscale representation, ensuring consistent integration of both fine and coarse details. The fusion mode is further refined using a local similarity–based strategy that adaptively adjusts for decomposed image components. A key element of the methodology is the incorporation of edge-preserving processing, designed to maintain critical structural boundaries. This is accomplished by combining nonlinear algorithms with linear low-pass filtering to identify and preserve regions with high amplitude and optimally scaled edges. The morphologically processed residuals derived from linear filter outputs serve as a reliable basis for selecting significant regions. Morphological operations such as reconstruction and area opening are employed to retain the original shape of edges, while blending the low-pass filter results with the selected regions ensures accurate structural preservation. Moreover, the method provides flexibility for contrast adjustment in the fused image, enhancing its diagnostic utility. Comprehensive experiments on diverse medical datasets demonstrate that the proposed framework outperforms several state-of-the-art techniques, delivering fused results with improved information richness, structural fidelity, and preservation of fine details from the source modalities.

Data availability

The dataset used in this study is publicly available from the Harvard Medical School MIF Database and can be accessed at: http://www.med.harvard.edu/AANLIB/home.html.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

CRediT authorship contribution statement

Supinder Kaur: Conceptualization, Data curation, Formal analysis, Funding acquisition. **Parminder Singh:** Formal analysis, Investigation. **Rajinder Vir:** Project administration, Resources, Software. **Arun Singh:** Writing – review & editing, Writing – original draft, Methodology. **Harpreet Kaur:** Supervision, Validation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We are thankful to the reviewers and editors for their valuable suggestions. We appreciate their insightful comments which helped us in improving the manuscript significantly.

