Application of U-Net Based Multiparameter Magnetic Resonance Image Fusion in the Diagnosis of Prostate Cancer

XUNAN HUANG[1], **BO ZHANG**[2], **XIAOLING ZHANG**[3], **MIN TANG**[3],
QIGUANG MIAO[1], **(Senior Member, IEEE), TANPING LI**[4], **AND GUANG JIA**[1]

[1]School of Computer Science and Technology, Xidian University, Xi'an 710071, China
[2]Tangdu Hospital, Air Force Military Medical University, Xi'an 710038, China
[3]MRI Room, Shaanxi Provincial People's Hospital, Xi'an 710068, China
[4]School of Physics and Optoelectronic Engineering, Xidian University, Xi'an 710071, China

Corresponding authors: Guang Jia (gjia@xidian.edu.cn) and Tanping Li (pli@xidian.edu.cn)

ABSTRACT Medical image fusion technology has been widely used in clinical practice by doctors to better understand lesion regions through the fusion of multiparametric medical images. This paper proposes an automated fusion method based on a U-Net. Through neural network learning, a weight distribution is generated based on the relationship between the image feature information and the multifocus training target. The MRI image pair of prostate cancer (axial T2-weighted and ADC map) is fused using a strategy based on local similarity and Gaussian pyramid transformation. Experimental results show that the fusion method can enhance the appearance of prostate cancer in terms of both visual quality and objective evaluation.

INDEX TERMS Medical image, image fusion, U-Net network, Laplacian pyramid, prostate cancer.

I. INTRODUCTION

With the rapid development of medical imaging technology, medical imaging has become an integral part of clinical disease diagnosis and treatment planning. Doctors may need to use multimodal medical imaging modalities, such as computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET), and single-photon emission computed tomography (SPECT), to comprehensively diagnose diseases. Due to its advantages of high soft-tissue contrast, high image resolution, and no radiation, MRI has been widely used in clinical diagnosis. Different tissue contrasts can be achieved by selecting pulse sequence scan parameters for MRI. It is difficult to detect prostate cancer without MRI when the prostate-specific antigen (PSA) level increases and/or there is a suspected abnormal digital rectal examination. Prostate cancer can be localized only through systematic gland-puncture sampling [1]. The gold standard for prostate cancer diagnosis is biopsy. The most widely used prostate biopsy technique is performed under the guidance of a trans-rectal ultrasound probe (TRUS). Although TRUS-guided prostate biopsy is the standard

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

method for the diagnosis of prostate cancer, its false-negative rate is high due to the inaccuracy of sampling [2]. MRI provides high-quality images for doctors to observe pelvic tissue, such as structures surrounding the prostate and the anatomical area around the gland [3]. Multiparametric MRI (mpMRI), which combines T2-weighted imaging (T2W) and functional pulse sequences (such as diffusion-weighted imaging (DWI) and dynamic contrast-enhanced (DCE) imaging), shows a better capability of detecting prostate cancer than regular MRI [4]. Biparametric MRI (bpMRI), including T2-weighted and apparent diffusion coefficient (ADC) maps generated from DWI, has been proposed as an accurate tool for localizing prostate cancer [5]. Internationally recognized consensus documents have been developed, such as prostate imaging reports and data system versions (PIRADS), and these have standardized the collection and reporting of prostate MRI results [3].

With the development of imaging technology, the role of mpMRI has been extended to tumor detection, disease monitoring and therapy follow-up. Some studies have confirmed the joint role of T2W MRI, DWI, and DCE-MRI in determining the invasiveness of prostate tumors [6]–[8]. Li *et al.* compared the performance of mpMRI and the Partin table and concluded that mpMRI achieves high accuracy

in prostate cancer staging [9]. There is an unmet need for intelligent mpMRI data processing for clinical applications. The traditional method used to diagnose prostate cancer is to manually browse different mpMRI images and switch between them. In a study by Steenbergen *et al.*, the tumor contours drawn by six observers in mpMRI (T2W imaging, DWI, and DCE-MRI) were compared using the data of 20 patients. The results showed that 69 satellite lesions were missed by all observers [10]. This variability among observers emphasizes the importance of the automatic fusion of prostate multimodal MRI images to enhance the appearance of satellite lesions to assist diagnosis. Prostate cancer may exhibit different signal intensity features in MRI images compared with normal prostate tissues. In a T2W image, for example, the prostate cancer region typically has lower SI than the surrounding tissue due to its higher cell density in T2W images (Figure 1). In an ADC map, low ADC value indicates limited diffusion, while high ADC value comes from tissue with relatively free diffusion [11]. It has been shown that incorporating ADC images with T2W images significantly improves the performance of prostate cancer detection compared with using T2W images alone [12]. Figure 1 shows that the relationship between the prostate and the surrounding tissue can be observed from the T2W image; however, the tumor boundary may be unclear due to benign diseases. The tumor boundary is more obvious in the ADC image; however, the surrounding tissue information may be lacking. The purpose of this experiment is to combine T2W images and ADC images through effective image fusion algorithms to help doctors more conveniently determine tumor location and size.
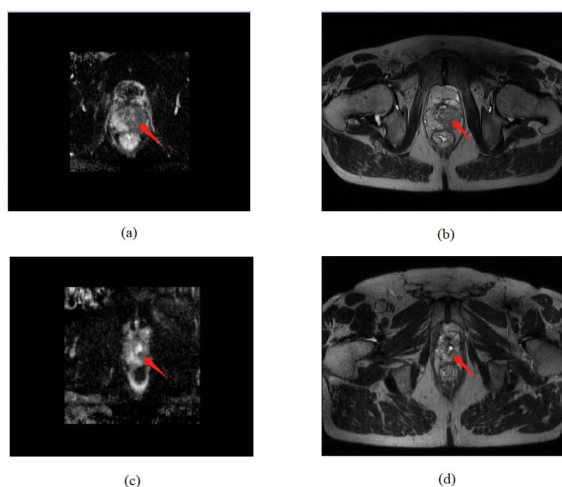
FIGURE 1. (a) The low-signal area of the ADC image of patient 4 shows the real cancer tissue. (b) The hypointense area in the T2W image alone is difficult to distinguish between benign prostatic hyperplasia and prostate cancer (red arrow). The appearance of patient 6's prostate cancer lesions (red arrows) in an axial ADC image (c) and axial T2W image (d).

The multiscale decomposition [13] method and sparse representation method [14] are commonly used image fusion methods based on the transform domain. Typical multiscale

decomposition methods include pyramid transformation [15] and wavelet transformation. The image obtained by the multiscale decomposition method contains much detailed information and exhibits high image contrast [16].

Medical image fusion is the process of merging and combining multimodal or multiparametric images to improve the applicability of medical images to clinical diagnosis and therapeutic monitoring [17]. There are two main types of image fusion methods: spatial domain-based and transform domain-based methods. The spatial domain-based methods are intuitive, simple, and performed directly on the image space but are sensitive to noise. The image fusion methods based on the transform domain are less sensitive to noise and more robust [16]. In recent years, image fusion research has focused on the transform domain, which has the advantages of good structure and distortion avoidance [14]. The Laplacian pyramid transform based on the Gaussian transform is widely used in medical image fusion [18]–[20].

With increased data generation speed and amount of imaging data, there is an urgent need for computerized methods to automatically extract clinical information from mpMRI data [21]. In recent years, convolutional neural networks (CNNs) have been used in diagnostic imaging of the urogenital tract [22], [23]. A CNN has been proved to be a powerful method for adaptive pattern recognition [24]. Image fusion based on deep learning has been widely used in multiparametric medical imaging [25], infrared and visible imaging [26], and remote sensing imaging [27]. The application of deep learning-based fusion of medical images is concentrated on multimodal image fusion (e.g., CT, PET, MRI) [16], [25]. Deep learning enables a better strategy for designing fusion rules than the traditional image fusion method—i.e., where the U-Net algorithm extracts relevant information from two source images and generates a weighted image using the end-to-end characteristics. Li *et al.* applied an artificial neural network at the pixel level for the fusion of multifocus images obtained from the same scene [28]. Liang *et al.* input the two source images into a deep learning network through MCFNet [16] and output fused images. The training dataset and labels needed for the neural network are difficult to obtain, which is a major limitation of MCFNet in applications. Liu *et al.* developed a medical image fusion algorithm based on a CNN [29], which outputs a fusion weight vector through a trained CNN model. This algorithm averaged the overlapping patches to convert the weight vector into a weight map with the same size as the input image. Wang *et al.* proposed a multiparametric segmentation and fusion network by utilizing 3D context information [30]. The method averaged the softmax output of a single task to fuse the segmentation results from three orthogonal views.

