Letter

Glioma Segmentation-Oriented Multi-Modal MR Image Fusion With Adversarial Learning

Yu Liu, *Member, IEEE*, Yu Shi, Fuhao Mu, Juan Cheng, and Xun Chen, *Senior Member, IEEE*

Dear Editor,

In recent years, multi-modal medical image fusion has received widespread attention in the image processing community. However, existing works on medical image fusion methods are mostly devoted to pursuing high performance on visual perception and objective fusion metrics, while ignoring the specific purpose in clinical applications. In this letter, we propose a glioma segmentation-oriented multi-modal magnetic resonance (MR) image fusion method using an adversarial learning framework, which adopts a segmentation network as the discriminator to achieve more meaningful fusion results from the perspective of the segmentation task. Experimental results demonstrate the advantage of the proposed method over some state-of-the-art medical image fusion methods.

Multi-modal medical image fusion aims to combine the complementary information contained in the source images of different modalities by generating a composite fused image, which is expected to be more informative for human or machine perception. During the past few decades, a variety of medical image fusion methods have been proposed. Most existing methods are developed under a popular three-phase image fusion framework, namely, decomposition, fusion and reconstruction [1]. According to the decomposition approach adopted, conventional multi-modal image fusion methods can be divided into several categories including multi-scale decomposition (MSD)-based methods [2]–[4], sparse representation (SR)-based methods [5]–[7], hybrid transform-based methods [8], [9], spatial domain methods [10], [11], etc. Recently, deep learning (DL) has emerged as a hotspot in the field of image fusion [12], [13] and some DL-based medical image fusion methods have been introduced in the literature, such as the general image fusion framework via the convolutional neural network (IFCNN) [14], the enhanced medical image fusion (EMFusion) method [15], the dual-discriminator conditional generative adversarial network (DDcGAN)-based method [16], and the unified and unsupervised image fusion (U2Fusion) method [17].

Although the study on medical image fusion has achieved considerable progress in recent years, it is worth noting that current works suffer from a common drawback, namely, there is a severe lack of clinical problem-oriented study. The primary target of most existing medical image fusion methods is to achieve fused images with pleasing visual quality and high performance on objective fusion metrics that are used in a broader range of image fusion tasks (i.e., not limited to medical image fusion). However, they ignore the specific purpose of the corresponding images in clinical applications, which limits the practical value of image fusion methods to a great extent.

As the most common primary brain malignancy, glioma has always been a serious health hazard to human. In clinical practice, the automatic segmentation of gliomas from multi-modal MR images is of great significance to the diagnosis and treatment of this disease. In this letter, we present a glioma segmentation-oriented multi-modal MR image fusion method with adversarial learning. By introducing a segmentation network as the discriminator to guide the fusion model, the fused images obtained are more meaningful in terms of the segmentation task. The fused modality can strengthen the association of different pathological information of tumors captured by multiple source modalities by integrating them into a composite image, which is helpful to the segmentation task as well as physician observation.

Methodology:

Overall framework: In the glioma segmentation task, the contrast-enhanced T1-weighted (T1c) and the fluid attenuated inversion recovery (Flair) are two frequently used magnetic resonance imaging (MRI) modalities. The former one can well characterize the tumor core areas, while the latter one can effectively capture the edema areas that surround the tumor core. Therefore, we mainly concentrate on the fusion of T1c and Flair modalities in this work. Motivated by the recent progress in generative adversarial network (GAN)-based image fusion such as the FusionGAN [18] and DDcGAN [16], we propose an adversarial learning framework for multi-modal MR image fusion. Unlike all the existing GAN-based fusion methods that adopt a classification network as the discriminator model, a semantic segmentation network is used as the discriminator in our fusion framework to distinguish the fused image from the source images, aiming to assist the fusion network (i.e., the generator) to extract sufficient pathological information that is related to tumor segmentation from the source images.