References

[1] M.A. Azam, et al., A review on multimodal medical image fusion: compendious analysis of medical modalities, multimodal databases, fusion techniques and quality metrics, Comput. Biol. Med. 144 (May 2022) 105253, https://doi.org/10.1016/J.COMPBIOMED.2022.105253.
[2] S. Ullah Khan, M. Ahmad Khan, M. Azhar, F. Khan, Y. Lee, M. Javed, Multimodal medical image fusion towards future research: a review, J. King Saud Univ. - Comput. Inf. Sci. 35 (8) (Sep. 2023) 101733, https://doi.org/10.1016/J.JKSUCI.2023.101733.
[3] S. Basu, S. Singhal, D. Singh, A systematic literature review on multimodal medical image fusion, Multimed. Tools. Appl. 83 (6) (Feb. 2024) 15845–15913, https://doi.org/10.1007/S11042-023-15913-W/TABLES/9.
[4] B. Goyal, D. Chyophel Lepcha, A. Dogra, V. Bhateja, A. Lay-Ekuakille, Measurement and analysis of multi-modal image fusion metrics based on structure awareness using domain transform filtering, Measurement 182 (Sep. 2021) 109663, https://doi.org/10.1016/J.MEASUREMENT.2021.109663.
[5] D.C. Lepcha, et al., Multimodal medical image fusion based on pixel significance using anisotropic diffusion and cross bilateral filter, Hum.-centric Comput. Inf. Sci. 12 (2022), https://doi.org/10.22967/HCIS.2022.12.015.
[6] G. Pajares, J.M. de la Cruz, A wavelet-based image fusion tutorial, Pattern. Recognit. 37 (9) (Sep. 2004) 1855–1872, https://doi.org/10.1016/J.PATCOG.2004.03.010.
[7] P.J. BURT, E.H. ADELSON, The Laplacian Pyramid as a Compact Image Code. Readings in Computer Vision, Jan. 1987, pp. 671–679, https://doi.org/10.1016/B978-0-08-051581-6.50065-9.
[8] Y. Yang, D.S. Park, S. Huang, N. Rao, Medical image fusion via an effective wavelet-based approach, EURASIP. J. Adv. Signal. Process. 2010 (1) (Apr. 2010) 1–13, https://doi.org/10.1155/2010/579341. *2010 2010:1*.
[9] H. Tian, Y.N. Fu, P.G. Wang, Image fusion algorithm based on regional variance and multi-wavelet bases, in: Proceedings of the 2010 2nd International Conference on Future Computer and Communication, ICFCC 2010 2, 2010, https://doi.org/10.1109/ICFCC.2010.5497628.
[10] X. Xu, Y. Wang, S. Chen, Medical image fusion using discrete fractional wavelet transform, Biomed. Signal. Process. Control 27 (May 2016) 103–111, https://doi.org/10.1016/J.BSPC.2016.02.008.
[11] Y. Liu, S. Liu, Z. Wang, A general framework for image fusion based on multi-scale transform and sparse representation, Inf. Fusion 24 (Jul. 2015) 147–164, https://doi.org/10.1016/J.INFFUS.2014.09.004.
[12] P. Ganasala, A.D. Prasad, Medical image fusion based on laws of texture energy measures in stationary wavelet transform domain, Int. J. ImAging Syst. Technol. 30 (3) (Sep. 2020) 544–557, https://doi.org/10.1002/IMA.22393.
[13] P. Ganasala, V. Kumar, CT and MR image fusion scheme in nonsubsampled contourlet transform domain, J. Digit. Imaging 27 (3) (Jan. 2014) 407–418, https://doi.org/10.1007/S10278-013-9664-X/METRICS.
[14] Z. Zhu, M. Zheng, G. Qi, D. Wang, Y. Xiang, A phase congruency and local LAPLACIAN energy based multi-modality medical image fusion method in NSCT domain, IEEe Access. 7 (2019) 20811–20824, https://doi.org/10.1109/ACCESS.2019.2898111.
[15] X. Liu, W. Mei, H. Du, Structure tensor and nonsubsampled shearlet transform based algorithm for CT and MRI image fusion, Neurocomputing. 235 (Apr. 2017) 131–139, https://doi.org/10.1016/J.NEUCOM.2017.01.006.
[16] S. Singh, R.S. Anand, D. Gupta, CT and MR image information fusion scheme using a cascaded framework in ripplet and NSST domain, IET. Image Process. 12 (5) (May 2018) 696–707, https://doi.org/10.1049/IET-IPR.2017.0214.
[17] D. Gai, X. Shen, H. Cheng, H. Chen, Medical image fusion via PCNN Based on edge preservation and improved sparse representation in NSST domain, IEEe Access. 7 (2019) 85413–85429, https://doi.org/10.1109/ACCESS.2019.2925424.
[18] W. Huang, Z. Jing, Multi-focus image fusion using pulse coupled neural network, Pattern. Recognit. Lett. 28 (9) (Jul. 2007) 1123–1132, https://doi.org/10.1016/J.PATREC.2007.01.013.
[19] C. Panigrahy, A. Seal, N.K. Mahato, MRI and SPECT image fusion using a weighted parameter adaptive dual channel PCNN, IEEe Signal. Process. Lett. 27 (2020) 690–694, https://doi.org/10.1109/LSP.2020.2989054.
[20] K. Wang, M. Zheng, H. Wei, G. Qi, Y. Li, Multi-modality medical image fusion using convolutional neural network and contrast pyramid, Sensors 20 (8) (Apr. 2020) 2169, https://doi.org/10.3390/S20082169, 2020, Vol. 20, Page 2169.
[21] Z. Chao, D. Kim, H.J. Kim, Multi-modality image fusion based on enhanced fuzzy radial basis function neural networks, Physica Medica 48 (Apr. 2018) 11–20, https://doi.org/10.1016/J.EJMP.2018.03.008.
[22] A. Sabeeh Yousif, Z. Omar, U.Ullah Sheikh, An improved approach for medical image fusion using sparse representation and Siamese convolutional neural network, Biomed. Signal. Process. Control 72 (Feb. 2022) 103357, https://doi.org/10.1016/J.BSPC.2021.103357.
[23] Z. Chao, X. Duan, S. Jia, X. Guo, H. Liu, F. Jia, Medical image fusion via discrete stationary wavelet transform and an enhanced radial basis function neural network, Appl. Soft. Comput. 118 (Mar. 2022) 108542, https://doi.org/10.1016/J.ASOC.2022.108542.
[24] H. Xu, J. Ma, EMFusion: an unsupervised enhanced medical image fusion network, Inf. Fusion 76 (Dec. 2021) 177–186, https://doi.org/10.1016/J.INFFUS.2021.06.001.
[25] W. Tang, F. He, Y. Liu, Y. Duan, MATR: multimodal medical image fusion via multiscale adaptive transformer, IEEE Trans. Image Process. 31 (2022) 5134–5149, https://doi.org/10.1109/TIP.2022.3193288.