The method automatically learns image features from a large amount of labeled data. U-Net is an efficient end-to-end neural network structure, which is famous for its "U" shape. Due to its excellent performance in medical image segmentation and its elegant architecture [31], [32], U-Net has quickly become the mainstream method for the segmentation
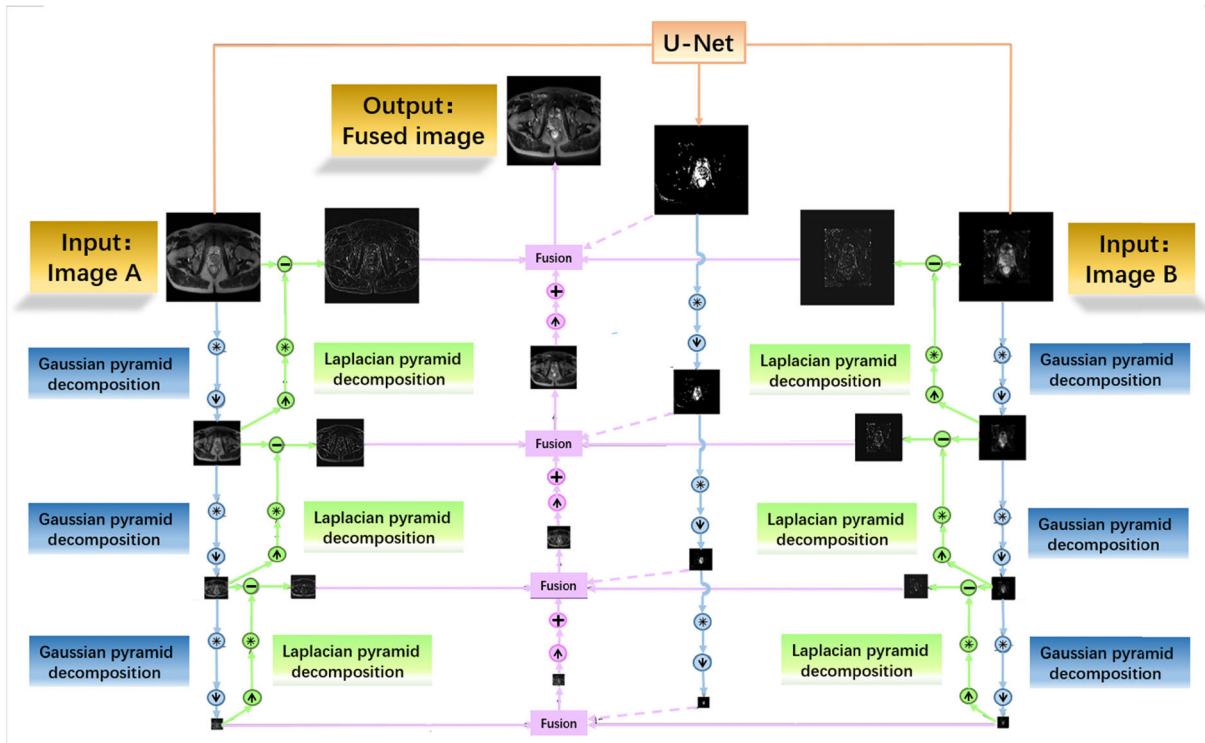
FIGURE 2. Flowchart of image fusion steps. The blue region represents Gaussian pyramid decomposition, and the green region represents Laplacian pyramid decomposition. The innovation of the algorithm is that the fusion weight map generated by U-Net, represented by the orange region, is brought into the fusion rules represented by the pink region, and the fusion image is then reconstructed by an inverse Laplacian pyramid.

of medical images and pathological images. For example, in the promise12 competition, U-Net performed well in segmenting the prostate [33], [34]. In recent years, CNNs and U-Net have been applied to medical image fusion [35].

Inspired by the above work, we propose an algorithm to fuse mpMRI images. This study combines the traditional Gaussian pyramid transform fusion framework in the transform domain with U-Net based on deep learning as a novel method for fusing medical images. The Laplacian pyramid has the advantage of enhancing salient image features (such as contrast and intensity), but it is sensitive to noise and cannot effectively process the edge information of images [18]. It is difficult to establish an appropriate fusion rule to process the semantic information of the image. U-shaped and other types of convolutional networks are also used to reduce noise and improve the quality of medical images [36], [37]. U-Net can effectively process and extract semantic image information [38]. The main contribution of this paper can be summarized as follows:

- We propose a method for fusion of mpMRI images. It is hoped that this application can aid doctors in finding prostate lesions in fused MRI images.
- This study applies the developed fusion method by combining the Laplacian pyramid transform and U-Net to fuse dual-parameter magnetic resonance images of the prostate.
- The experiment shows that our proposed method can provide fusion results with better tumor appearance in terms

of subjective quality assessment and objective evaluation than other methods used for comparison in this article.

The remainder of this article is organized as follows. The second section introduces the design scheme of the algorithm and the proposed method for data generation. The third section introduces training methods and objective evaluation indicators. At the end of the third section, the fusion results are introduced. A comparative analysis and future directions are discussed in the fourth section. The conclusion is presented in section 5.

II. THE PROPOSED ALGORITHM

This research proposes an mpMRI fusion framework based on the Laplacian pyramid transform, in which the fusion rules used are the weight maps generated by U-Net. Figure 2 shows the structure of the entire fusion framework. The fusion process is mainly composed of four steps, which are represented by different colors in the figure. The first step (orange arrows) indicates that U-Net is used to generate a weight map based on two source images. The details of this step are shown in Figure 2. (In the experiment, the Gaussian pyramid has a total of 8 layers, and only 4 layers are shown in the figure.) The second step is to perform Gaussian image decomposition on the weight map (indicated by blue arrows) and Laplacian pyramid decomposition on the two input source images (indicated by green crops). Gaussian pyramid decomposition is performed by first using the Gaussian window function and image to perform the convolution operation (indicated by ⊛),

TABLE 1. The networks structure of U-Net.

	Layer	Output size		Layer	Output size
Input	Image A&B	(256, 256, 2)		Convtranspose	(8, 8, 256)
	Convblock	(256, 256, 8)		Concatenate	(8, 8, 512)
	Maxpooling	(128, 128, 8)		Convblock	(8, 8, 256)
	Convblock	(128, 128, 16)		Convtranspose	(16, 16, 128)
	Maxpooling	(64, 64, 16)		Concatenate	(16, 16, 256)
	Convblock	(64, 64, 32)		Convblock	(16, 16, 128)
	Maxpooling	(32, 32, 32)		Convtranspose	(32, 32, 64)
	Convblock	(32, 32, 64)		Concatenate	(32, 32, 128)
	Maxpooling	(16, 16, 64)	Image Reconstruction	Convblock	(32, 32, 64)
	Convblock	(16, 16, 128)		Convtranspose	(64, 64, 32)
	Maxpooling	(8, 8, 128)		Concatenate	(64, 64, 64)
	Convblock	(8, 8, 256)		Convblock	(64, 64, 32)
	Maxpooling	(4, 4, 256)		Convtranspose	(128, 128, 16)
	Convblock	(4, 4, 512)		Concatenate	(128, 128, 32)
	Maxpooling	(2, 2, 512)		Convblock	(128, 128, 16)
	Convblock	(2, 2, 1024)		Convtranspose	(256, 256, 8)
Image Reconstruction	Convtranspose	(4, 4, 512)		Concatenate	(256, 256, 16)
	Concatenate	(4, 4, 1024)		Convblock	(256, 256, 8)
	Convblock	(4, 4, 512)	Output	Sigmoid	(256, 256, 1)

and interlacing and inter-column downsampling operations (indicated by ⤓) are then performed on the results obtained. Each time Gaussian gold tower decomposition is performed, the image size decreases 4-fold. The Gaussian pyramid consists of images G_0, G_1,…, G_7. Taking G_0 (i.e., the bottom of the Gaussian pyramid) as the source image, the other layer image G_k is the result of Gaussian pyramid decomposition on the k-1 layer image G_{k-1}. The two operations (Gaussian blur and downsampling) in the Gaussian pyramid construction process will lose part of the high-frequency details of the image. The pyramid is used to describe this lost high-frequency information. The process of constructing the Laplacian tower requires an upsampling operation (represented by ⤒), and G_k increases 4-fold to restore it to the same size as G_{k-1}. The same Gaussian kernel w is then used in downsampling to perform the convolution operation on G_k to obtain G_k', and the difference between G_{k-1} and G_k ' (represented by ⊖) is assessed to obtain L_k.

In the third step, the purple arrow is used to indicate that the local energy similarity is calculated for each layer of the image obtained by the pyramid decomposition, the weights in the weight map obtained in the first step are then brought into the calculation according to the similarity, and the results of each image layer are synthesized. Then, through the inverse Laplacian transform, the composite image of each layer is iteratively converted into the final fused image. The specific details and calculation formula are introduced in Part II.B of the article.

A. NETWORK STRUCTURE

U-Net was proposed by Ronneberger *et al.* and exhibits a strong ability to represent and extract features [39]. The network is used for pixel-level regression and end-to-end image segmentation with robustness and generalizability. We used U-Net to generate a graph of fusion weights in this study. Synthesized multifocus image datasets are used to train

U-Net to generate a fusion weight map, as shown in Figure 3 (256 × 256).

The multifocus image pairs are superimposed as the input of the neural network for feature extraction and fusion. The network automatically extracts relevant image information from the input image pair and outputs a weighted image with the same size as the input image. Two images with a size of 256 × 256 are merged into dual-channel input data. Seven upsampling layers and seven downsampling layers are used in the U-shaped network. As shown in Figure 3, the left side is a shrinking path composed of repeated convolution blocks and maximum pooling layers. The ReLU activation function is used after each convolutional layer. During the downsampling process, the number of channels gradually increases.

The upsampling process of the gradual expansion of the image size is on the right. Each layer is skip connected with the feature maps of the corresponding shrinking layer, and the deconvolution is then gradually expanded. In the upsampling process, the generated high-dimensional features are merged with the low-level features to preserve the details of the image to the greatest extent.

In the structure, the light-yellow box represents the multichannel feature map after convolution, and the blue box represents the copied feature map. The number of channels is displayed at the bottom of the box. The red box represents the feature map generated after the 2 × 2 maximum pooling downsampling operation with strides of 2. The network uses the Dice coefficient as the cost function to increase the speed of the weight update, thereby effectively increasing the training speed. In the last layer, to obtain a single-channel weight map, we assign a sigmoid convolutional layer, which is represented by the purple box in the figure.