Fig. 1 shows the overall framework of the proposed multi-modal MR image fusion method. The source images $\mathbf{I}_{T1c}$ and $\mathbf{I}_{\text{Flair}}$ are concatenated and fed to a fusion network to generate the fused image $\mathbf{I}_f$. Two segmentation network-based discriminators are adopted to further improve the capability of the generator in preserving tumor pathological information. The discriminator D_{T1c} aims to distinguish $\mathbf{I}_f$ from $\mathbf{I}_{T1c}$ via the segmentation results on the tumor core area, while the discriminator D_{Flair} aims to distinguish $\mathbf{I}_f$ from $\mathbf{I}_{\text{Flair}}$ via the segmentation results on the whole tumor area. The generator is encouraged to output a fused image with rich pathological information about the tumor area to fool the discriminators. By this means, the proposed framework builds an adversarial mechanism that is mainly concerned with the tumor regions, instead of the entire images considered in previous GAN-based fusion methods. During the training phase, the fused image gradually absorbs the information of the tumor area contained in the source images via alternating training of the generator and two discriminators. The training procedure is summarized in Algorithm 1. In each iteration, the two discriminators are both trained k times sequentially and then the

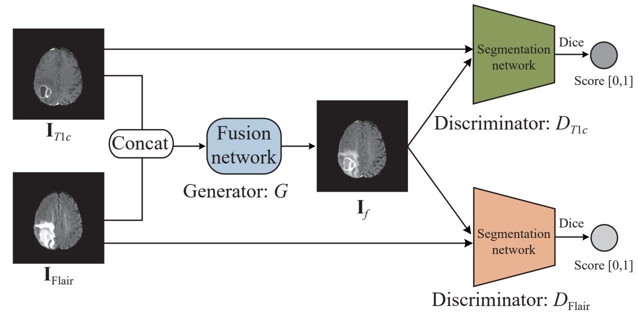


Fig. 1. The overall framework of the proposed fusion method.

Corresponding author: Xun Chen.

Citation: Y. Liu, Y. Shi, F. H. Mu, J. Cheng, and X. Chen, "Glioma segmentation-oriented multi-modal MR image fusion with adversarial learning," *IEEE/CAA J. Autom. Sinica*, vol. 9, no. 8, pp. 1528–1531, Aug. 2022.

Y. Liu, Y. Shi, F. H. Mu and J. Cheng are with the Department of Biomedical Engineering, Hefei University of Technology, Hefei 230009, and also with the Anhui Province Key Laboratory of Measuring Theory and Precision Instrument, Hefei University of Technology, Hefei 230009, China (e-mail: yuliu@hfut.edu.cn; yushi@mail.hfut.edu.cn; fuhaomu@mail.hfut.edu.cn; chengjuan@hfut.edu.cn).

X. Chen is with the Department of Neurosurgery, the First Affiliated Hospital of USTC, Division of Life Sciences and Medicine, and also with the Department of Electronic Engineering and Information Science, University of Science and Technology of China, Hefei 230001, China (e-mail: xunchen@ustc.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JAS.2022.105770

