IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 7 July 2022, accepted 10 August 2022, date of publication 22 August 2022, date of current version 29 August 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3200752

RESEARCH ARTICLE

Texture Aware Deep Feature Map Based Linear Weighted Medical Image Fusion

VIJAYARAJAN RAJANGAM[1], (Senior Member, IEEE), DHEERAJ KANDIKATTU[2], UTKARSH[2], MUKUL KUMAR[2], AND ALEX NOEL JOSEPH RAJ[3], (Member, IEEE)

[1]Centre for Healthcare Advancement, Innovation and Research, Vellore Institute of Technology, Chennai 600127, India
[2]School of Electronics Engineering, Vellore Institute of Technology, Chennai 600127, India
[3]Department of Electronic Engineering, College of Engineering, Shantou University, Shantou 515063, China

Corresponding author: Vijayarajan Rajangam (viraj2k@gmail.com)

This work was supported by the Vellore Institute of Technology, Chennai, India.

ABSTRACT Medical image analysis is a critical job for clinicians and radiologists to attain minute insights for proper diagnosis. The presence of complementary details of the region of interest (ROI) from multiple medical imaging modalities instigates the researchers to integrate or combine the pathological details for the ease of clinical diagnosis. In this paper, the objective is to obtain a comprehensive image that presents composite image details from the two multimodal images of the same ROI. The basic idea is to generate robust fusion weights in the form of individually weighted matrices that could potentially superintend the fused outcome from the input image matrices. The extraction of texture features comes into play with the employment of the fast gray level co-occurrence matrix-mean technique. The feature maps of the source images are derived from the convolution layers on which the texture analysis is done to evaluate a weight map. Linear weights-based spatial domain fusion is employed using the weight map. Post auditioning several relevant fusion strategies and baseline hyper-parameter tuning, the obtained sets of outputs are validated via objective analysis in terms of standard metrics and compared with other fusion methods.

INDEX TERMS Feature map, GLCM, medical image fusion, texture map, deep learning.

I. INTRODUCTION

Multimodal medical image fusion, being an auxiliary approach, assists doctors to diagnose smoothly by leveraging information enhancement from multiple imaging modalities. The objective of image fusion is to integrate details from different parent images of the ROI to derive a comprehensive image that provides composite visual details from the multimodal images [1], [2]. When compared with the parent images, the visual information contained in the fused image is found to be much more detailed. It has the capacity to enhance the amount of visual information which will reduce the redundancy of information present in two or more images. Image fusion is predominantly employed in medical image diagnosis, remote sensing, agriculture, surveillance, and navigation.

The pathological analysis of ROI for disease diagnosis is possible by inspecting multimodal medical images such as Computed Tomography (CT) image, Magnetic Resonance Imaging (MRI), X-ray, Positron Emission of Tomography (PET), and Single Photon Emission Computed Tomography (SPECT) [3]. The fusion of CT and MRI presents anatomical and functional information in the composite image that makes the diagnosis less laborious for the clinicians. Image fusion methods are classified into pixel level fusion, feature level fusion, and decision level fusion [4]. In the first category, we tend to process raw pixel values with the parent image details and optimally retain a good chunk of original information. The method of feature level fusion operates at the point, angle, edge, texture, and other features extracted from the source images. The decision level fusion is carried out on the information extracted via low and mid-level image processing. In decision level fusion, both redundancy and uncertain information can be reduced while retaining the useful information present in the source images to serve image analysis better. This paper focuses on obtaining a single image, which presents better information by fusing two multimodal medical images. The medical modality known as MRI reveals

The associate editor coordinating the review of this manuscript and approving it for publication was Rajeswari Sundararajan.

the functional abnormalities of organs/tissues, whereas CT exposes them on an anatomical level. Thus, for more detailing in one go, the proposed image fusion technique stands apart and can be performed in vivid variants.

The exploration of the robust capability of a Deep Learning (DL) network helps to extract informative features and data representation. DL has been leading the state-of-the-art results in several computer vision and image processing operations [5]. The standard fusion practices follow step-wise max fusion strategy to club individual fused feature maps at the end. On the other hand, we are going with the formation of individual textural matrices with the aid of fast Gray Level Co-occurrence Matrix (GLCM)-mean technique followed by the genesis of individual weight matrices containing the fusion weights [6]. The feature maps of MRI and CT are derived using two levels of convolution layers. The average of feature maps is obtained concerning individual modality and then the fast-GLCM is applied to extract the texture feature maps of MRI and CT. Upon applying specific criteria, a weight map is obtained from the two fast-GLCM feature maps. This weight map is used to carry out spatial domain fusion for the source images. The performance of the proposed fusion method is compared with other fusion methods using the standard fusion metrics.

This paper is organized into different sections explaining the vivid angles of the proposition, starting with the literature review. In the literature review, a study of multimodal image fusion strategies using deep learning is conducted. Three types of decision functions are presented in section 3. Upon dataset acquisition, the intuitive interpretation is backed by the implementation and the results are attested in Section 4. This is followed by a conclusion in section 5.

A. LITERATURE SURVEY

Fayez and Sabine *et al.* proposed a novel image fusion model which is based on the Visual Geometry Group (VGG)-19 and softmax operator [7]. The proposed fusion model uses the weighted fusion technique. VGG-19 is used to extract feature maps from CT and MRI images, which are then processed by the softmax operator to generate weights needed for weighted fusion.

The most primitive setup with which we commenced our fusion algorithm is the Zero Learning Medical Image Fusion (ZLMIF) technique [7]. As discussed above, the fundamental idea is to provide individual deep feature maps in terms of numeric vectors as potential inputs to the softmax operator, which then would convert them into a vector of probabilities. The normalized numerics can be employed as fusion weights for individual feature maps, respectively, which would give rise to probable weight maps followed by clubbing them all to attain a final fused image.

Zhang *et al.* published a method that revolves around the proposition of a general fusion framework for varied forms of datasets which include infrared and visible images, multifocus images, MRI/CT images of the brain and multi-exposure images [4]. They used different fusion strategies for each type of input dataset. For performing a comprehensive fusion of infrared, multifocus, and medical images, max fusion is employed whereas for the multi-exposure images, mean fusion is used.

Inspired by this framework which is solely based on transform-domain image fusion algorithms, we moved ahead with a convolutional neural network, which would consist of feature extraction module, feature fusion module, and image regeneration module [4].

Fu *et al.* proposed a fusion model that uses a rolling guidance filter and VGG-16 convolutional network [8]. The rolling guidance filter produces a base image and a detail image. The convolutional neural network (CNN) produces a perceptual image. MRI and CT images are given as input to a rolling guidance filter and CNN to produce altogether three pairs of images. Base images are fused by local energy maximum fusion rule, detail images by local variance max fusion rule and perceptual images using sum modified Laplacian maximum fusion rule. At last, all the three fused images are bundled to get the final fused output.

Nishant *et al.* presented an unsupervised CNN model for the fusion of high and low-frequency components of MRI-PET source image pairs by exploiting structural similarity index (SSIM) as the loss function during training [9]. The authors suggested an application of color coding to visualize the outcome upon respective quantification of each input image in terms of the partial derivatives of the fused image.