Due to the limitation of GTX 1060 GPU memory, the batch size was selected as 32 in this experiment. To select the appropriate learning rate for this batch size, we set three learning
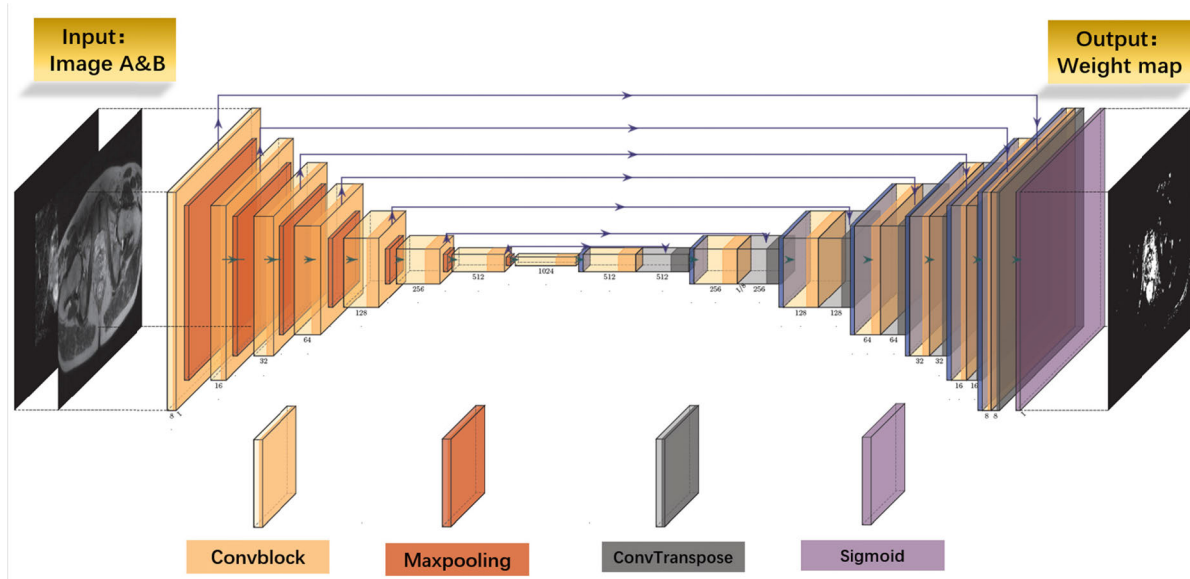
FIGURE 3. The architecture of the U-Net proposed in our work. The input of the network is a dual-channel grayscale image, and the output is a single-channel weighted image. The number of channels of each layer of convolution is identified below the corresponding convolution block. The arrow indicates skip connection.

rates of 0.01, 0.001 and 0.0001 for comparison. The results show that when the learning rate is large (0.01), the curve of target loss function fluctuates. When the learning rate is small, the training time is long. Therefore, we use 0.001 as the initial learning rate of the Adam optimizer for training. After U-Net training for 30 epochs, the Dice coefficient can reach more than 0.98. To balance the training effect and training time, we set the number of epochs to 30. The training process is performed within the Keras deep learning framework.

B. GENERATION OF A COLLECTION OF MULTIFOCUS IMAGES FOR TRAINING

Due to the lack of a ground truth for medical image fusion, medical images cannot be used as a training dataset. In this experiment, the contents of the ImageSets and Segmentation-Class folders in the PASCAL VOC 2012 dataset are used to synthesize multifocus images for training U-Net. There are 2913 original images in the JPEGImages folder, which are deemed to be the ground-truth images. The corresponding segmentation images in the SegmentationClass folder are used to define target and background areas within the original images. The multifocus image synthesis process includes four steps: (1) conversion of source images into grayscale images, (2) Gaussian filtering of grayscale images for the generation of blurred images, (3) generation of focused regions (target and background) by using binarization of the segmented images, and (4) synthesis of multifocus images. To mimic grayscale medical images, the color images in the JPEGImages folder are converted into grayscale images. For each grayscale image, we use a Gaussian filter (with its standard deviation set to 2 and window size set to 7) to generate blurred images. Next, we binarize the segmented image in

the SegmentationClass folder and define a mask. In the mask, the background pixels have a value of 0, and the target pixels have a value of 1. Finally, the grayscale image and the blurred image are superimposed onto one another according to the mask to generate focus image A, in which the value of the background-area pixel equals the value of the unblurred grayscale-image pixel, and the target-area pixel corresponds to the blurred-image pixel. Similarly, focus image B is generated by setting the value of the background-area pixel equal to the value of the blurred-image pixel, where the target-area pixel corresponds to the unblurred grayscale-image pixel.

The production of the training dataset by using the above steps is equivalent to generating two focus images (A and B) from an unblurred grayscale image. The mask separating the target from the background is deemed the ground-truth training target. The task of U-Net is to find pixels in the target area from the multifocus images (A and B) and generate a weight map, in which the pixels with a value of 1 represent the greatest possibility of being located in the target area, and the pixels with a value of 0 represent the greatest possibility of being in the background area.

C. IMAGE PYRAMID FUSION STEPS

The entire image fusion process is divided into four steps: generation of a weight map by U-Net, Gaussian transform and Laplacian transform of images, rule-based fusion, and image reconstruction.

Step 1: Calculation of the fusion weight map. Multi-parametric medical images A and B are combined into a dual-channel image and sent to the trained U-Net to obtain a weight map. The pixel values in the weight map represent the focus attributes at the pixel positions of the source images.

The closer the weight value is to 1, the greater the representation of the pixel intensity in A, and the closer the weight is to 0, the greater the representation of the pixel intensity in B. A weight of 0.5 indicates that images A and B have an identical weight at the pixel point. The gray values of A and B at the pixel point are averaged to generate the fusion result. The pixel value is used as a weight in step 3.

Step 2: Pyramid decomposition. The two source medical images are decomposed according to the Laplacian pyramid. Gaussian pyramid decomposition is performed on the weight map obtained in the first step. The maximum number of decomposition levels is determined according to the size of the medical image. Let the image size be H×W. The maximum possible total number of decomposition levels for each pyramid is $log_2 min(h, w)$. The image size in this experiment is 256; thus, the maximum number of decomposition layers is 8. In the Gaussian pyramid, the image size of the G_0 layer is 256 × 256. The image size of the G_1 layer is reduced 2-fold in length and width—i.e., 128 ×128—and so on. The image size of the G_7 layer is 2 ×2.

Step 3: Image fusion based on the local energy features and weight map. According to the fusion rules, image fusion is performed at each decomposition level.

In the formulaic description, L_k^A and L_k^B represent the Laplacian decomposition images corresponding to source images A and B in layer k of the Laplacian pyramid, respectively. At the highest pyramid level, the Gaussian decomposition images of images A and B, G_0^A and G_0^B, are calculated and combined. In the experiment presented in this paper, the threshold value of the similarity measure in Formulas 3 and 4 is set to 0.6, and the detailed explanation of the threshold selection is provided in section III.E.

(1). Calculate the local energy features of the Laplacian (or Gaussian for the final layer) images A and B. For each image, the local energy at (i, j) is the sum of the squared pixel values in the m × n window centered on the point.

$$\begin{cases} E_A = \sum_M \sum_N L_k^A(i + m, j + n)^2 \\ E_B = \sum_M \sum_N L_k^B(i + m, j + n)^2 \end{cases} \quad (1)$$

(2). Calculate the matching degree of two images $M_{AB}(i, j)$.

$$M_{AB}(i, j) = \frac{\left(\sum_M \sum_N L_k^A(i + m, j + n) L_k^B(i + m, j + n)\right)^2}{E_A \times E_B} \quad (2)$$

(3). If the matching degree at this point $M_{AB}(i, j) < thd$, select the pixel with the highest energy at this point.

$$L\{F\}(i, j) = \begin{cases} L_k^A(i, j), & if\ E_A > E_B \\ L_k^B(i, j), & if\ E_B > E_A \end{cases} \quad (3)$$

(4). If the matching degree at this point $M_{AB}(i, j) > thd$, the image is fused according to the fusion weight graph W

$$L\{F\}(i, j) = G_k^W(i, j) L_k^A(i, j) + (1-)_k^W(i, j)L_k^B(i, j) \quad (4)$$

Step 4: Laplacian pyramid reconstruction. The fused image obtained in the third step is reconstructed according to the Laplacian pyramid inverse transform.

Algorithm 1 Fusion

$$E_A = \sum_M \sum_N G_0^A(i + m.j + n)^2$$

$$E_B = \sum_M \sum_N G_0^B(i + m.j + n)^2$$

$$M_{AB}(i, j) = \frac{\left(\sum_M \sum_N G_0^A(i + m, j + n) G_0^B(i + m, j + n)\right)^2}{E_A \times E_B}$$

If $M_{AB}(i, j) > $ thd **then**
$$M_1(i, j) = 1$$
else then
$$M_1(i, j) = 0$$
If $E_A > E_B$ **then**
$M_2(i, j) = 1$
else then
$M_2(i, j) = 0$

$$Y_P(i, j) = M_1(i, j)(G_0^W(i, j) G_0^A(i, j)$$
$$+ (1 - G_0^W(i, j))G_0^B(i, j)$$
$$Y_Q(i, j) = (1 - M_1(i, j))(M_2(i, j) G_0^A(i, j)$$
$$+ (1 - M_2(i, j))G_0^B(i, j)$$
$$Y_0(i, j) = Y_P(i, j) + Y_Q(i, j)$$

for k=1 **to** nlev **do (nlev is the decomposition level of the Gaussian pyramid)**

$$E_A = \sum_M \sum_N L_k^A(i + m.j + n)^2$$

$$E_B = \sum_M \sum_N L_k^B(i + m.j + n)^2$$

$$M_{AB}(i, j) = \frac{\left(\sum_M \sum_N L_k^A(i + m, j + n) L_k^B(i + m, j + n)\right)^2}{E_A \times E_B}$$

If $M_{AB}(i, j) > $ thd **then**
$M_1(i, j) = 1$
else then
$$M_1(i, j) = 0$$
If $E_A > E_B$ **then**
$M_2(i, j) = 1$
else then
$M_2(i, j) = 0$

$$Y_P(i, j) = M_1(i, j)(G_k^W(i, j) L_k^A(i, j)$$
$$+ (1 - G_k^W(i, j))L_k^B(i, j)$$
$$Y_Q(i, j) = (1 - M_1(i, j))(M_2(i, j) L_k^A(i, j)$$
$$+ (1 - M_2(i, j))L_k^B(i, j)$$
$$Y_k(i, j) = Y_P(i, j) + Y_Q(i, j) Y_{k-1}(i, j)$$

end

III. EXPERIMENT AND RESULTS

In the training phase, the network is trained using synthetic multifocus images. Specifically, multifocus images A and B are combined into an H × W × 2 matrix pair as input and sent to U-Net. U-Net generates an H × W weight map and performs the corresponding ground-truth supervision.