[26] J. Fu, W. Li, J. Du, Y. Huang, A multiscale residual pyramid attention network for medical image fusion, Biomed. Signal. Process. Control 66 (Apr. 2021) 102488, https://doi.org/10.1016/J.BSPC.2021.102488.
[27] B. Goyal, et al., Multi-modality image fusion for medical assistive technology management based on hybrid domain filtering, Expert. Syst. Appl. 209 (Dec. 2022) 118283, https://doi.org/10.1016/J.ESWA.2022.118283.
[28] P.H. Dinh, Medical image fusion based on enhanced three-layer image decomposition and Chameleon swarm algorithm, Biomed. Signal. Process. Control 84 (Jul. 2023) 104740, https://doi.org/10.1016/J.BSPC.2023.104740.
[29] G. Zhang, R. Nie, J. Cao, L. Chen, Y. Zhu, FDGNet: a pair feature difference guided network for multimodal medical image fusion, Biomed. Signal. Process. Control 81 (Mar. 2023) 104545, https://doi.org/10.1016/J.BSPC.2022.104545.
[30] W. Li, Y. Zhang, G. Wang, Y. Huang, R. Li, DFENet: a dual-branch feature enhanced network integrating transformers and convolutional feature learning for multimodal medical image fusion, Biomed. Signal. Process. Control 80 (Feb. 2023) 104402, https://doi.org/10.1016/J.BSPC.2022.104402.
[31] D.C. Lepcha, B. Goyal, A. Dogra, A. Alkhayyat, S.K. Shah, V. Kukreja, A robust medical image fusion based on synthetic focusing degree criterion and special kernel set for clinical diagnosis, J. Comput. Sci. 20 (4) (Feb. 2024) 389–399, https://doi.org/10.3844/JCSSP.2024.389.399.
[32] Y. Xie, L. Yu, C. Ding, CFIFUSION: dual-branch complementary feature injection network for medical image fusion, Int. J. ImAging Syst. Technol. 34 (4) (Jul. 2024) e23144, https://doi.org/10.1002/IMA.23144.
[33] Y. Song, et al., DESTRANS: a medical image fusion method based on transformer and improved DENSENET, Comput. Biol. Med. 174 (May 2024) 108463, https://doi.org/10.1016/J.COMPBIOMED.2024.108463.
[34] Y. Liu, X. Chen, H. Peng, Z. Wang, Multi-focus image fusion with a deep convolutional neural network, Inf. Fusion 36 (Jul. 2017) 191–207, https://doi.org/10.1016/J.INFFUS.2016.12.001.
[35] Y. Jia, et al., Caffe: convolutional architecture for fast feature embedding, in: Proceedings of the MM 2014 - Proceedings of the 2014 ACM Conference on Multimedia, Nov. 2014, pp. 675–678, https://doi.org/10.1145/2647868.2654889.
[36] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, Y. LeCun, OverFeat: integrated recognition, localization and detection using convolutional networks, in: Proceedings of the International Conference on Learning Representations, 2013.
[37] Y. Liu, X. Chen, J. Cheng, H. Peng, A medical image fusion method based on convolutional neural networks, in: Proceedings of the 20th International Conference on Information Fusion, Fusion 2017 - Proceedings, Aug. 2017, https://doi.org/10.23919/ICIF.2017.8009769.
[38] M. Iwanowski, "Edge-Aware Color Image Manipulation by Combination of Low-Pass Linear Filter and Morphological Processing of Its Residuals," Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 12334 LNCS, pp. 59–71, 2020, doi: 10.1007/978-3-030-59006-2_6.
[39] D.C. Lepcha, B. Goyal, A. Dogra, S.H. Wang, J.S. Chohan, Medical image enhancement strategy based on morphologically processing of residuals using a special kernel, Expert. Syst. (2022) e13207, https://doi.org/10.1111/EXSY.13207.
[40] P. Soille, Morphological image analysis, Morphol. Image Anal. (2004), https://doi.org/10.1007/978-3-662-05088-0.
[41] W. Wang and F. Chang, "A multi-focus image fusion method based on laplacian pyramid," 2011, doi: 10.4304/jcp.6.12.2559-2566.
[42] P.J. Burt, R.J. Kolczynski, Enhanced image capture through fusion, in: Proceedings of the 1993 IEEE 4th International Conference on Computer Vision, 1993, pp. 173–182, https://doi.org/10.1109/ICCV.1993.378222.
[43] "The Whole Brain Atlas." Accessed: Oct. 12, 2024. [Online]. Available: https://www.med.harvard.edu/aanlib/home.html.
[44] Haozhe Xiang, Han Zhang, Yu Cheng, Xiongwen Quan, and Wanwan Huang. "SMFUSION: semantic-preserving fusion of multimodal medical images for enhanced clinical diagnosis." Arxiv Preprint *arXiv:2505.12251* (2025).
[45] Peng Peng, Yaohua Luo, Multimodal medical image fusion using a progressive parallel strategy based on deep learning, Electronics. (Basel) 14 (11) (2025) 2266.
[46] Zhe Li, Haiwei Pan, Kejia Zhang, Yuhua Wang, and Fengming Yu. "Mambadfuse: a mamba-based dual-phase model for multi-modality image fusion." Arxiv Preprint *arXiv:2404.08406* (2024).
[47] A. Sholehkerdar, J. Tavakoli, Z. Liu, Theoretical analysis of TSALLIS entropy-based quality measure for weighted averaging image fusion, Inf. Fusion 58 (Jun. 2020) 69–81, https://doi.org/10.1016/J.INFFUS.2019.12.010.