Zhang *et al.* proposed a medical image fusion model that is specifically based on DenseNet, which aspires for feature reuse by interconnecting the features over channels. This enables the algorithm to perform better than conventional models with fewer parameters and calculation costs [10]. Nasrin and Ahmad proposed a method using VGG19, a pretrained network, for the fusion of MRI and PET scans. The weights for the fusion were extracted from the features of pretrained CNN layers [11].

B. DECISION FUNCTIONS

Based on the analysis of existing fusion methods, it is decided to specifically focus on the fusion module of the architecture and hence, went on to scrutinize the three different fusion strategies. Their respective intuitions can be briefly illustrated as follows:

1) GLCM ENERGY- BASED DECISION FUNCTION

The features are extracted from the source images using two convolutional layers. The depth of the first convolution layer is 64. Hence, 64 feature maps are generated for the MRI and CT images. For each feature map, energy is evaluated. E_1^i and E_2^i are the energy of the feature maps, where 'i' changes from 1 to 64. The feature fusion is governed by the following criteria.

- **if** $E_1^i > E_2^i$
  Append CT feature map

else
Append MRI feature map
end

2) GLCM ENERGY AND CONTRAST-BASED DECISION FUNCTION

For each feature map, energy and contract values are evaluated. C_1^i and C_2^i are the contrast of the feature maps of CT and MRI source images respectively. where 'i' represents the number of feature maps changing from 1 to 64. Similarly E_1^i and E_2^i are the energy of the feature maps. The fusion strategy is presented as follows.

- Let a, b, c, and d be the variables initialized to zero. The count of the variables will be increased according to the stated criteria.
- **if** $C_1^i > C_2^i$
  a = a+1
  else $C_1^i < C_2^i$
  b = b+1
  end
  if $E_1^i > E_2^i$
  c = c+1
  else $E_1^i < E_2^i$
  d = d+1
  end
- **if** $(a - b) < (c - d)$
  if $E_1^i > E_2^i$
  Append CT feature map
  else
  Append MRI feature map
  end
  else if $C_1^i > C_2^i$
  Append CT feature map
  else
  Append MRI feature map
  end
  end

C. SSIM-BASED DECISION FUNCTION

SSIM is evaluated between the fused and ground truth images, thus returning a numeric oscillating in the range of 0 to 1. The score can be computed each time by taking a specific ground truth image as a reference image and one of the medical modalities in the form of a feature map of the same scene as the processed image [12]. Thus, the basic idea is to exploit this concept to accumulate robust local feature maps with ground truth images. The fusion decision function based on the SSIM score is stated below

- **if** $SSIM_1^i > SSIM_2^i$
  Append CT feature map
  else
  Append MRI feature map
  end

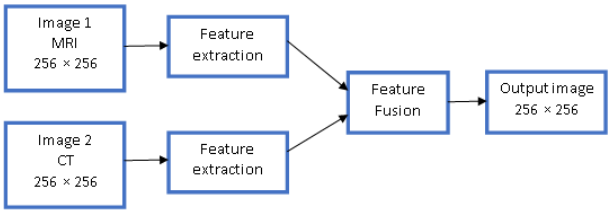

FIGURE 1. Block diagram of feature level fusion.

In the above-said decision functions for fusion, it is observed that the feature map selection is carried out using GLCM and SSIM. From the selected feature maps of the source medical images, the fused image is reconstructed using a reconstruction module. Lastly, we tried to explore the technicalities in the regeneration phase of the setup, and as a result, we moved to the FunFuseAn framework. The post feature extraction in which the fusion of high and low-frequency components of MRI-CT grayscale image pairs can be done separately by exploiting SSIM as the loss function during training [8]. The idea of separately handling the frequency components is executed to avoid loss as well as the mismatch of information contained in the fused outcome.

II. PROPOSED METHODOLOGY

In the DL-based fusion networks, the features are extracted by the convolution layers and fused using specific fusion criteria, as shown in Fig. 1. Then, the reconstruction module delivers the fused image from the fused features. In this paper, we proposed a Texture aware Deep Feature map-based linear weighted Image Fusion model (TDFIF). The model tends to work in two primed phases, namely the training phase and the fusion phase. The medical imaging modalities as potential inputs are primarily fed into the proposed network followed by the training procedure being done on it. In the fusion phase, a single pair of MRI and CT images is given as input to the trained model to get the fused output.

The basic idea is to generate robust fusion weights in the form of a weight matrix that could potentially superintend the fusion outcome upon encountering the input image matrices. To be precise, three specific decision rules can be decided based on the probable inequalities between the corresponding pixels of two texture matrices. It is for the generation of robust fusion weights in the form of individual weight matrices. Finally, upon respective encounter with the input image matrices followed by a linear weighted addition, the final fused image can be obtained which might be potentially vouched for its composite image details, unlike the source images.

A. FEATURE EXTRACTION MODULE

In this module, there are two convolution layers; the first convolution layer has a kernel of size 3 × 3 with one input as well as 64 output channels, whereas the second layer consists of a kernel of size 3 × 3 with both the input as well as output channels having frequency 64. Moreover, the padding and stride are fixed as unity to make sure that the

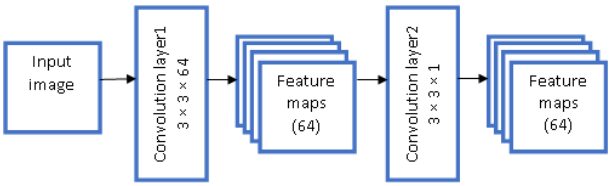

FIGURE 2. Block diagram of a feature extraction module.

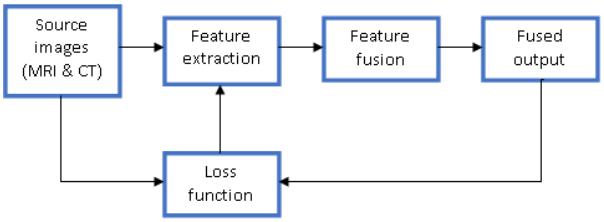

FIGURE 3. Block diagram of a training phase.

size of the feature map doesn't slip. Upon feature extraction of individual modalities, the obtained respective feature maps are summed up independently to produce two summed-up maps. The detailed block diagram of the feature extraction module is shown in Fig. 2.

B. LOSS FUNCTION

The subjective analysis of the fusion outcome depends on the local luminance, contrast, and structural properties of the image. That's the reason for considering SSIM [13] as a loss function that is solely based on human perception. The fusion method with loss function is presented in Fig. 3.