A. MRI IMAGE

The MRI data were downloaded from I2CVB (http://i2cvb.github.io/#prostate-data). The dataset includes a total of 17 patients with prostate cancer confirmed by biopsy. Of these 17 patients, 12 had prostate cancer in the peripheral zone (PZ), 3 had prostate cancer in the central gland (CG), and 2 had infiltrating prostate cancer in both the PZ and CG areas. An experienced radiologist segmented the prostate and prostate cancer areas in the image.

All data were obtained using a 3-T MR scanner with a pelvic phased-array coil (Magnetom Trio and Skyla, Siemens Healthcare, Erlangen, Germany). The T2W image on the axial plane was obtained using a three-dimensional T2W fast spin-echo sequence (TR: 3600 ms, TE: 143 ms, ETL: 109, slice thickness: 1:25 mm). The nominal matrix and FOV of the 3D T2W fast spin-echo image were 320 mm×256 mm and 280 mm×240 mm, respectively. The single-shot spin-echo echo-planar imaging (EPI) technique was used to obtain DWI images. The diffusion-coded gradient was obtained by the pulse-gradient spin-echo technique, and the diffusion image was obtained in three orthogonal directions. Sequential sampling in k-space was used, with TE = 101 ms, TR = 4200 ms and bandwidth = 1180 Hz. Other parameters include the following: FOV, 240 mm; acquisition matrix size, 128 × 128; and slice thickness, 3.5 mm. ADC mapping was performed by the Siemens workstation directly from the original data of each pixel.

B. PARAMETER SETTING

The mean squared error loss function is used as the training target. For each convolutional layer, we initialize the weights using the variance scaling method and initialize the bias to 0. The optimizer uses adaptive moment estimation (Adam), and its basic learning rate is set to 0.001. We use a batch learning strategy to train the model with the parameter batch size set to 32. We train the model from scratch on 30 epochs of data. The performance evaluation during training uses Dice coefficients.

C. EXPERIMENTAL ENVIRONMENT

Our experiment was conducted on a PC with an Intel Core i5-6500 CPU at 3.20 GHz, 16 GB of RAM, and an NVIDIA GeForce GTX 1060 GPU with 6 GB of memory. In the Python 3.5 environment, U-Net model training and feature extraction experiments were performed using the TensorFlow framework to generate weight maps. Image pyramid decomposition and Laplacian inverse transformation

were implemented using MATLAB R2016a to complete the final fusion process.

D. EVALUATION METHODS

The evaluation methods of image fusion use five indexes: information entropy (EN), the similarity measure based on the gradient ($Q_{AB/F}$), structural similarity (MS_SSIM), the mutual information measure (MI), and feature-mutual information (FMI_w). For each index, the higher the value, the better the quality of the fused image.

EN is a measure of the amount of data contained in an image [40]. The EN of the discrete random variable i is defined as follows:

$$EN = -\sum_{i=1}^{L} h(i)\, log_2 h(i) \tag{5}$$

where L is the number of gray levels, and h(i) denotes the normalized histogram of the fused image. In this study, L is set to 256.

$Q_{AB/F}$ is a gradient-based measurement that evaluates fusion quality by measuring the spatial details of the source image and the fused image [41]. It is defined as follows:

$$
\begin{aligned}
&Q_{AB/F} \\
&= \frac{\sum_{n=1}^{N}\sum_{m=1}^{M} Q_{AF}(n,m)\omega_A(n,m) + Q_{BF}(n,m)\omega_B(n,m)}{\sum_{n=1}^{N}\sum_{m=1}^{M}(\omega_A(n,m) + \omega_B(n,m))}
\end{aligned}
\tag{6}
$$

where N and M are the length and width of the input images, respectively. $Q_{AF}(n,m)$ and $Q_{BF}(n,m)$ are edge information retention values, weighted by $\omega_A(n,m)$ and $\omega_B(n,m)$, respectively.

Structural similarity (SSIM) is a measure of the similarity between two images [42]. Its definition is as follows:

$$SSIM_{AB}^F = \frac{SSIM_{AF} + SSIM_{BF}}{2} \tag{7}$$

$$SSIM_{AF} = \frac{(2\mu_A\mu_F + C_1)(2\sigma_{AF} + C_2)}{(\mu_A^2 + \mu_F^2 + C_1)(\sigma_A^2 + 2\sigma_F^2 + C_2)} \tag{8}$$

MI is a measure of the total information transferred from the input images to the fused image and is denoted as follows [43]:

$$MI = \frac{MI_{AF} + MI_{BF}}{2} \tag{9}$$

$$MI_{AF} = \sum_m \sum_n p_{AF}(m,n) log_2 \frac{p_{AF}(m,n)}{p_A(m)p_F(n)} \tag{10}$$

FMI is a measure of MI between the features of the input image and the features of the fused image, with the following definition [40]:

$$FMI = \frac{FMI_{AF} + FMI_{BF}}{2} \tag{11}$$

$$FMI_{AF} = \frac{2}{L}\sum_{i=1}^{L} \frac{MI_{AF}}{EN_A + EN_F} \tag{12}$$

FMI_ w is defined as the value of FMI, which is calculated after the wavelet transform of the image is performed by using a sliding window. In this study, the default 3 × 3 sliding window is used.

E. LOCAL SIMILARITY

In the image pyramid fusion process, the third step involves the calculation of local similarity and the selection of the matching-degree threshold. Taking different thresholds, fusion results are checked by applying objective evaluation criteria (Table 2). It is concluded that the optimal threshold is between 0.4 and 0.6. In this experiment, 0.6 is selected for the image fusion process.

TABLE 2. Average score of the objective evaluation of various fusion thresholds.

Threshold	EN	FMI_w	MI	MS_SSIM	$Q_{AB/F}$
0.1	6.9844	0.4697	13.968	0.9748	0.7281
0.2	7.1383	0.5048	14.276	0.9762	0.7002
0.3	7.1584	0.5154	14.316	0.9781	0.6212
0.4	7.1585	0.5158	14.317	0.9785	0.6628
0.5	7.1553	0.5133	14.310	0.9797	0.634
0.6	7.1358	0.5055	14.271	0.9803	0.6782
0.7	7.1041	0.486	14.208	0.9806	0.7055
0.8	7.0583	0.478	14.116	0.9809	0.7066

F. RESULTS

In the experiment, we selected a total of 30 pairs of MRI images (axial T2W and ADC images) from 10 prostate cancer patients as experimental data (3 pairs per patient).

In the training stage, the synthesized multifocus image dataset is used to train the network. To expand the image data and increase the robustness of the network, for the MRI image and the corresponding label image in the training set, we randomly select several operations for the image from the following strategies: random rotation in the range of 0 to 10 degrees, from 0 to randomly move along the X and Y axes between 0.1 times the image size, and randomly flip the image horizontally or vertically, stretching 1 to 1.2 times randomly along the X and Y axes. A total of 30 epochs are trained, each requiring approximately 75 s. The 30 rounds of training take a total of 38 min. After 30 epochs of training, the Dice coefficient is 0.9826. This shows that U-Net can effectively segment the clearest parts from the multifocus gray image. Then, the MRI image is input into the trained U-Net to generate a weight map.

T2W images are sensitive but not specific to prostate cancer. It is difficult to distinguish prostate cancer from prostatic intraepithelial neoplasia, hemorrhage, changes after radiotherapy or prostatitis in T2W images. The ADC value of prostate cancer is lower than that of the surrounding healthy prostate tissue, showing a low signal in the ADC map. By fusing ADC and T2W images, the location of prostate cancer and the relationship between tumor and surrounding tissue can be identified, and the image specificity of prostate cancer can be improved. It is convenient for doctors to judge the tumor stage and determine the location of the tumor, which reduces the workload of doctors who read the map.

Our proposed algorithm and some comparison algorithms are used to perform image fusion on all test image pairs with an objective and subjective evaluation of the fusion results.

The compared algorithms include the weighted average algorithm (AVE) [46], domain multiparametric medical image fusion method (NSCT) [47], zero-learning fast medical image fusion method (ZLF) [48], and CNN-based method (CNN)[29].

Figures 4 and 5 show the fusion results from two pairs of images. The mpMRI images are T2W and ADC images, which are usually used for prostate tumor diagnosis. The red areas in Figure 4(b) and Figure 5(b) define the prostate tumor regions as marked by the radiologist. The ADC image is converted into a pseudo-color image and superimposed on T2W images (Figure 4(c) and Figure 5(c)).