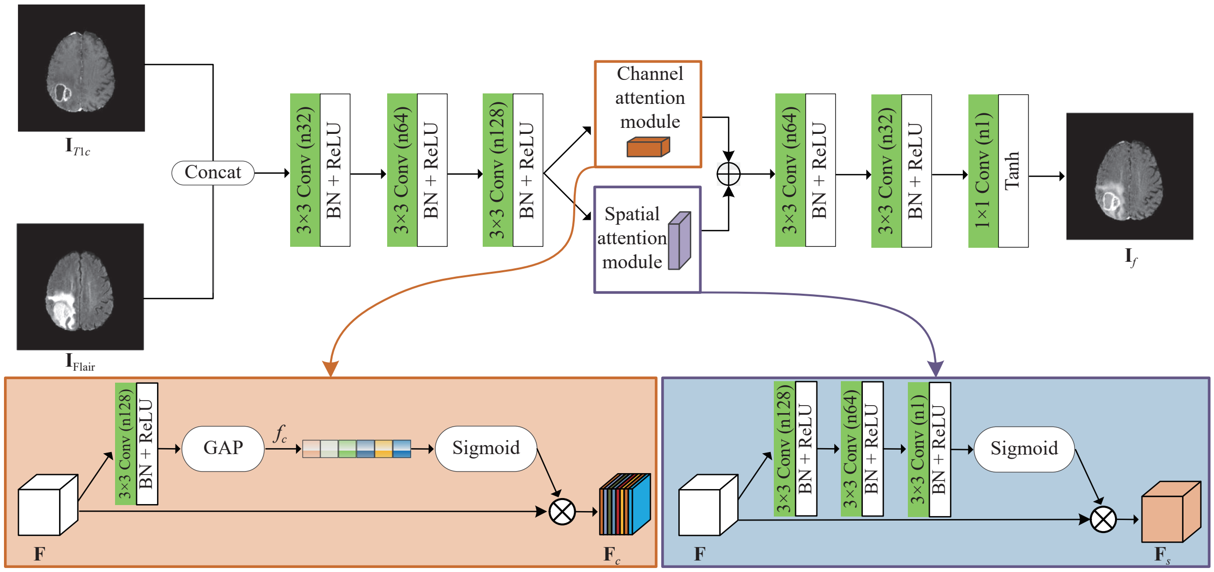


Fig. 2. The architecture of our generator network. Conv(nk): Convolutional layer with k filters, BN: Batch normalization, ReLU and Tanh: Activation layers; GAP: Global average pooling, f_c: Fully connected layer.

generator is trained. Once the generated fused image cannot be distinguished by the discriminators, we obtain the generator as the trained fusion network, which can generate expected fused image that contains sufficient pathological information about the tumor area.

Algorithm 1 Training Procedure of the Proposed Method

Initialize the parameters of discriminator D_{T1c} and D_{Flair} and generator G, k is set to 2.
for number of training iterations **do**
  # Train the discriminator D_{T1c}:
  for k steps **do**
    Select m fused images $\{\mathbf{I}_f^1, ..., \mathbf{I}_f^m\}$ from generator G.
    Select m $T1c$ images $\mathbf{I}_{T1c}^1, ..., \mathbf{I}_{T1c}^m\}$.
    Update discriminator D_{T1c} parameters by AdamOptimizer to minimize $L_{D_{T1c}}$ in (7).
  end for
  # Train the discriminator D_{Flair}:
  for k steps **do**
    Select m fused images $\{\mathbf{I}_f^1, ..., \mathbf{I}_f^m\}$ from generator G.
    Select m Flair images $\mathbf{I}_{\text{Flair}}^1, ..., \mathbf{I}_{\text{Flair}}^m\}$.
    Update discriminator D_{Flair} parameters by AdamOptimizer to minimize $L_{D_{\text{Flair}}}$ in (7).
  end for
  # Train the generator G:
  Select m $T1c$ images $\{\mathbf{I}_{T1c}^1, ..., \mathbf{I}_{T1c}^m\}$ and m Flair images $\{\mathbf{I}_{\text{Flair}}^1, ..., \mathbf{I}_{\text{Flair}}^m\}$ from the dataset.
  Update generator parameters by AdamOptimizer to minimize L_G in (1).
end for

Network architecture: In this work, we adopt relatively plain architectures to design the generator and discriminator, and the results demonstrate that they have been able to achieve good performance. The network architecture of the generator is shown in Fig. 2. First, the source images are concatenated and fed to three consecutive convolutional layers for feature extraction. Then, a parallel attention structure that consists of a channel attention module (CAM) and a spatial attention module (SAM) is adopted for feature refinement. The features obtained by the two attention branches are further fused by an addition operation. Finally, the refined feature maps pass through three convolutional layers to reconstruct the fused image. For the CAM, the input feature $\mathbf{F} \in \mathbb{R}^{H\times W\times C}$ is first fed to a convolutional layer to obtain $\mathbf{F}_1 \in \mathbb{R}^{H\times W\times C}$. Then, the global average pooling (GAP) operation is performed to generate a feature vector $\mathbf{z} \in \mathbb{R}^{1\times 1\times C}$ by shrinking $\mathbf{F}_1$ through its spatial dimensions $H \times W$. Next, $\mathbf{z}$ passes through a fully connected layer and a sigmoid function to generate the activations of the channel attention module. Finally, the output $\mathbf{F}_c \in \mathbb{R}^{H\times W\times C}$ of the CAM branch is obtained by rescaling $\mathbf{F}$ with the activations. The goal of the SAM branch is to generate a spatial attention map to recalibrate the input feature $\mathbf{F}$. It mainly consists of three convolutional layers and a sigmoid function to obtain the spatial attention map $\mathbf{Y} \in \mathbb{R}^{H\times W\times 1}$. $\mathbf{F}$ is weighted by the attention map to obtain the output $\mathbf{F}_s \in \mathbb{R}^{H\times W\times C}$ of the SAM. For each discriminator, the U-Net architecture [19] is used due to its popularity in medical image segmentation.