SSIM requires two images, a reference image and a processed image, and returns a numeric oscillating in the range of 0 to 1. The score can be computed each time by taking a specific ground truth image as a reference image and one of the medical modalities (in the form of a feature map) of the same ROI as a processed image. The mathematical interpretation for the same is mentioned below.

$$SSIM(I_1, I_2) = \frac{1}{N} \sum_{i_k, j_k \epsilon} [l(i_k, j_k)]^{\alpha} \cdot [c(i_k, j_k)]^{\beta} \cdot [s(i_k, j_k)]^{\gamma} \tag{1}$$

Here, I_1 and I_2 are the two potential inputs (medical imaging modalities), N is the number of local windows in I_1 and I_2, i_k and j_k are k^{th} local image contents of images, I_1 and I_2, respectively. We have assumed the values of α, β and γ as unity stating the clear message that all the three primed properties, namely; structural, contrast, and luminance are given the same weightage.

$$l(i_k, j_k) = \frac{2\mu_{i_k}\mu_{j_k} + C_l}{\mu_{i_k}^2 \mu_{j_k}^2 + C_l} \tag{2a}$$

$$c(i_k, j_k) = \frac{2\sigma_{i_k}\sigma_{j_k} + C_c}{\sigma_{i_k}^2 \sigma_{j_k}^2 + C_c} \tag{2b}$$

$$s(i_k, j_k) = \frac{2\sigma_{i_k j_k} + C_s}{\sigma_{i_k} + \sigma_{j_k} + C_s} \tag{2c}$$

The above equations $2a$, $2b$, $2c$ describe the luminance, contrast and structural properties of local image contents i_k and j_k. μ_{i_k} and μ_{j_k} are mean; σ_{i_k} and σ_{j_k} are the standard deviations of the image pixel values; σ_g is the standard deviation of the Gaussian filter, and $\sigma_{i_k j_k}$ is the correlation coefficient.

The pixel loss, $L2$ which tends to preserve better luminance, is experimented in addition to SSIM The steerable total loss function is expressed as:

$$L_{total} = \lambda * LSSIM + (1 - \lambda) * L2 \tag{3}$$

where,

$$LSSIM = (1 - SSIM(I_1, F)) + (1 - SSIM(I_2, F)) \tag{4a}$$

$$L2 = \|(F - I_1)\|^2 + \|(F - I_2)\|^2 \tag{4b}$$

where I_1 and I_2 are the two source images and F is the final fused image.

C. FEATURE FUSION MODULE

After feature extraction from the individual modalities, the obtained respective feature maps are summed up independently to produce two summed-up maps. The feature map sum is then divided by the number of feature maps to get a feature map average F_{avg}. Here, F_{sum} is the feature map sum, F_i is i^{th} feature map. The normalization of grey levels is done to adjust the numeric in the feature map sum to a common scale, without distorting differences in the range of values and hence, we attain the average of all the feature maps as depicted in the equations below.

$$F_{sum} = \sum_{i=1}^{64} F_i \tag{5}$$

$$F_{avg} = F_{sum}/64 \tag{6}$$

Then, texture feature extraction is employed with the use of fast GLCM-Mean technique assisted by pre-computed numerical value in the form of F_{avg}. In this way, the two independent textural matrices for individual modalities are obtained.

We have employed GLCM texture features based decision rule for the fusion of decided modalities [6]. Specifically, we used the fast GLCM-Mean technique to extract the second-order statistical texture features from the brain image. The texture in an image is all about how one level is co-occurring with the other. GLCM is a matrix containing all the probable frequencies of co-occurrences of each neighbouring level. The numerics in the GLCM signify the frequency of occurrences of a specific pair of pixels with a particular value concerning a specific spatial relationship. Preserving texture in the fused image obtained from the source modalities is alarmingly essential in the case of medical image fusion, as texture details help in classifying whether the image contains abnormalities or not. The equation for calculating GLCM mean for k^{th} local image content is

$$\mu_k = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} i \cdot (P_k(i, j)) \tag{7}$$

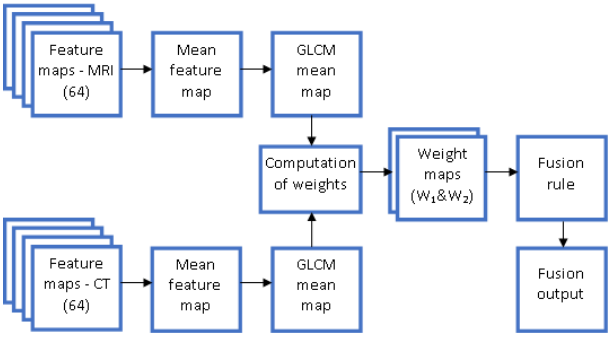

FIGURE 4. Block diagram of feature fusion module.

where, μ_k is the GLCM mean of k^{th} local image content, P is the GLCM matrix of k^{th} image, and i being the reference pixel value. Here, $P_k(i,j)$ represents the probability of pixel value i and j occurring side by side in the k^{th} local image. A new texture feature map is formed by assigning the numerical value μ_k at the center of the local image window in the form of $T_A : MRI$ and $T_B : CT$.

The fused image is obtained through the weights that are derived from the three fusion rules based on the probable inequalities between the corresponding pixels of two textural matrices. It is done for the generation of robust fusion weights in the form of individual weight matrices.

$$W_A(i,j) = \begin{cases} 0 & \text{if } T_A(i,j) < T_B(i,j) \\ 0.5 & \text{if } T_A(i,j) = T_B(i,j) \\ 1 & \text{if } T_A(i,j) > T_B(i,j) \end{cases}$$

$$W_B(i,j) = \begin{cases} 0 & \text{if } T_A(i,j) > T_B(i,j) \\ 0.5 & \text{if } T_A(i,j) = T_B(i,j) \\ 1 & \text{if } T_A(i,j) < T_B(i,j) \end{cases}$$

where W_A and W_B are weight maps of MRI and CT respectively. T_A and T_B are textural matrices of MRI and CT respectively.

Finally, upon respective encounters with I_1 and I_2, the source image matrices, followed by linear addition, the final fused image matrix is attained, as mentioned in equation 8, shown in Fig. 4 & Fig. 5. This output image seems to be a potential candidate of possessing comprehensive richness for pathological analysis.

$$F_{fused} = I_1 \times W_A + I_2 \times W_B \tag{8}$$

III. RESULTS AND ANALYSIS

The performance of the proposed method is validated by the set of images and analyzed with other fusion methods. Dense shift invariant transform (DSIFT), sparse representation (SR) fusion, ZLMIF, image fusion framework based on CNN (IFCNN) FunFuseAn, and VGG19 [7] are the methods used for comparison. The first experiment among the three experiments is about analyzing the fusion metrics for the four image pairs in the dataset. The source image pairs and fusion outputs are presented in Fig. 6, 7, 8 & 9.

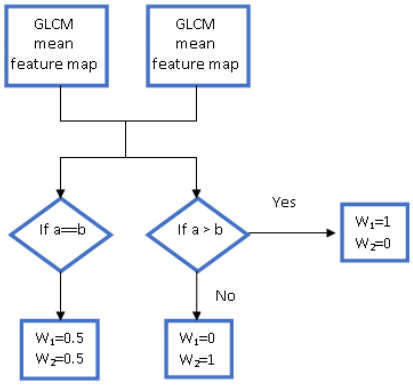

FIGURE 5. Weight Calculation using GLCM feature maps of MRI and CT. Where a and b are the mean values in the feature map according to the pixel coordinates.

Algorithm 1 Fusion of Extracted Feature Map Average F_A and F_B

Input Extracted mean feature maps F_A and F_B.