Figure 4 shows the fusion of the ADC map and T2W image of patient 4. The tumor is located in the PZ and CG of the prostate. The normal PZ of the prostate presents two symmetrical crescent-shaped high-signal areas in the T2W image. In the ADC image, the tumor area has a low signal, which has a clear contrast with the surrounding normal area signal, but the tissue structure cannot be displayed well in the image. In the T2W image, the tumor area shows a low signal, and the rectum below the tumor and the bladder tissue above the tumor are clearly displayed.

The main disadvantage of image (d) is the limited ability to retain details, with poor contrast and brightness. The overall brightness value of the image is lower than that of the source image. The feature of some fat on the hip is missing. Image (e) has good overall performance, and the details such as fat texture on the hip are clear. However, the bladder area with high gray values in the ADC image and T2W image is mainly merged, so the contrast of the prostate area is not significantly improved. Image (f) contains some undesirable artifacts, and the overall image fusion effect is distorted. Image (g) performs well in the details of the prostate area, but the overall low-frequency components of the image show less detail than the source image, and the surrounding tissue structure is not clearly visible, which affects the diagnosis. Image (h) not only contains complete details but also has good contrast and clarity. The contrast of the prostate area is increased, and the gray value division is more uniform, which is of great help in identifying prostate cancer. The lower part of the tumor has grown to the junction of the prostate and rectum. The image corresponds to a tumor larger than T2.

Figure 5 shows the fusion the ADC map and T2W image of patient 3. This tumor's shape is irregularly and has invaded both the PZ and CG. In the T2W image, the tumor area represented by the low signal can be seen in the PZ on both sides. However, there is a suspected area between high and low signals in the lower right part of the prostate area in the image. ADC images are needed to help confirm whether there are malignant lesions at this location. In the ADC image, the suspected domain presents a high signal, and it is judged that there are no malignant cells in this area. The overall brightness of image (d) is lower than that of the source image, but the prostate cancer area can be better distinguished from the surrounding area. The overall image (e) contrast is relatively high, and the image texture and other details are
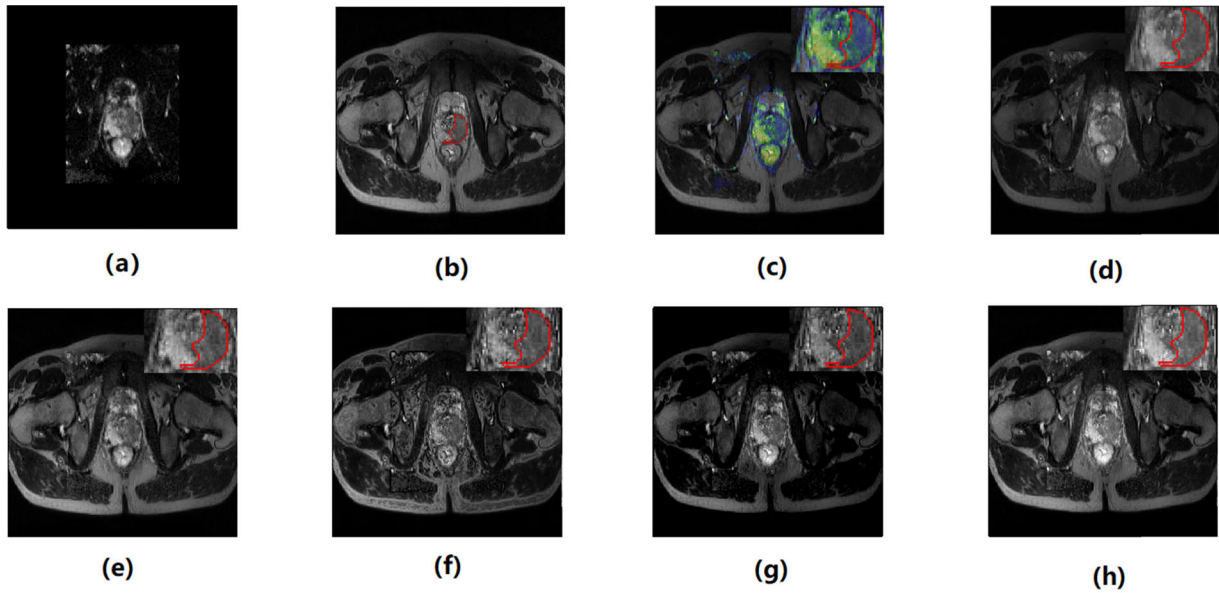
FIGURE 4. Fusion of magnetic resonance images from the ADC sequence and T2W sequence of patient 4. (a) Source 1 (ADC). (b) Source 2 (T2W). (c) Superposition of the ADC pseudocolor image and the T2W gray image. (d) AVE. (e) NSCT. (f) ZLF. (g) CNN. (h) Our proposed method.
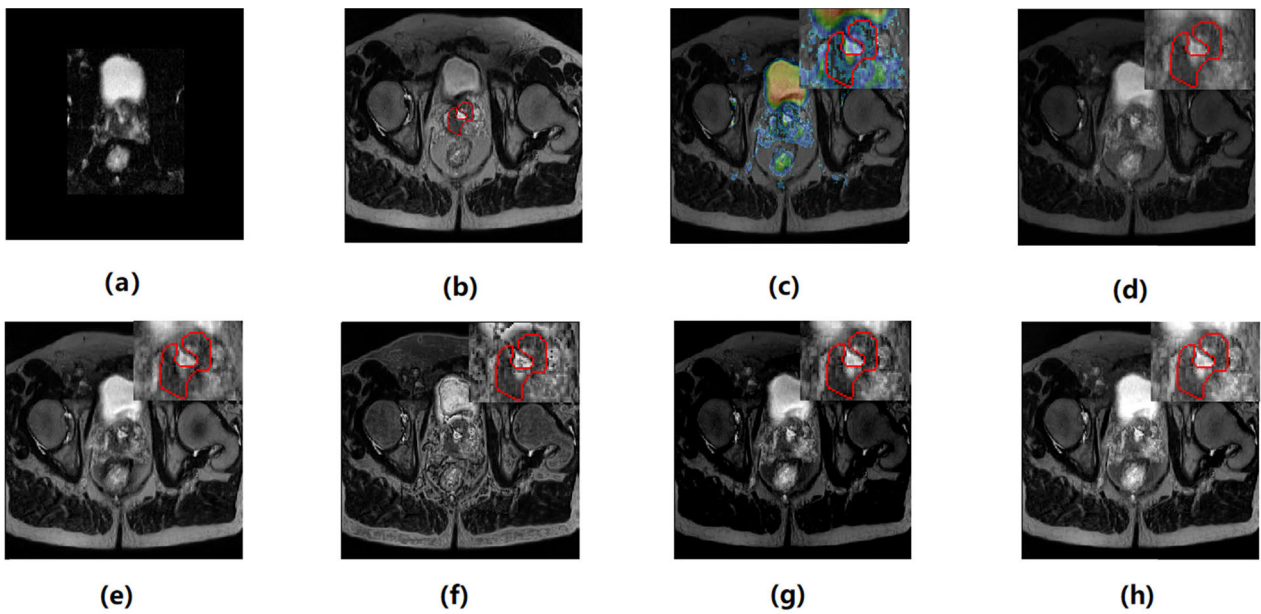
FIGURE 5. Fusion of magnetic resonance images from the ADC sequence and T2W sequence of patient 3. (a) Source 1 (ADC). (b) Source 2 (T2W). (c) Superposition of the ADC pseudocolor image and the T2W gray image. (d) AVE. (e) NSCT. (f) ZLF. (g) CNN. (h) Our proposed method.

clear. The contrast of the prostate area is not significantly improved. The lower right corner of the red area of prostate cancer in the figure shows a low signal, but there is no tumor in this part, which affects the diagnosis. Image (f) contains some undesirable artifacts, which can significantly reduce the visual perception ability of the observers. There are black blocks, and the overall image fusion effect is distorted. Image (g) performs well in the details of the prostate area, with high contrast. The main disadvantage is that some details are lost,

and the muscles and fat tissues of the rectum and bladder are not clearly visible. Image (h) not only contains complete details but also has good contrast and clarity. The contrast of the prostate area increases, so the low signal inside the tumor boundary and the high signal of normal tissue outside the boundary are visible, which can help the doctor confirm the tumor boundary.

Table 3 shows the quantitative performance evaluation of the different fusion methods. The proposed algorithm

TABLE 3. Average scores in the objective evaluation of various fusion methods.

	AVE	NSCT	ZLF	CNN	Our Proposed
EN	6.0579	6.7953	6.5448	6.7929	**6.84851**
FMI_w	0.4437	0.4832	0.3329	0.4668	**0.4930**
MI	12.1158	13.6449	13.0896	13.5819	**13.7150**
MS_SSIM	0.8572	0.9554	0.8604	0.9647	**0.9679**
$Q_{AB/F}$	0.4207	0.5621	0.6367	**0.6678**	0.6414

achieves better performance than the other four algorithms for most indexes except the $Q_{AB/F}$ index (our method is second only to CNN in this case). Our fusion algorithm is superior to all other algorithms based on both visual perception and objective evaluation.

Figure 6 shows the objective index values of the fusion results for the images of 10 patients using the algorithm presented in this article and the compared algorithms. We can observe from Figure 6 that our method almost leads the other methods in terms of scores. The quantitative results are consistent with the visualization results described in the above analysis.