Loss function: The loss function of the proposed method consists of three parts, i.e., the loss function of the generator, the loss function of the discriminator D_{T1c} and the discriminator D_{Flair}. The least squares-based GAN loss model [20] is adopted in this work because of its stability for training. The loss function of the generator is defined as

$$L_G = L_{adv} + \alpha L_{\text{content}} \tag{1}$$

where L_{adv} denotes the adversarial loss between the generator and the discriminators, while L_{content} denotes content loss of image fusion. The parameter α is used to balance these two terms. The adversarial loss is formulated as

$$L_{adv} = (\text{Dice}(D_{T1c}(\mathbf{I}_f)) - c)^2 + (\text{Dice}(D_{\text{Flair}}(\mathbf{I}_f)) - c)^2 \tag{2}$$

where Dice($\cdot$) denotes the calculation of Dice coefficient between the given segmentation map and the ground truth. It is defined as

$$Dice = \frac{2\sum_{i=1}^{N} p_i g_i + \varepsilon}{\sum_{i=1}^{N} p_i^2 + \sum_{i=1}^{N} g_i^2 + \varepsilon} \tag{3}$$

where $p_i \in P$ is the predicted segmentation map, $g_i \in G$ is the ground truth segmentation map, N is the number of pixels, and ε is a small constant to avoid dividing by 0. The Dice coefficient is a commonly used approach to measure the segmentation accuracy, and its value is between 0 and 1. The parameter c in (2) denotes the score that the generator expects the discriminators to obtain from the generated fused image. The content loss L_{content} is defined as

$$L_{\text{content}} = L_{\text{ssim}} + \beta L_{\text{pixel}} \tag{4}$$

where β is used to control the balance between the structural similarity (SSIM) [21] based term L_{ssim} and the pixel-based term L_{pixel}. The L_{ssim} is formulated as

$$L_{\text{ssim}} = 2(1 - \text{SSIM}(\mathbf{I}_{T1c}, \mathbf{I}_f)) + (1 - \text{SSIM}(\mathbf{I}_{\text{Flair}}, \mathbf{I}_f)) \tag{5}$$

where SSIM($\cdot$) denotes the SSIM measure between two images. Considering that the T1c images contains more structural information about anatomic tissues, the weight for $\mathbf{I}_{T1c}$ is set to a larger value (i.e., 2). The pixel loss L_{pixel} is defined as

$$L_{\text{pixel}} = \frac{1}{H \times W}(\|\mathbf{I}_f - \mathbf{I}_{T1c}\|_F + 2\|\mathbf{I}_f - \mathbf{I}_{\text{Flair}}\|_F) \tag{6}$$

where $H \times W$ denote the spatial resolution of the input image and $\|\cdot\|_F$ denotes the Frobenius norm. The weight for $\mathbf{I}_{\mathrm{Flair}}$ is set to a larger value (i.e., 2) to preserve more intensity information from the Flair image since the lesions are likely to have distinct intensity in the image of this modality.

The loss function of each discriminator is defined as

$$L_{D_m} = (\mathrm{Dice}(D_m(\mathbf{I}_f)) - a)^2 + (\mathrm{Dice}(D_m(\mathbf{I}_m)) - b)^2 \quad (7)$$

where D_m denotes a certain discriminator network and $\mathbf{I}_m$ is the corresponding source image. The parameters a and b are the expected scores of the discriminator to obtain from the fused image and the source image, respectively.