Steps

1) Extract GLCM mean feature map from F_A and F_B. P_A and P_B are GLCM matrices of F_A and F_B respectively.

$$T_{A_k} = \mu_{A_k} = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} i \cdot (P_{A_k}(i,j))$$

$$T_{B_k} = \mu_{B_k} = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} i \cdot (P_{B_k}(i,j))$$

2) Using the SSIM- based decision functions, generate weight maps W_A and W_B

$$W_A(i,j) = \begin{cases} 0 & \text{if } T_A(i,j) < T_B(i,j) \\ 0.5 & \text{if } T_A(i,j) = T_B(i,j) \\ 1 & \text{if } T_A(i,j) > T_B(i,j) \end{cases}$$

$$W_B(i,j) = \begin{cases} 0 & \text{if } T_A(i,j) > T_B(i,j) \\ 0.5 & \text{if } T_A(i,j) = T_B(i,j) \\ 1 & \text{if } T_A(i,j) < T_B(i,j) \end{cases}$$

3) Generate fused image by multiplying W_A and W_B with A and B and then adding the products, where A and B are MRI and CT images respectively.

$$F_{fused} = A \times W_A + B \times W_B$$

The metrics used for performance analysis are quality metric (Q_{mi}) [6], [14], and feature mutual information (FMI)-pixel [6], [14]. The second experiment is about edge preservation analysis using detect correct similarity (DCS) [14] metric, contrast based metric based on local similarity (Q_Y), Contrast based quality metric(Q_{cb}) [15], and SSIM [9]. The proposed method is tested on all the image pairs and the mean values of the standard metrics are evaluated in the third experiment.

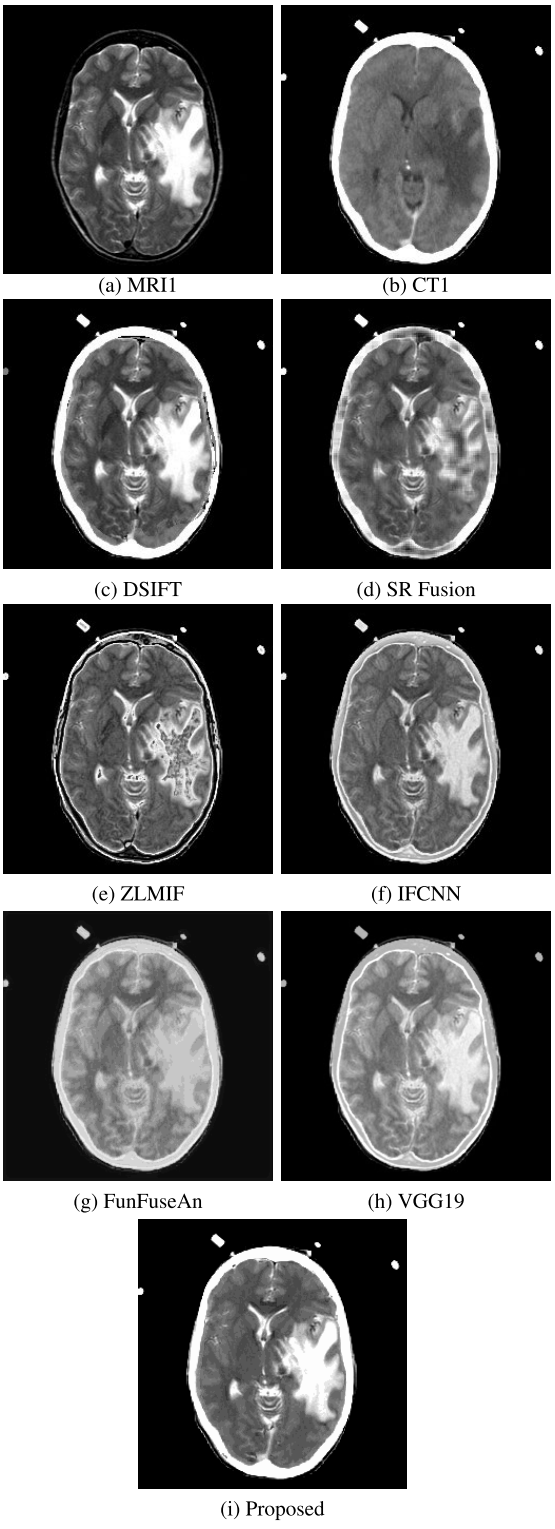

FIGURE 6. 1st set of output for the fusion methods.

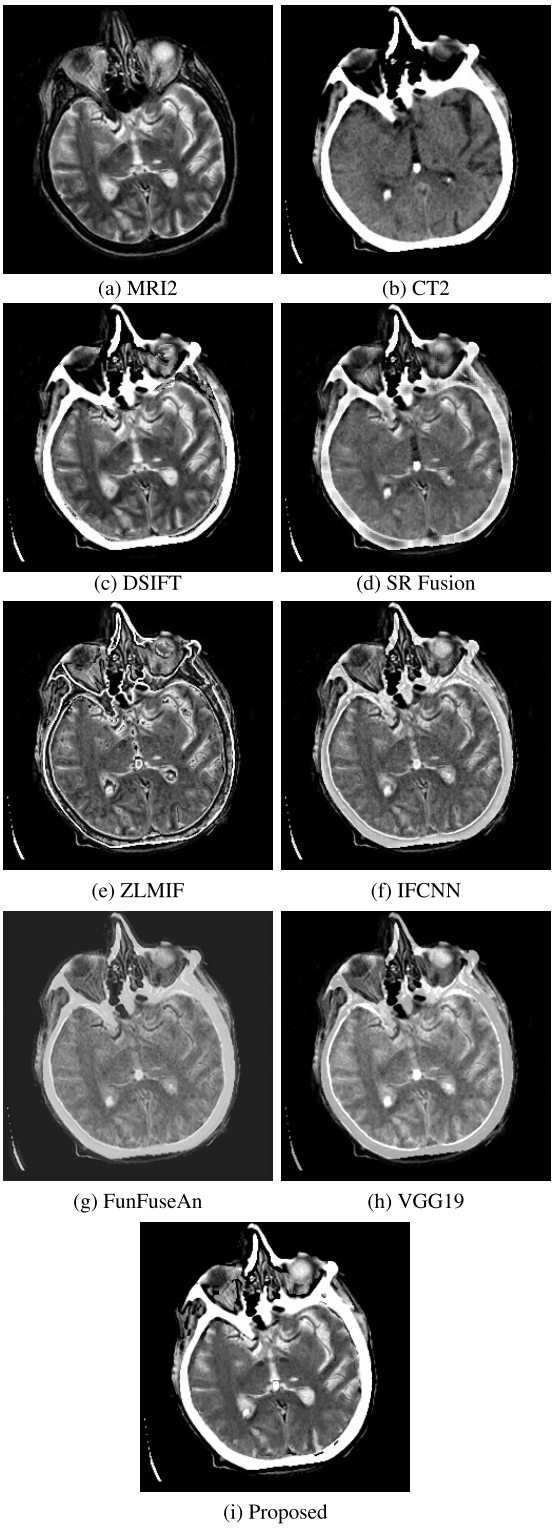

FIGURE 7. 2nd set of output for the fusion methods.