IV. DISCUSSION

The AVE image fusion method is based on the manual selection of fusion rules. The weighted average fusion method combines all pixels according to a simple weight. The AVE algorithm is the fastest among the five methods due to its simple calculation. Since the pixel value around the clear area in the ADC image is 0, the contrast of the clear area in the fusion image decreases as the weight increases. Because the pixel value around the periprostate region in the ADC image is 0, the contrast of the periprostate region in the fused image decreases according to the overlap in the weight additions, and some information in the T2W image is lost. The NSCT method uses a nonsubsampled, contoured transform to decompose the image into low frequency and high frequency, uses different fusion rules according to these frequencies, separates the high-frequency part and the low-frequency part of the image, and designs different fusion rules for the high-pass subband and low-pass subband. The fused image has rich information and clear edges. Subjectively, the fused image is the closest to our fusion result. The MI value is second only to our algorithm in objective evaluation, but the calculation is complicated. The ZLF is a CNN trained by a large amount of data, and it can directly input medical images of different modalities to generate weight maps for fusion. Our experiments show that for the fusion of multimodal MRI images, there are many black spots in the images, and the fusion effect is not impressive. In the dataset used for this study, there are many black background areas around prostate tissue in the ADC images, but there is image information in the corresponding areas of the T2W images. This may be because the ZLF algorithm directly convolves the input image and does not consider the difference in information

content between two images at the same pixel position. Liu *et al.* proposed a method for medical image fusion, which added the assistance of the CNN network in the framework of the Laplacian pyramid. The CNN network is used to learn the features of the two input images, and the learning results are then inversely transformed into a weight map to participate in the Laplacian pyramid fusion. From the intuitive comparison, we can see that the fusion effect of the prostate region is better; however, the information in the T2W image outside the prostate region is ignored, and the fusion effect is too concentrated on the prostate region. Thus, the Qabf value is the greatest of all the methods, which is not conducive to doctors observing the tissues around the prostate from the image. We introduce U-Net into our method for the fusion of prostate MRI images to allow doctors to clearly identify prostate tumors. By using the characteristics of U-Net to segment the image, the areas with clear details of the MRI image are extracted and fused. In this paper, we add batch normalization and cascading to U-Net, which deepens the network structure and generates more accurate results than other methods when processing image details.

There are some limitations to this study. This study is based on a set of single-site MRI images. The method requires further validation on multicenter image trials with different anatomical areas, image resolution, or image quality by adapting a more robust and generalized deep learning network. Another limitation is that the study mainly focuses on image fusion methods for the ADC map and T2W sequence. Additional mpMRI images have been developed to detect tumors, such as DCE-MRI, magnetic resonance elastography, and chemical exchange saturation transfer MRI. Future studies may investigate how our proposed fusion method can be effectively applied to mpMRI images besides T2W and ADC images. The training dataset in the U-Net training experiment transforms the voc2012 dataset into a multifocus image. In the future, the medical image dataset will be changed to a multifocus image dataset for use in network training, and the network obtained may be more suitable for processing medical images. Based on the CNN method [29], Wang *et al.* obtained better image fusion results by using two sets of local similarity thresholds [51]. We can further investigate how to set different local similarity thresholds according to the image size of each layer of the pyramid to improve the quality of the fused image. Qi *et al.* proposed an algorithm based on EN to improve the quality of medical image fusion [52]. An image with high EN may have more meaningful information than an image with only high energy. In step 3, we may use the local energy entropy around the point (i, j) as the reference and take the random image blocks into the fusion process, rather than just according to the energy value at point (i, j). In this study, the main purpose of fusion imaging is to assist doctors in identifying areas with prostate cancer from the image. Radiologists may be invited to blindly evaluate the quality of the fused image and further optimize the fusion rules. Our method can be merged into a computer-aided diagnosis system to achieve effective and
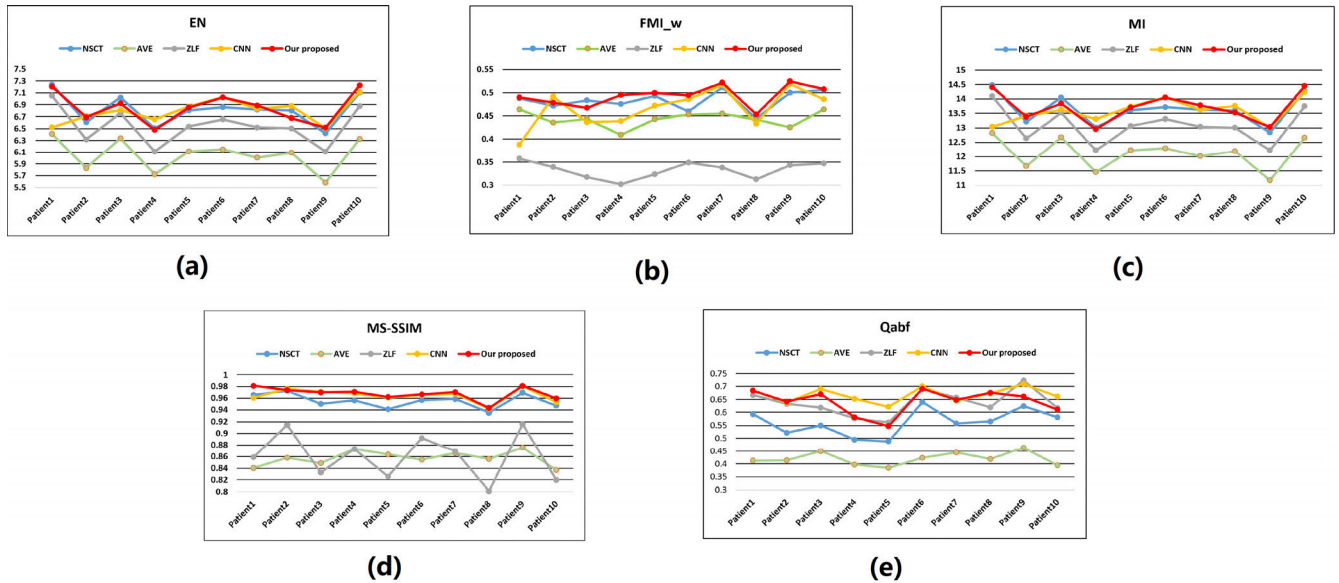
FIGURE 6. Line graph of scores for each pair of source images in terms of each metric (A) EN. (B) FMI_w. (C) MI. (D) MS_SSIM. (E) Q_(AB/F).

accurate automatic recognition and segmentation of prostate cancer.

V. CONCLUSION

this paper proposes an mpmri image fusion method based on cnns to enhance the appearances of tumors inside the prostate. inspired by its feature representation and extraction, a u-net is designed to generate weight maps based on pixel information. to effectively train u-net, we synthesize large-scale training samples with the voc 2012 natural image dataset, design image pairs with multifocus areas (target and background), and use the original unsmoothed images as fusion targets. the main innovation of this method is that a fusion weight map can be automatically generated through a u-net based on the relationship between image feature information and training targets. a weighted fusion strategy is proposed based on the weight map with the goal of avoiding artifact generation and energy loss. the effectiveness of this method is illustrated by fusing t2w and adc mri images from prostate cancer patients. the results show that our proposed method is superior to other similar methods in terms of not only overall visual perception but also the appearance of prostate cancer in the images. the python implementation of the proposed u-net image fusion method is available online at https://github.com/huangxunan/medfusion-unet.

REFERENCES

[1] O. Rouviere and P. C. Moldovan, "The current role of prostate multiparametric magnetic resonance imaging," *Asian J. Urol.*, vol. 6, no. 2, pp. 137–145, Apr. 2019, doi: 10.1016/j.ajur.2018.12.001.

[2] E. C. Serefoglu, S. Altinova, N. S. Ugras, E. Akincioglu, E. Asil, and M. D. Balbay, "How reliable is 12-core prostate biopsy procedure in the detection of prostate cancer?" *Eur. Urology Supplements*, vol. 9, no. 2, pp. 54–55, 2010, doi: 10.1016/S1569-9056(10)60075-9.

[3] C. V. Stempel, L. Dickinson, and D. Pendsé, "MRI in the management of prostate cancer," *Seminars Ultrasound, CT MRI*, vol. 41, no. 4, pp. 366–372, Aug. 2020, doi: 10.1053/j.sult.2020.04.003.

[4] V. Kumar, G. S. Bora, R. Kumar, and N. R. Jagannathan, "Multiparametric (mp) MRI of prostate cancer," *Prog. Nucl. Magn. Reson. Spectrosc.*, vol. 105, pp. 23–40, Apr. 2018, doi: 10.1016/j.pnmrs.2018.01.001.

[5] L. Xu, G. Zhang, B. Shi, Y. Liu, T. Zou, W. Yan, Y. Xiao, H. Xue, F. Feng, J. Lei, Z. Jin, and H. Sun, "Comparison of biparametric and multiparametric MRI in the diagnosis of prostate cancer," *Cancer Imag.*, vol. 19, no. 1, p. 394, Dec. 2019, doi: 10.1186/s40644-019-0274-9.

[6] T. Kobus, P. C. Vos, T. Hambrock, M. De Rooij, C. A. Hulsbergen—Van de Kaa, J. O. Barentsz, A. Heerschap, and T. W. J. Scheenen, "Prostate cancer aggressiveness: *In vivo* assessment of MR spectroscopy and diffusion-weighted imaging at 3 T," *Radiology*, vol. 265, no. 2, pp. 457–467, Nov. 2012, doi: 10.1148/radiol. 12111744.

[7] Y. Peng, Y. Jiang, C. Yang, J. B. Brown, T. Antic, I. Sethi, C. Schmid-Tannwald, M. L. Giger, S. E. Eggener, and A. Oto, "Quantitative analysis of multiparametric prostate MR images: Differentiation between prostate cancer and normal tissue and correlation with Gleason score—A computer-aided diagnosis development study," *Radiology*, vol. 267, no. 3, pp. 787–796, Jun. 2013, doi: 10.1148/radiol.13121454.