Experiments:

Experimental setup: The proposed fusion model is trained and validated on the multi-modal MRI glioma dataset released by the MICCAI brain tumor segmentation (BraTS) 2019 challenge. The training set that contains 335 multi-modal MRI scans of size $240 \times 240 \times 155$ in this dataset is mainly adopted to train our fusion model. We totally generate 4397 pairs of T1c and Flair slices from these scans for network training, based on the principle that the selected slices must contain the tumor regions. The size of each slice is 240×240. 30 pairs of T1c and Flair slices are employed to validate the effectiveness of the trained fusion network. They are selected from 30 sets of multi-modal MRI scans from the BraTS 2019 validation set. The parameters α and β in (1) and (4) are experimentally set to 2 and 450, respectively. For easier training, soft labels are used to define the parameters a, b and c [20]. In our experiments, a is a random value ranging from 0 to 0.3, while b and c randomly range from 0.7 to 1.2. The batch size is set to 1 and the learning rate is fixed at 1×10^{-4} for both the generator and the discriminators. It takes about 11 hours to complete the training process lasting 10 epoches on an NVIDIA TITAN RTX GPU. The proposed method is compared with four representative MSD-based or DL-based medical image fusion methods: DTCWT [22], LPCNN [23], U2Fusion [17] and IFCNN [14]. For objective evaluation, since existing image fusion metrics can be divided into four categories: information theory-based, image feature-based, SSIM-based and human visual perception-based [24], we select one popular metric from each category to make a comprehensive assessment. Specifically, they are the normalized mutual information Q_{MI} [25], the gradient-based metric Q_G [26], the SSIM-based metric Q_W [27] and the visual information fidelity fusion metric *VIFF* [28].

Fusion results: The objective assessment results of different fusion methods is listed in Table 1, in which the average score over 30 testing samples in each case is reported. For each metric, the optimal value is indicated in bold and the suboptimal one is underlined. It can be seen that the proposed method outperforms other four methods on Q_W and *VIFF* with obvious advantages. It ranks the second place on Q_{MI}, but the gap to the best performing one (i.e., U2Fusion) is very small. For the metric Q_G, the DTCWT and LPCNN methods obtain higher performance, which is mainly owing to that the MSD framework adopted is good at detail extraction. Our method outperforms two DL-based methods on Q_G. A more detailed sample-wise comparison on the objective performance of different fusion methods is shown in Fig. 3. Overall, the proposed method obtains more competitive performance on objective evaluation. Fig. 4 shows three sets of fusion results obtained by different methods. In comparison to the objective evaluation, the superiority of the proposed method on visual quality in terms of the tumor area is more significant. In fact, we believe that the latter factor is more important to a specific purpose-oriented image fusion problem. It can be seen from Fig. 4 that the fusion results of the four comparison methods can well preserve the pathological information from the T1c modality in T1c, but lose a large amount of lesion information about the edema regions captured by the Flair image, leading to high difficulty to distinguish the whole tumors. By contrast, the fused images obtained by the proposed method can simultaneously capture different categories of tumor pathological information from both T1c and Flair modalities. Therefore, the fusion results of the proposed method have higher potential to benefit the segmentation task as well as physician observation. The average running time of the proposed method for fusing a pair of images of size is 240×240 pixels is about 0.12 seconds with GPU acceleration using the PyTorch framework, which is generally acceptable in practical usage.

Ablation study: We conduct a set of ablation experiments to verify the effect of adding segmentation networks as discriminators in the

Table 1. Objective Assessment Results of Different Fusion Methods

	Q_{MI}	Q_G	Q_W	*VIFF*
DTCWT	0.9961	<u>0.5151</u>	0.7414	0.5639
LPCNN	1.0700	**0.5382**	0.7544	0.5424
U2Fusion	**1.0880**	0.4508	0.7017	0.5966
IFCNN	1.0660	0.4811	<u>0.7815</u>	<u>0.6637</u>
Proposed	<u>1.0868</u>	0.4815	**0.7967**	**0.7727**