A. DATA ACQUISITION

A total of 268 pairs of MRI-CT human brain images are acquired. The split-up is random and made in such a way that 204 pairs are devoted to the training phase and the remaining 64 image pairs are employed for checking the robustness of the proposed fusion model. The Whole Brain Atlas, an open resource for central nervous system imaging, has made the MRI-CT images public for research purpose. Axial acquisition plane is used to prepare the dataset, and the acquisition Type is two-dimensional. MRI-T1, MRI-T2, and MRI-PD are

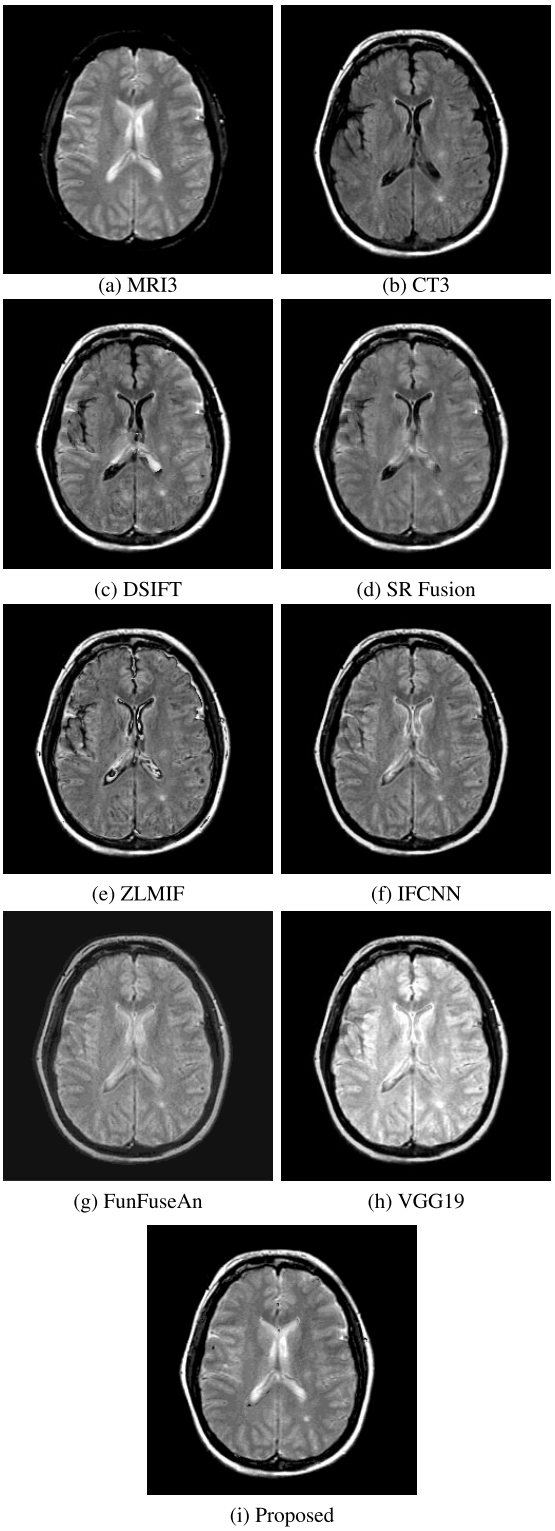

FIGURE 8. 3rd set of output for the fusion methods.

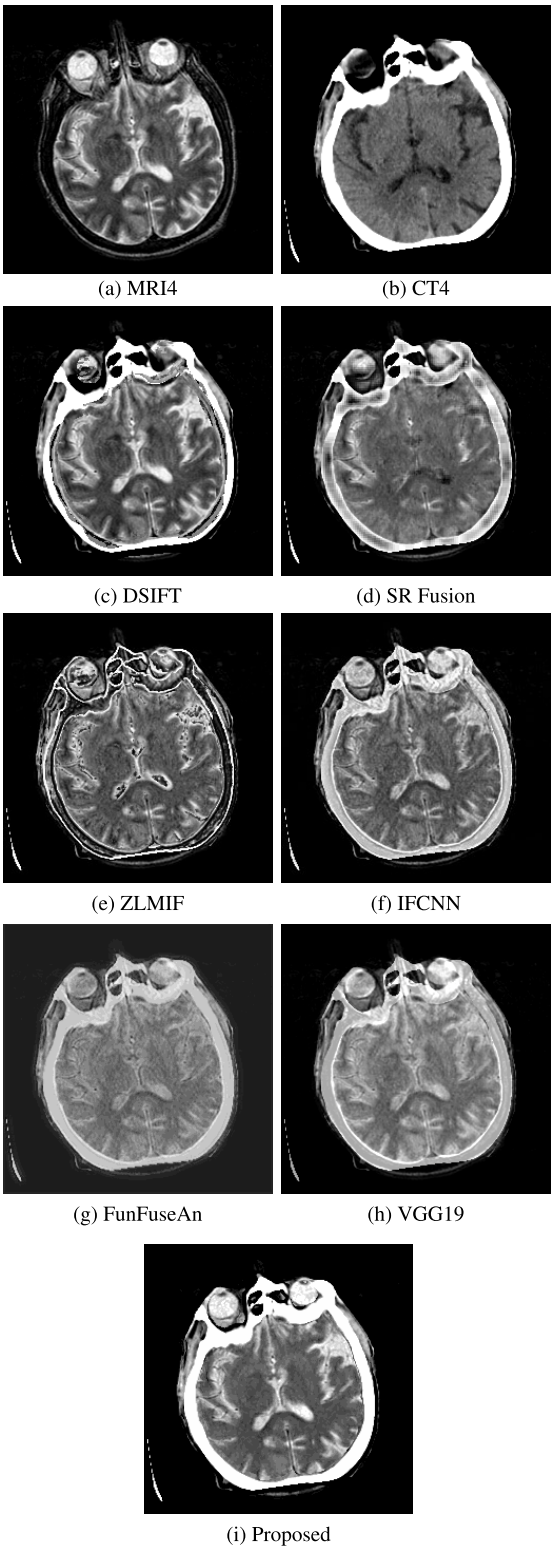

FIGURE 9. Obtained 4th set of output for the fusion methods.

the MRI sequences that are solely considered for the training purpose.

The images are registered brain images of different modalities. In some cases, the multimodal brain images are offered with fused ground truth images. The complete fusion dataset for multimodal images is not available, and hence the registered multimodal brain images are selectively taken with appropriate preregistration. The 268 source image pairs are obtained similarly. The fusion dataset derived from the website is used for training and testing. The ground truth images

are derived from the well-known fusion strategies and subsequently used for training and testing.

B. PERFORMANCE METRICS

The performance evaluation metrics adhere to the category of non-reference-based metrics. Q_{mi} and FMI-pixel deliver the metric score based on mutual statistical information. DCS metric tends to extract the edge similarity between parent modalities and the fused outcome. SSIM and Q_{cb} are quality assessment metrics based on Human Visual System (HVS), which employs structural similarity to compute the metric score. We have chosen three deep learning-based models, namely, ZLMIF, IFCNN, and FunfuseAn, for a relative assessment report concerning our proposed framework.