[8] R. Nagarajan, D. Margolis, S. Raman, M. K. Sarma, K. Sheng, C. R. King, G. Verma, J. Sayre, R. E. Reiter, and M. A. Thomas, "MR spectroscopic imaging and diffusion-weighted imaging of prostate cancer with gleason scores," *J. Magn. Reson. Imag.*, vol. 36, no. 3, pp. 697–703, Sep. 2012, doi: 10.1002/jmri.23676.

[9] H. Augustin, G. A. Fritz, T. Ehammer, M. Auprich, and K. Pummer, "Accuracy of 3-Tesla magnetic resonance imaging for the staging of prostate cancer in comparison to the partin tables," *Acta Radiologica*, vol. 50, no. 5, pp. 562–569, Jun. 2009, doi: 10.1080/02841850902889846.

[10] P. Steenbergen, K. Haustermans, E. Lerut, R. Oyen, L. De Wever, L. Van den Bergh, L. G. W. Kerkmeijer, F. A. Pameijer, W. B. Veldhuis, J. R. N. van der Voort van Zyp, F. J. Pos, S. W. Heijmink, R. Kalisvaart, H. J. Teertstra, C. V. Dinh, G. Ghobadi, and U. A. van der Heide, "Prostate tumor delineation using multiparametric magnetic resonance imaging: Inter-observer variability and pathology validation," *Radiotherapy Oncol.*, vol. 115, no. 2, pp. 186–190, May 2015, doi: 10.1016/j.radonc.2015.04.012.

[11] M. A. Jacobs, R. Ouwerkerk, K. Petrowski, and K. J. Macura, "Diffusion-weighted imaging with apparent diffusion coefficient mapping and spectroscopy in prostate cancer," *Topics Magn. Reson. Imag.*, vol. 19, no. 6, pp. 261–272, Nov. 2008, doi: 10.1097/RMR.0b013e3181aa6b50.

[12] A. G. A. Sorial, O. F. Kamel, and M. G. H. Ismail, "Role of ADC map MR imaging in prediction of local aggressiveness of prostate cancer," *Egyptian J. Hospital Med.*, vol. 71, no. 2, pp. 2490–2497, Apr. 2018, doi: 10.12816/0045646.

[13] D. P. Bavirisetti and R. Dhuli, "Multi-focus image fusion using multi-scale image decomposition and saliency detection," *Ain Shams Eng. J.*, vol. 9, no. 4, pp. 1103–1117, Dec. 2018, doi: 10.1016/j.asej.2016.06.011.

[14] H. Yin, Y. Li, Y. Chai, Z. Liu, and Z. Zhu, "A novel sparse-representation-based multi-focus image fusion approach," *Neurocomputing*, vol. 216, pp. 216–229, Dec. 2016, doi: 10.1016/j.neucom.2016.07.039.

[15] J. Sun, Q. Han, L. Kou, L. Zhang, K. Zhang, and Z. Jin, "Multi-focus image fusion algorithm based on Laplacian pyramids," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 35, no. 3, pp. 480–490, 2018, doi: 10.1364/JOSAA.35.000480.

[16] X. Liang, P. Hu, L. Zhang, J. Sun, and G. Yin, "MCFNet: Multi-layer concatenation fusion network for medical images fusion," *IEEE Sensors J.*, vol. 19, no. 16, pp. 7107–7119, Aug. 2019, doi: 10.1109/JSEN.2019.2913281.

[17] A. P. James and B. V. Dasarathy, "Medical image fusion: A survey of the state of the art," *Inf. Fusion*, vol. 19, pp. 4–19, Sep. 2014, doi: 10.1016/j.inffus.2013.12.002.

[18] W. Li, J. Du, Z. Zhao, and J. Long, "Fusion of medical sensors using adaptive cloud model in local Laplacian pyramid domain," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 4, pp. 1172–1183, Apr. 2019, doi: 10.1109/TBME.2018.2869432.

[19] J. Du, W. Li, B. Xiao, and Q. Nawaz, "Union Laplacian pyramid with multiple features for medical image fusion," *Neurocomputing*, vol. 194, pp. 326–339, Jun. 2016, doi: 10.1016/j.neucom.2016.02.047.

[20] A. Sahu, V. Bhateja, A. Krishn, and Himanshi, "Medical image fusion with Laplacian pyramids," in *Proc. Int. Conf. Med. Imag., m-Health Emerg. Commun. Syst. (MedCom)*, Greater Noida, India, Nov. 2014, pp. 448–453.

[21] Y. Sun, H. M. Reynolds, B. Parameswaran, D. Wraith, M. E. Finnegan, S. Williams, and A. Haworth, "Multiparametric MRI and radiomics in prostate cancer: A review," *Australas. Phys. Eng. Sci. Med.*, vol. 42, no. 1, pp. 3–25, Mar. 2019, doi: 10.1007/s13246-019-00730-z.

[22] X. Wang, W. Yang, J. Weinreb, J. Han, Q. Li, X. Kong, Y. Yan, Z. Ke, B. Luo, T. Liu, and L. Wang, "Searching for prostate cancer by fully automated magnetic resonance imaging classification: Deep learning versus non-deep learning," *Sci. Rep.*, vol. 7, no. 1, p. 15415, Dec. 2017, doi: 10.1038/s41598-017-15720-y.

[23] K. H. Cha, L. M. Hadjiiski, R. K. Samala, H.-P. Chan, R. H. Cohan, E. M. Caoili, C. Paramagul, A. Alva, and A. Z. Weizer, "Bladder cancer segmentation in CT for treatment response assessment: Application of deep-learning convolution neural network—A pilot study," *Tomography*, vol. 2, no. 4, pp. 421–429, 2016, doi: 10.18383/j.tom.2016.00184.

[24] V. R and A. K, "Image fusion a review of methods and applications," *Int. J. Comput. Sci. Eng.*, vol. 6, no. 9, pp. 556–566, Sep. 2018, doi: 10.26438/ijcse/v6i9.556566.

[25] M. Kaur and D. Singh, "Fusion of medical images using deep belief networks," *Cluster Comput.*, vol. 23, no. 2, pp. 1439–1453, Jun. 2020, doi: 10.1007/S10586-019-02999-X.

[26] Y. Cui, H. Du, and W. Mei, "Infrared and visible image fusion using detail enhanced channel attention network," *IEEE Access*, vol. 7, pp. 182185–182197, 2019, doi: 10.1109/ACCESS.2019.2959034.

[27] F. Palsson, J. R. Sveinsson, and M. O. Ulfarsson, "Multispectral and hyperspectral image fusion using a 3-D-convolutional neural network," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 639–643, May 2017, doi: 10.1109/LGRS.2017.2668299.

[28] S. Li, J. T. Kwok, and Y. Wang, "Multifocus image fusion using artificial neural networks," *Pattern Recognit. Lett.*, vol. 23, no. 8, pp. 985–997, 2002, doi: 10.1016/S0167-8655(02)00029-6.

[29] Y. Liu, X. Chen, J. Cheng, and H. Peng, "A medical image fusion method based on convolutional neural networks," in *Proc. 20th Int. Conf. Inf. Fusion (Fusion)*, Xi'an, China, Jul. 2017, pp. 1–7.

[30] G. Wang, W. Li, S. Ourselin, and T. Vercauteren, "Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks," in *Proc. Brainlesion, Glioma, Multiple Sclerosis, Stroke Traumatic Brain Injuries, Int. MICCAI Brainlesion Workshop 3rd Int. Workshop*, Quebec City, QC, Canada, vol. 10670, A. Crimi, S. Bakas, H. Kuijf, B. Menze, and M. Reyes, Eds. Cham, Switzerland: Springer, Sep. 2017, pp. 178–190, doi: 10.1007/978-3-319-75238-9_16.

[31] T. Falk *et al.*, "U-net: Deep learning for cell counting, detection, and morphometry," *Nature Methods*, vol. 16, no. 1, pp. 67–70, Jan. 2019, doi: 10.1038/s41592-018-0261-2.

[32] X. Pan, L. Li, D. Yang, Y. He, Z. Liu, and H. Yang, "An accurate nuclei segmentation algorithm in pathological image based on deep semantic network," *IEEE Access*, vol. 7, pp. 110674–110686, 2019, doi: 10.1109/ACCESS.2019.2934486.

[33] M. Drozdzal, G. Chartrand, E. Vorontsov, M. Shakeri, L. D. Jorio, A. Tang, A. Romero, Y. Bengio, C. Pal, and S. Kadoury, "Learning normalized inputs for iterative estimation in medical image segmentation," *Med. Image Anal.*, vol. 44, pp. 1–13, Feb. 2018, doi: 10.1016/j.media.2017.11.005.

[34] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proc. 3DV 4th Int. Conf. 3D Vis.*, Stanford, CA, USA, Oct. 2016, pp. 565–571.

[35] S. Zheng, P. Qi, S. Chen, and X. Yang, "Fusion methods for CNN-based automatic modulation classification," *IEEE Access*, vol. 7, pp. 66496–66504, 2019, doi: 10.1109/ACCESS.2019.2918136.

[36] H. Chen, Y. Zhang, M. K. Kalra, F. Lin, Y. Chen, P. Liao, J. Zhou, and G. Wang, "Low-dose CT with a residual encoder-decoder convolutional neural network," *IEEE Trans. Med. Imag.*, vol. 36, no. 12, pp. 2524–2535, Dec. 2017, doi: 10.1109/TMI.2017.2715284.

[37] C. You, W. Cong, G. Wang, Q. Yang, H. Shan, L. Gjesteby, G. Li, S. Ju, Z. Zhang, Z. Zhao, and Y. Zhang, "Structurally-sensitive multi-scale deep neural network for low-dose CT denoising," *IEEE Access*, vol. 6, pp. 41839–41855, 2018, doi: 10.1109/ACCESS.2018.2858196.