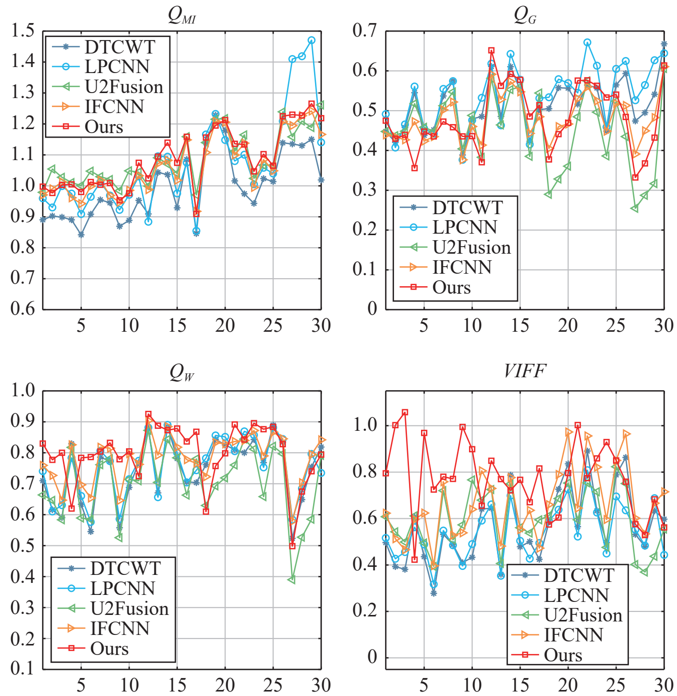
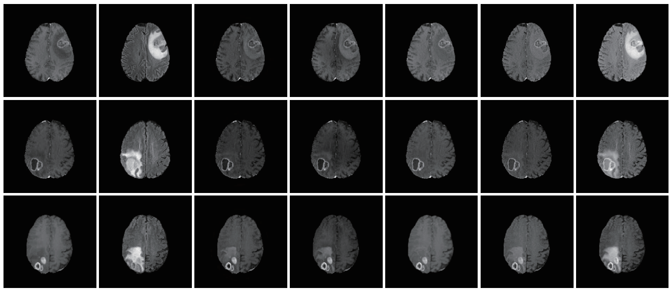


Fig. 3. A more detailed sample-wise comparison on the objective performance of different fusion methods.

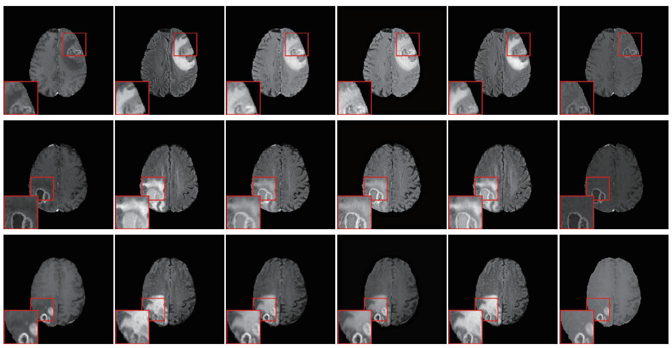
proposed fusion framework. The corresponding results are shown in Fig. 5. The first and second columns are the source images of $T1c$ and Flair modalities. The last four columns give the fusion results obtained by the completed model, the model with two discriminators both removed, the model with the discriminator D_{T1c} removed and the model with the discriminator D_{Flair} removed, respectively. In comparison to the fusion results obtained by the completed model, the results of the model with two discriminators both removed suffer from blurry effect within edema regions and around tumor boundaries. The fusion results of the model with the discriminator D_{T1c} removed are too close to the input Flair images, leading to the loss of pathological information contained in the T1c images. The fusion results of the model with the discriminator D_{Flair} removed almost lose all the edema information captured in the Flair images.