1) QUALITY METRIC

Q_{mi} is a quality measurement metric that evaluates the amount of information transferred between the source images and the fused image. $H(I_1)$, $H(I_2)$ and $H(I_F)$ are the marginal entropies of source images I_1, I_2, and fused image I_F respectively. $H(I_1, I_F)$ and $H(I_2, I_F)$ are the joint entropies.

$$Q_{mi} = 2 \times [\frac{MI(I_1, I_F)}{H(I_1) + H(I_F)} + \frac{MI(I_2, I_F)}{H(I_2) + H(I_F)}] \quad (9)$$

2) DETECT CORRECT SIMILARITY

It is the ratio of edge pixels present in both I_1 & I_2 and edge pixels present in I_1 but not in I_2. DCS reveals the similarity between two images based on edge pixels. Higher the value of DCS, better the similarity between two images [16].

$$DCS = \frac{Edge\ pixels\ present\ in\ both\ I_1\ and\ I_2}{Edge\ pixels\ present\ in\ I_1\ but\ not\ in\ I_2} \quad (10)$$

3) STRUCTURAL SIMILARITY INDEX

SSIM is a quality evaluation metric that considers contrast, variance and luminance to measure the structural similarity between images [17]. SSIM takes the values ranging from 0 to 1. Values of SSIM close to 0 reveal less similarity whereas values close to 1 reveal the high similarity between the images.

4) QUALITY METRIC BASED ON LOCAL SIMILARITY

Q_Y is a metric that employs local structural similarity between source images as a measure. The local structural similarities of a window w are calculated which are SSIM(x, y_w), SSIM(x, f_w) and SSIM(y, f_w) where x and y are source images and f is fused image.

$$Q(I_1^w, I_2^w, I_F^w) = \begin{cases} \lambda(w) \times SSIM(I_1, I_F^w) \\ +(1 - \lambda(w))\,\mathrm{SSIM}(I_2, I_F^w) \\ \quad \text{for } \mathrm{SSIM}(I_1, I_2^w) \geqslant 0.75 \\ \max\{\mathrm{SSIM}(I_1, I_F^w), \mathrm{SSIM}(I_2, I_F^w\} \\ \text{for } \mathrm{SSIM}(I_1, I_2^w) < 0.75 \end{cases}$$

$$\text{where, } \lambda(w) = \frac{s(I_1^w)}{s(I_1^w) + s(I_2^w)} \text{ is the local weight}$$

$s(I_1^w)$ and $s(I_2^w)$ are local variances of window w.

5) FEATURE MUTUAL INFORMATION

FMI_Pixel is mutual information-based metric that calculates the mutual information and entropies regionally.

$$I(x, y) = \sum_{xy} p(x, y) \frac{p(x, y)}{p(x) \cdot p(y)}$$

$$FMI_F^{AB} = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{I_i(CT; F)}{H_i(CT) + H_i(F)} + \frac{I_i(MRI; F)}{H_i(MRI) + H_i(F)} \right)$$

where $H_i(CT)$, $H_i(MRI)$, and $H_i(F)$ are the entropies evaluated locally of the source images CT,MRI, and fused image F respectively. $p(x, y)$ is the joint probability distribution of random variables x and y. $p(x)$ and $p(y)$ are probability distribution functions of random variables x and y respectively. I_i is mutual information. n is the number of local regions.

6) CONTRAST- BASED QUALITY METRIC

Q_{cb} employs the major features in the human visual system model which is a perceptual quality measure. It uses the contrast sensitivity function to describe human sensitivity to contrast.

$$Q_{AF}(x, y) = \begin{cases} \frac{C'_A(x, y)}{C'_F(x, y)} & \text{if } C'_A(x, y) < C'_F \\ \frac{C'_F(x, y)}{C'_A(x, y)} & \text{otherwise.} \end{cases}$$

$$\lambda_A(x, y) = \frac{C_A^2(x, y)}{C_A^2(x, y) + C_B^2(x, y)}.$$

$$Q_{cb}(x, y) = \lambda_A(x, y) Q_{AF}(x, y) + \lambda_B(x, y) Q_{BF}(x, y)$$

where C_A, C_B, and C_F are contrast maps of source images A,B, and fused image F respectively.

C. FUSION PERFORMANCE ANALYSIS

The fusion methods are employed on the four sets of source image pairs and the metrics are presented below.

1) MUTUAL INFORMATION-BASED METRICS

It is observed from the metrics that the TDFIF delivers good information transfer from the source images to the fused image. This could be observed by analyzing Q_{mi}. The TDFIF delivers superior results compared to all other methods. This metric is evaluated considering a complete source and fused images. But, $FMI - pixel$ is the metric evaluated based on the mutual information of local regions. Due to the contributions of the few local regions, the average information is high for the other methods. The performance of the proposed method for all the image pairs in the dataset is presented in Table 1 & 2.

D. QUALITATIVE PERFORMANCE ASSESSMENT

The qualitative performance analysis of TDFIF is carried out using edge preservation, contrast, variance, and the structural similarity between the fused image and source images.

TABLE 1. Performance analysis: Q_{mi} for the four sets of source images.

Methods	Set1	Set2	Set3	Set4
DSIFT	0.9738	0.7346	0.8233	0.7815
SR Fusion	0.8562	0.6722	0.8209	0.6949
ZL	0.8285	0.6172	0.7776	0.6605
IFCNN	0.8022	0.6046	0.7370	0.6527
FunFuseAn	0.9552	0.7226	0.8090	0.7630
VGG19	0.8649	0.7334	0.6878	0.6727
TDFIF	1.0914	0.8818	1.0705	0.9486

TABLE 2. Performance analysis: *FMI – pixel* for the four sets of source images.

Methods	Set1	Set2	Set3	Set4
DSIFT	0.8993	0.7007	0.8940	0.8385
SR Fusion	0.8950	0.7128	0.8935	0.8788
ZL	0.8642	0.7596	0.8874	0.8128
IFCNN	0.8913	0.7793	0.8877	0.8606
FunFuseAn	0.8162	0.7025	0.8837	0.8601
VGG19	0.9248	0.8994	0.9244	0.9056
TDFIF	0.8250	0.7118	0.9034	0.8773

TABLE 3. Performance analysis: DCS for the four sets of source images.

Methods	Set1	Set2	Set3	Set4
DSIFT	6.7422	4.6513	5.8470	4.9865
SR Fusion	3.2540	2.4468	3.2147	2.4537
ZL	2.7826	1.8097	2.6494	1.8009
IFCNN	2.9459	2.0509	2.8673	2.1264
FunFuseAn	2.7865	2.1418	2.7695	2.1283
VGG19	3.3050	2.0217	2.8193	2.1798
TDFIF	3.4908	2.3998	3.0862	3.0049

TABLE 4. Performance analysis: Q_Y for the four sets of source images.

Methods	Set1	Set2	Set3	Set4
DSIFT	0.4621	0.1736	0.0023	0.1434
SR Fusion	0.0832	0.0309	0.0815	0.1905
ZL	0.1369	0.1812	0.0212	0.0469
IFCNN	0.1326	0.2109	0.0458	0.1314
FunFuseAn	0.3948	0.3323	0.0190	0.0355
VGG19	0.2598	0.1777	0.2989	0.1711
TDFIF	0.4918	0.4411	0.0348	0.0793

TABLE 5. Performance analysis: Q_{cb} for the four sets of source images.