[38] S. Chen, H. Yang, J. Fu, W. Mei, S. Ren, Y. Liu, Z. Zhu, L. Liu, H. Li, and H. Chen, "U-net plus: Deep semantic segmentation for esophagus and esophageal cancer in computed tomography images," *IEEE Access*, vol. 7, pp. 82867–82877, 2019, doi: 10.1109/ACCESS.2019.2923760.

[39] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI* (Lecture Notes in Computer Science), vol. 9351, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds., 1st ed. Cham, Switzerland: Springer, 2015, pp. 234–241.

[40] 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, 2011, doi: 10.1016/j.compeleceng.2011.07.012.

[41] B. Bondzulic and V. Petrovic, "Objective image fusion performance measures," *Vojnotehnicki glasnik*, vol. 56, no. 2, pp. 181–193, 2008, doi: 10.5937/vojtehg0802181B.

[42] 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.

[43] G. Qu, D. Zhang, and P. Yan, "Information measure for performance of image fusion," *Electron. Lett.*, vol. 38, no. 7, p. 313, 2002, doi: 10.1049/el:20020212.

[44] A. P. S. Kirkham, M. Emberton, and C. Allen, "How good is MRI at detecting and characterising cancer within the prostate?" *Eur. Urol.*, vol. 50, no. 6, pp. 1163–1175, Dec. 2006, doi: 10.1016/j.eururo.2006.06.025.

[45] K. Hosseinzadeh and S. D. Schwarz, "Endorectal diffusion-weighted imaging in prostate cancer to differentiate malignant and benign peripheral zone tissue," *J. Magn. Reson. Imag.*, vol. 20, no. 4, pp. 654–661, Oct. 2004, doi: 10.1002/jmri.20159.

[46] H. Qu, H. Wang, and H. Xu, "Medical image fusion based on statistical modeling," *J. Phys. Conf. Ser.*, vol. 1187, no. 4, p. 42042, 2019, doi: 10.1088/1742-6596/1187/4/042042.

[47] Z. Zhu, M. Zheng, G. Qi, D. Wang, and Y. Xiang, "A phase congruency and local Laplacian energy based multi-modality medical image fusion method in NSCT domain," *IEEE Access*, vol. 7, pp. 20811–20824, 2019, doi: 10.1109/ACCESS.2019.2898111.

[48] F. Lahoud and S. Süsstrunk, "Zero-learning fast medical image fusion," in *Proc. 22nd Int. Conf. Inf. Fusion (FUSION)*. Ottawa, ON, Canada: IEEE Press, Jul. 2019, pp. 1–8.

[49] X. Huang, H. Chafi, K. L. Matthews, O. Carmichael, T. Li, Q. Miao, S. Wang, and G. Jia, "Magnetic resonance elastography of the brain: A study of feasibility and reproducibility using an ergonomic pillow-like passive driver," *Magn. Reson. Imag.*, vol. 59, pp. 68–76, Jun. 2019, doi: 10.1016/j.mri.2019.03.009.

[50] X. Huang, R. N. Schurr, S. Wang, Q. Miao, T. Li, and G. Jia, "Development of radiofrequency saturation amplitude-independent quantitative markers for magnetization transfer MRI of prostate cancer," *Current Med. Imag. Formerly Current Med. Imag. Rev.*, vol. 16, no. 6, pp. 695–702, Jul. 2020, doi: 10.2174/1573405615666190318153328.

[51] K. Wang, M. Zheng, H. Wei, G. Qi, and Y. Li, "Multi-modality medical image fusion using convolutional neural network and contrast pyramid," *Sensors*, vol. 20, no. 8, p. 2169, Apr. 2020, doi: 10.3390/s20082169.

[52] G. Qi, J. Wang, Q. Zhang, F. Zeng, and Z. Zhu, "An integrated dictionary-learning entropy-based medical image fusion framework," *Future Internet*, vol. 9, no. 4, p. 61, Oct. 2017, doi: 10.3390/fi9040061.

XUNAN HUANG received the M.S. degree from the School of Aerospace Science and Technology, Xidian University, China, in 2015, where he is currently pursuing the Ph.D. degree with the School of Computer Science and Technology. His current interests include medical image processing, pattern recognition, and deep learning.

BO ZHANG received the B.S. degree in clinical medicine from the Fourth Military Medical University, Xi'an, in 1990, and the M.S. and Ph.D. degrees in urology from Xijing Hospital, Fourth Military Medical University, in 2002 and 2005, respectively.

From 2001 to 2004, he was a Visiting Scholar with The Ohio State University Medical Center, Columbus, OH, USA. From 2004 to 2007, he was the Director of the Kidney Transplantation Center, Tangdu Hospital, Fourth Military Medical University. Since 2004, he has been an Assistant Professor with the Urology Surgery Department, Tangdu Hospital, Fourth Military Medical University. In 2018, he served as the Urology Director of Tangdu Hospital, Fourth Military Medical University. He has been supported by two general projects of the National Natural Science Foundation of China and seven projects of Shaanxi Province. He is the author of four books, more than 80 articles, and more than seven inventions. His research interests include clinical research work of 3D printing in urology, renal autotransplantation, urogenital tumors, urological endoscopy, and minimally invasive robotic surgery. He was a recipient of the Second Prize of Military Medical Achievements. He is an Associate Editor of the *Journal of Stem Cells*.

XIAOLING ZHANG received the Ph.D. degree from the School of Medicine, Xi'an Jiao Tong University, in 2019. She is currently the Director of the MRI Department with Shaanxi Provincial People's Hospital and a Professor of Xi'an Medical College. She is the author of more than 70 articles. Her current research interests include medical imaging in neurological and mental diseases and clinical research on functional magnetic resonance imaging, which is supported by two National Natural Science Grants and two scientific projects of Shaanxi Province.
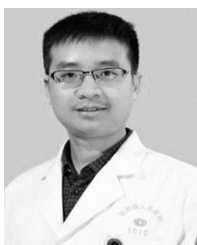
MIN TANG is currently pursuing the M.D. degree with the Medical School, Xi'an Jiaotong University. He is currently working with the Department of MRI, Shaanxi Provincial People's Hospital. He is engaged in functional magnetic resonance imaging and new technology research.

QIGUANG MIAO (Senior Member, IEEE) received the Ph.D. degree in computer application technology from Xidian University, in 2005.

He is currently a Professor and a Ph.D. Student Supervisor with the School of Computer Science and Technology, Xidian University. In 2012, he was supported by the Program for New Century Excellent Talents at the University by the Ministry of Education. As the Principal Investigator, he is conducting or has completed four projects through the NSFC, two projects through the Shaanxi Provincial Natural Science Fund, over ten projects through the National Defence Preresearch Foundation, and 863 projects through the Weapons and Equipment Fund. He has hosted one project supported by the Fundamental Research Funds for the Central Universities by MOE In recent years, he has published over 100 papers in significant domestic and international journals or conferences, including the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, the *Journal of Visual Communication and Image Representation*, *NeuroComputing*, and the *IET Image Processing*, *Knowledge Based Systems*, 30 of which are indexed by SCI, and over 40 of which are included in EI. His research interests include computer vision, machine learning, and Big Data.

He is a member of the Professor Committee, a Committee Member of the CCF and CCF Computer Vision, and the Vice Chairman of CCF YOCSEF. In the field of teaching, he was named the Pacemaker of Ten Excellent Teacher in 2008, 2011, and 2014. He has served as the Committee Chairman for the First CCF Youth Elite Association, CNCC2008, CIS 2012, CCFAI 2013, CCDM2014 PC member, and CIS 2013 Special Session Chair. He is a Committee Member of the Editorial Board of the Internet of Things and the Assessment Expert of the State Science and Technology Prizes and the National Defense Basic Scientific Research Project. He was a recipient of a prize at the Ministerial and Provincial Level twice.

TANPING LI received the B.S. and M.S. degrees in theoretical physics from Lanzhou University, Lanzhou, China, in 1997 and 2000, respectively, and the Ph.D. degree in biophysics from The Ohio State University, Columbus, OH, USA, in 2011.

From 2011 to 2012, she was a Postdoctoral Researcher with the Department of Chemistry, The Ohio State University. From 2013 to 2015, she was a Research Associate with the Department of Chemistry, Louisiana State University, Baton Rouge, LA, USA. Since 2017, she has been an Associate Professor with the School of Physics and Optoelectronic Engineering, Xidian University, Xi'an, China. Her research interests include computer simulation and image processing.
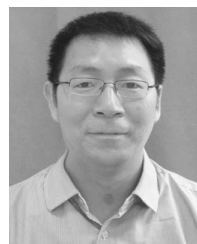
GUANG JIA received the B.S. degree in theoretical physics from Lanzhou University, in 1997, and the Ph.D. degree in biophysics from The Ohio State University, Columbus, OH, USA, in 2006.

From 2006 to 2008, he was a Postdoctoral Researcher with the Department of Radiology, The Ohio State University. From 2008 to 2011, he was an Instructor with the Department of Radiology, The Ohio State University. From 2011 to 2013, he was an Assistant Professor with the Department of Radiology, The Ohio State University. From 2013 to 2016, he was an Associate Professor with the Department of Physics, Louisiana State University, Baton Rouge, LA, USA. Since 2017, he has been a Professor with the School of Computer Science and Technology, Xidian University, Xi'an, China. He is the author of 50 articles and ten inventions. His research interests include diagnostic medical physics and AI-based medical image processing.

Dr. Jia was a recipient of the Radiological Society of North America Research Trainee Prize, in 2008, the American Urological Association Foundation Outstanding Graduate Scholar Award, in 2011, and the American College of Radiology Imaging Network Young Investigator Award, in 2012.

• • •