Evaluation from the perspective of segmentation: We finally conduct a segmentation experiment to further evaluate the effectiveness of the proposed fusion method. To exclude the impact of other factors, we only use the fused image as the input of the segmentation model and compare the accuracy using the fused images obtained by different methods. In practice, the fused modality and source modalities can be employed as the segmentation input together to pursue higher accuracy. Considering that the ground truth segmentation label is only provided by the training set in the BraTS dataset, we use the training set as the experimental data here and generate a set of source images from it using the similar approach mentioned above (for simplicity, we mainly focus on 2D image segmentation in this experiment). The generated samples are further divided into training, validation and testing sets with a ratio of 6:2:2. The fused images are then obtained by different fusion methods. For each method, the segmentation model is trained and then the accuracy is tested. The popular U-Net is adopted as the segmentation network. Two popular metrics (the Dice coefficient (DC) and the Hausdorff distance (HD)) are used to evaluate the segmentation accuracy on the whole tumor area. Table 2 lists the segmentation



*T*1*c* Flair DTCWT LPCNN U2Fusion IFCNN Proposed

Fig. 4. Three sets of $T1c$ and Flair image fusion results obtained by different methods.



*T*1*c* Flair Completed model w/o two discriminators w/o discriminator D_{T1c} w/o discriminator D_{Flair}

Fig. 5. Three sets of $T1c$ and Flair image fusion results in the ablation study.

Table 2. The Segmentation Accuracy Using the Fused Images Obtained by Different Fusion Methods

	DTCWT	LPCNN	U2Fusion	IFCNN	Proposed
DC	0.5578	0.5903	0.5006	0.5861	**0.7694**
HD	27.64	30.43	29.32	24.35	**17.51**

accuracy using the fused images obtained by different fusion methods. It can be seen that the fused images of the proposed method achieve much higher segmentation accuracy than those of the other four comparison methods. This observation is in accord with results shown in Fig. 4 , which further verifies the significance of the proposed fusion method.

Conclusions: This letter presents a glioma-oriented multi-modal MR image fusion method. The proposed method is based on an adversarial learning framework, in which the segmentation network is introduced as the discriminators to guide the fusion network to achieve more meaningful results from the viewpoint of tumor segmentation. Experimental results demonstrate the advantage of the proposed method in terms of objective evaluation, visual quality and significance to segmentation. This work is expected to provide some new thoughts to the future study of medical image fusion.

Acknowledgments: This work was supported by the National Natural Science Foundation of China (62176081, 61922075, 6217 1176); the Fundamental Research Funds for the Central Universities (JZ2020HGPA0111, JZ2021HGPA0061); and the USTC Research Funds of the Double First-Class Initiative (KY2100000123).

References

[1] S. Li, X. Kang, L. Fang, J. Hu, and H. Yin, "Pixel-level image fusion: A survey of the state of the art," *Inf. Fusion*, vol. 33, pp. 100–112, 2017.

[2] X. Qu, J. Yan, H. Xiao, and Z. Zhu, "Image fusion algorithm based on spatial frequency-motivated pulse coupled neural networks in nonsubsampled contourlet transform domain," *Acta Automatic Sinica*, vol. 34, no. 12, pp. 1508–1514, 2008.

[3] J. Du, W. Li, and B. Xiao, " Anatomical-functional image fusion by information of interest in local laplacian filtering domain," *IEEE Trans. Image Process.*, vol. 26, no. 12, pp. 5855–5866, 2017.

[4] M. Yin, X. Liu, Y. Liu, and X. Chen, " Medical image fusion with parameter-adaptive pulse coupled neural network in nonsubsampled shearlet transform domain," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 1, pp. 49–64, 2019.

[5] B. Yang and S. Li, " Pixel-level image fusion with simultaneous orthogonal matching pursuit," *Inf. Fusion* , vol. 13, no. 1, pp. 10–19, 2012.

[6] Y. Liu, X. Chen, R. K. Ward, and Z. Jane Wang, "Image fusion with convolutional sparse representation," *IEEE Signal Process. Lett.*, vol. 23, no. 12, pp. 1882–1886, 2016.

[7] Y. Liu, X. Chen, R. K. Ward, and Z. J. Wang, "Medical image fusion via convolutional sparsity based morphological component analysis," *IEEE Signal Process. Lett.*, vol. 26, no. 3, pp. 485–489, 2019.

[8] S. Li and B. Yang, " Hybrid multiresolution method for multisensor multimodal image fusion," *IEEE Sensors J.*, vol. 10, no. 9, pp. 1519–1526, 2010.