Methods	Set1	Set2	Set3	Set4
DSIFT	0.7721	0.5641	0.7388	0.6498
SR Fusion	0.7255	0.5441	0.7172	0.5617
ZL	0.7045	0.5223	0.7131	0.5694
IFCNN	0.7213	0.5461	0.7220	0.5934
FunFuseAn	0.5560	0.4578	0.2836	0.2254
VGG19	0.6657	0.5453	0.6869	0.5469
TDFIF	0.6500	0.5319	0.7290	0.6124

TABLE 6. Performance analysis: SSIM for the four sets of source images.

Methods	Set1	Set2	Set3	Set4
DSIFT	0.6863	0.4245	0.5416	0.5009
SR Fusion	0.6986	0.4168	0.5465	0.5034
ZL	0.7314	0.4311	0.5642	0.5059
IFCNN	0.7931	0.4344	0.5883	0.5395
FunFuseAn	0.5372	0.4253	0.6656	0.5877
VGG19	0.8217	0.7386	0.7986	0.7469
TDFIF	0.5321	0.4922	0.7124	0.6343

1) EDGE PRESERVATION ANALYSIS

The edge preservation capability of the proposed method is analyzed by DCS using two edge operators. The edge similarity is evaluated and the mean value is observed for analysis [16]. The DCS metric for four sets of images is presented in Table 3. It could be observed from the tabulated values that the proposed method preserves edges better than deep learning-based fusion methods. The other two methods, DSIFT and SR fusion, perform better than TDFIF.

2) CONTRAST AND VARIANCE BASED ANALYSIS

Q_Y is the local structural similarity measure using SSIM. The proposed method delivers good local similarity for the two sets of image and performs moderately well for the other two image pairs as presented in Table 4. Q_{cb} is the contrast sensitivity-based metric in which DSIFT tops the performance metrics. Whereas, the proposed method performs moderately well among the DL-based methods. From Table 5, it could be observed that similar performance is reflected in DCS. SSIM is another qualitative metric that analyzes the structural similarity between the source and fused images considering contract, variance, and illumination. The evaluated values are presented in Table 6. The SSIM values of the proposed method are better compared to other fusion methods taken for analysis except for VGG19 based fusion method.

E. DEPICTION OF RELATIVE ASSESSMENT OF FUSION METHODS

The analysis of metrics among the DL-based methods leads to ranking the performance. The TDFIF method tops the ranking, as shown in Fig. 10, as the performance is good in Q_{mi}, Q_Y and $SSIM$. Compared to other DL-based methods, its performance is moderately good in other metrics. The output images of the proposed method are subjectively superior compared to other fusion methods.

F. ANALYSIS OF SEGMENTED ROI AFTER FUSION

The impact of fusion in segmentation is analyzed by segmenting the source and fused images using fuzzy C-Means (FCM) clustering algorithm. The images are segmented into five clusters, then the segmented regions are presented in Fig. 11. It could be observed that the details present in the source images are fused and presented in the segmented region of the fused image. This would help in analyzing the ROI from the single fused image.

G. FEASIBILITY AND FUTURE SCOPE

To throw light on the future scope of the existing proposition, one could ponder over strengthening the primed modules, specifically, feature fusion as well as feature regeneration. The former could be strengthened upon the employment of unique fusion strategies which could potentially impact the fused outcome. Moreover, one could also look for vivid classifiers apart from the concept of decision mapping which could potentially assist the process efficiently. Now, for strengthening the latter module, one could go for the in-detail examination of the regeneration layers to interpret the happenings within the FC phase of the network. The feature maps' comprehensive visualization would do in this case.

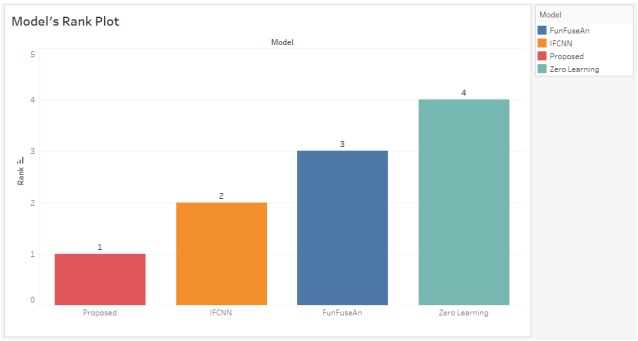

FIGURE 10. Relative Assessment of fusion models (Model Vs Relative Ranking) ***Lower the rank better the Model.***

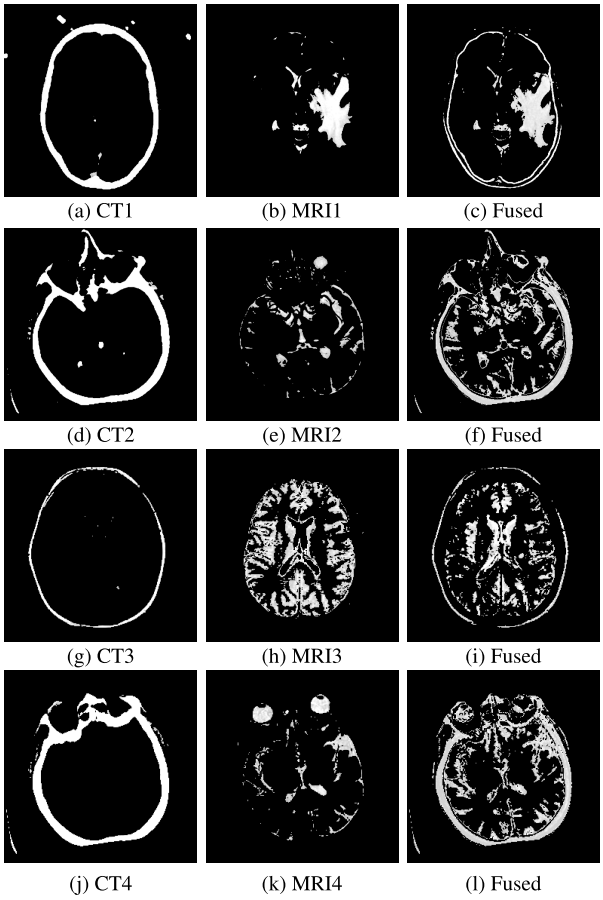

FIGURE 11. Impact of fusion in Segmentation (a-c) Set1, (d-f) Set2 (g-i) Set3 (j-l) Set4.

One could opt for making this module insignificant by vouching for weighted fusion as it eliminates the anticipated biases. Apart from strengthening the individual modules, one could try for baseline hyper-parameter tuning and robust training approaches (with the use of appropriate loss function, enhancing the frequency of the inputs via data augmentation, etc), multi-modal inputs such as PET, SPECT, etc apart from the standard medical imaging modalities, etc. Thus, the existing proposition can be potentially corroborated as robustly feasible as well as scalable from the technical stand-point as depicted in Tables 1 - 6.

TABLE 7. Mean metric values calculated over the entire dataset.

Metric	Mean Value
Q_y	0.5310
Q_{mi}	0.5198
Q_{cb}	0.5644
DCS	3.2828
FMI-pixel	0.7832
SSIM	0.5328

H. FUSION PERFORMANCE ON THE DATASET

The proposed TDFIF method is tested on all the image pairs in the dataset and the mean of metrics is presented in Table 7. It is observed that the qualitative and quantitative performances are moderately good for the proposed method.

IV. CONCLUSION

Multimodal fusion plays a vital role in combining complementary image details, thus eliminating the redundancy present among the multiple medical images of the same ROI. This paper evaluates statistical parameters from the GLCM matrices of the feature maps. The feature maps are derived from the source images using two sets of convolution layers. The decision function-based weights are derived from the GLCM matrix of the feature maps. It could be observed that the proposed TDFIF with SSIM-based decision function can deliver good fusion results subjectively. The performance is evaluated by the standard fusion metrics and also compared with other fusion algorithms. The objective evaluation is also good compared to other fusion methods.

REFERENCES

[1] R. Vijayarajan and S. Muttan, "Discrete wavelet transform based principal component averaging fusion for medical images," *AEU-Int. J. Electron. Commun.*, vol. 69, no. 6, pp. 896–902, Jun. 2015.
[2] R. Vijayarajan and S. Muttan, "Adaptive principal component analysis fusion schemes for multifocus and different optic condition images," *Int. J. Image Data Fusion*, vol. 7, no. 2, pp. 189–201, 2016, doi: 10.1080/19479832.2016.1149113.
[3] R. Vijayarajan and S. Muttan, "Iterative block level principal component averaging medical image fusion," *Optik*, vol. 125, no. 17, pp. 4751–4757, Sep. 2014, doi: 10.1016/j.ijleo.2014.04.068.
[4] Y. Zhang, Y. Liu, P. Sun, H. Yan, X. Zhao, and L. Zhang, "IFCNN: A general image fusion framework based on convolutional neural network," *Inf. Fusion*, vol. 54, pp. 99–118, Feb. 2020.
[5] Y. Huang, W. Li, and J. Du, "Anatomical-functional image fusion based on deep convolution neural networks in local Laplacian pyramid domain," *Int. J. Imag. Syst. Technol.*, vol. 31, no. 3, pp. 1246–1264, Sep. 2021.
[6] Z. Omar and T. Stathaki, "GLCM-based metric for image fusion assessment," in *Proc. 15th Int. Conf. Inf. Fusion*, Jul. 2012, pp. 376–381.
[7] F. Lahoud and S. Susstrunk, "Zero-learning fast medical image fusion," in *Proc. 22th Int. Conf. Inf. Fusion (FUSION)*, Jul. 2019, pp. 1–8.
[8] J. Fu, W. Li, A. Ouyang, and B. He, "Multimodal biomedical image fusion method via rolling guidance filter and deep convolutional neural networks," *Optik*, vol. 237, Jul. 2021, Art. no. 166726.
[9] N. Hoffmann, M. Oelschlägel, E. Koch, M. Kirsch, and S. Gumhold, "Structural similarity based anatomical and functional brain imaging fusion," in *Proc. Int. Workshop Multimodal Brain Image Anal.*, Aug. 2019, pp. 121–129.
[10] B. Zhang, C. Jiang, Y. Hu, and Z. Chen, "Medical image fusion based a densely connected convolutional networks," in *Proc. IEEE 5th Adv. Inf. Technol., Electron. Autom. Control Conf. (IAEAC)*, Mar. 2021, pp. 2164–2170, doi: 10.1109/IAEAC50856.2021.9390712.
[11] N. Amini and A. Mostaar, "Deep learning approach for fusion of magnetic resonance imaging-positron emission tomography image based on extract image features using pretrained network (VGG19)," *J. Med. Signals Sens.*, vol. 12, pp. 25–31, Jan. 2021.

[12] B. Ma, Y. Zhu, X. Yin, X. Ban, H. Huang, and M. Mukeshimana, "SESF-Fuse: An unsupervised deep model for multi-focus image fusion," *Neural Comput. Appl.*, vol. 33, no. 11, pp. 5793–5804, Sep. 2020.
[13] A. Ali Kiaei, H. Khotanlou, M. Abbasi, P. Kiaei, and Y. Bhrouzi, "An objective evaluation metric for image fusion based on Del operator," 2019, *arXiv:1905.07709*.
[14] G. Qu, D. Zhang, and P. Yan, "Information measure for performance of image fusion," *Electron. Lett.*, vol. 38, no. 7, pp. 313–315, 2002.
[15] M. B. A. Haghighat, A. Aghagolzadeh, and H. Seyedarabi, "A non-reference image fusion metric based on mutual information of image features," *Comput. Electr. Eng.*, vol. 37, no. 5, pp. 744–756, Sep. 2011.
[16] C. B. Gao, J. L. Zhou, J. R. Hu, and F. N. Lang, "Edge detection of colour image based on quaternion fractional differential," *IET Image Process.*, vol. 5, no. 3, pp. 261–272, Apr. 2011.
[17] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004, doi: 10.1109/TIP.2003.819861.

VIJAYARAJAN RAJANGAM (Senior Member, IEEE) received the Bachelor of Engineering degree in electronics and communication engineering from the University of Madras, in 1998, the master's degree in applied electronics from Madurai Kamarajar University, in 1999, and the Ph.D. degree in image fusion from Anna University, 2015. He is currently working as a Faculty Member with the Centre for Healthcare Advancement, Innovation, and Research, School of Electronics Engineering, VIT, Chennai. His research interests include computer vision, machine learning algorithms for signal and image processing, image fusion, bio-cryptography, emotion classification, and deep learning for image analysis. He is a fellow of IETE and a Life Member of ISTE. He is a Recognized Reviewer for Elsevier journals, IEEE Access, and IEEE Transactions on Instrumentation and Measurements.

DHEERAJ KANDIKATTU is currently pursuing the Bachelor of Technology degree in electronics and communication engineering with the Vellore Institute of Technology (VIT University), Chennai Campus. His research interests include deep learning and machine learning. He is also into competitive programming.

UTKARSH is currently pursuing the Bachelor of Technology degree in electronics and communication engineering with the Vellore Institute of Technology (VIT University), Chennai Campus. His research interests include data analytics, machine learning, and blockchain. More specifically, his research interests involve incorporating a data-driven approach in interpreting stuffs and attaining a standpoint, which is technically feasible and robust from business point of view.

MUKUL KUMAR is currently pursuing the Bachelor of Technology degree in electronics and communication engineering with the Vellore Institute of Technology (VIT University), Chennai Campus. His research interests include data science and software development. In particular, he has developed a strong interest in data analytics and machine learning techniques, which can be used to extract valuable insights from large data sets.

ALEX NOEL JOSEPH RAJ (Member, IEEE) received the B.E. degree in electrical engineering from Madras University, India, in 2001, the M.E. degree in applied electronics from Anna University, in 2005, and the Ph.D. degree in engineering from the University of Warwick, in 2009. From October 2009 to September 2011, he was a Design Engineer with Valeport Ltd., Totnes, U.K. From March 2013 to December 2016, he was a Professor with the Department of Embedded Technology, School of Electronics Engineering, Vellore Institute of Technology, Vellore, India. Since January 2017, he has been with the Department of Electronic Engineering, College of Engineering, Shantou University, China. His research interests include deep learning, signal and image processing, and FPGA implementations.