[9] Y. Liu, S. Liu, and Z. Wang, "A general framework for image fusion based on multi-scale transform and sparse representation," *Inf. Fusion*, vol. 24, pp. 147–164, 2015.

[10] W. Zhao and H. Lu, " Medical image fusion and denoising with alternating sequential filter and adaptive fractional order total variation," *IEEE Trans. Instrum. Meas.* , vol. 66, no. 9, pp. 2283–2294, 2017.

[11] G. Bhatnagar and Q. M. J. Wu, "A fractal dimension based framework for night vision fusion," *IEEE/CAA J. Autom. Sinica* , vol. 6, no. 1, pp. 220–227, 2019.

[12] Y. Liu, X. Chen, Z. Wang, Z. Wang, R. Ward, and X. Wang, "Deep learning for pixel-level image fusion: Recent advances and future prospects," *Inf. Fusion*, vol. 42, pp. 158–173, 2018.

[13] H. Zhang, H. Xu, X. Tian, J. Jiang, and J. Ma, " Image fusion meets deep learning: A survey and perspective," *Inf. Fusion*, vol. 76, pp. 323–336, 2021.

[14] 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, 2020.

[15] H. Xu and J. Ma, " EMFusion: An unsupervised enhanced medical image fusion network," *Inf. Fusion*, vol. 76, pp. 177–186, 2021.

[16] J. Ma, H. Xu, J. Jiang, X. Mei, and X.-P. Zhang, "DDcGAN: A dual-discriminator conditional generative adversarial network for multi-resolution image fusion," *IEEE Trans. Image Process.*, vol. 29, pp. 4980–4995, 2020.

[17] H. Xu, J. Ma, J. Jiang, X. Guo, and H. Ling, " U2Fusion: A unified unsupervised image fusion network," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, pp. 502–518, 2022.

[18] J. Ma, W. Yu, P. Liang, C. Li, and J. Jiang, "FusionGAN: A generative adversarial network for infrared and visible image fusion," *Inf. Fusion*, vol. 48, pp. 11–26, 2019.

[19] O. Ronneberger, P. Fischer, and T. Brox, " U-Net: Convolutional networks for biomedical image segmentation," in *Proc. MICCAI*, pp. 231–241, 2015.

[20] X. Mao, Q. Li, H. Xie, R. Y. Lau, Z. Wang, and S. Paul Smolley, "Least squares generative adversarial networks," in *Proc. IEEE Int. Conf. Computer Vision*, 2017, pp. 2794–2802.

[21] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, " Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, 2004.

[22] J. J. Lewis, R. J. O'Callaghan, S. G. Nikolov, D. R. Bull, and N. Canagarajah, "Pixel-based and region-based image fusion with complex wavelets," *Inf. Fusion*, vol. 8, no. 2, pp. 119–130, 2007.

[23] Y. Liu, X. Chen, J. Cheng, and H. Peng, " A medical image fusion method based on convolutional neural networks," in *Proc. Int. Conf. Inf. Fusion*, 2017, pp. 1–7.

[24] Z. Liu, E. Blasch, Z. Xue, J. Zhao, R. Laganire, and W. Wu, "Objective assessment of multiresolution image fusion algorithms for context enhancement in night vision: A comparative study," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 1, pp. 94–109, 2012.

[25] N. Cvejic, C. Canagarajah, and D. Bull, "Image fusion metric based on mutual information and Tsallis entropy," *Electronics Lett.* , vol. 42, no. 11, pp. 626–627, 2006.

[26] C. Xydeas and V. Petrovic, " Objective image fusion performance measure," *Electronics Lett.*, vol. 36, no. 4, pp. 308–309, 2000.

[27] G. Piella and H. Heijmans, "A new quality metric for image fusion," in *Proc. IEEE Int. Conf. Image Processing*, 2003, pp. 173–176.

[28] Y. Han, Y. Cai, Y. Cao, and X. Xu, "A new image fusion performance metric based on visual information fidelity," *Inf. Fusion*, vol. 14, no. 2, pp. 127–135, 2013.