Check for updates

Research on the Modality Transfer Method of Brain Imaging Based on Generative Adversarial Network

Dapeng Cheng [1,2]*, Nuan Qiu [1], Feng Zhao [1,2], Yanyan Mao [1,2] and Chengnuo Li [1]

[1] School of Computer Science and Technology, Shandong Technology and Business University, Yantai, China, [2] Shandong Co-Innovation Center of Future Intelligent Computing, Yantai, China

Brain imaging technology is an important means to study brain diseases. The commonly used brain imaging technologies are fMRI and EEG. Clinical practice has shown that although fMRI is superior to EEG in observing the anatomical details of some diseases that are difficult to diagnose, its costs are prohibitive. In particular, more and more patients who use metal implants cannot use this technology. In contrast, EEG technology is easier to implement. Therefore, to break through the limitations of fMRI technology, we propose a brain imaging modality transfer framework, namely BMT-GAN, based on a generative adversarial network. The framework introduces a new non-adversarial loss to reduce the perception and style difference between input and output images. It also realizes the conversion from EEG modality data to fMRI modality data and provides comprehensive reference information of EEG and fMRI for radiologists. Finally, a qualitative and quantitative comparison with the existing GAN-based brain imaging modality transfer approaches demonstrates the superiority of our framework.

Keywords: brain imaging modality transfer, EEG, fMRI, generative adversarial network, non-adversarial loss

1. INTRODUCTION

Brain imaging modality transfer has become popular in the field of medical imaging, providing a variety of reference information for early diagnosis, identification, treatment, and follow-up of the diseases (Yi et al., 2019). For example, the translation of computed tomography (CT) modality data to magnetic resonance (MR) modality data (Jin et al., 2019) solves the problem of lack of MR modality data or synthesizes positron emission tomography (PET) modality data from CT modality data (Choi and Lee, 2018) for cancer staging, detection, and treatment.

The modality transfer of brain imaging is to map one modality to another (Armanious et al., 2020). The information provided by an image obtained from a certain imaging method is often limited, and can only reflect the information of one modality, which generally cannot help the doctor to make an accurate diagnosis. The modality transfer technology is conducive to converting between different modality images to obtain multi-modality information. Combined with multi-modality images, it can provide a variety of information about diseased tissues or organs, and provide a strong theoretical basis for clinical medicine to make an accurate diagnosis. As for the modality transfer of brain imaging, the current research at home and abroad is mainly based on the image-to-image translation. Existing image-to-image translation methods are mainly divided into two categories: sparse representation-based method and learning-based method.

The method based on sparse representation (SR) is the process of using image block prediction. An image block is extracted from an atlas with the same modalities as the source image to sparsely

represent the source image, and then the obtained sparse coefficient is used to estimate the target image block of another modality image. Ye et al. (2013) used T1-weighted MR images to synthesize T2-weighted MR images. Cordier et al. (2016) used brain labeled images to predict FLAIR images. However, the sparse representation-based approach requires sparse coding optimization of image blocks at all locations, which will lead to a decrease in the efficiency of the prediction process, and sometimes there will be noise in the synthesis results.

The convolutional neural network (CNN) in the "two networks" shows good performance in the fields of computer vision and medical image analysis. Li et al. (2014) used CNN to predict PET images from MR. Han (2017) used deep CNN to synthesize CT with MR. But, the training of a robust CNN model requires a large data set and a long training time. However, there is an unavoidable problem in the CNN-based image modality migration task. Although the entire network training process is automatic, it still needs to design loss function, namely, design loss network parameters.

Since the GAN concept was proposed, it has been showing strong practicability in image synthesis. The work of Denton et al. (2015) and Salimans et al. (2016) has proved that GAN has been showing strong practicability in the image. GAN can solve the problems encountered in CNN well. According to the work of Zhao et al. (2016), it is proved that when using CNN, a specific loss function is needed to deal with the problem, and GAN can achieve good results. At the same time, a new GAN network structure—Conditional GANs was proposed, referring to the design pattern of the CNN network and adding Settings for specific applications when GAN was used. A lot of work confirmed that such a GAN network has a good effect in processing images, videos, and 3D data. Isola et al. (2017) have proposed an image translation structure Pix2Pix based on conditional GAN, which can achieve good results under supervised learning.

However, a network based on supervised learning requires paired data training, and patients should be tested for both EEG and fMRI at short intervals. Besides, further, post-processing is required, such as image registration (Yang et al., 2018). Obtaining these data is a great challenge. Whatsmore, with the development of computer technology and digital imaging technology, brain imaging methods are gradually mature and diversified, yet many of these technologies still have limitations. For example, functional magnetic resonance imaging types of equipment are expensive, and patients who use pacemakers cannot use the fMRI. Although fMRI can accurately locate the active area when "observing the active brain" and its spatial resolution can reach the millimeter level (Tang et al., 2019), its temporal resolution is far lower than EEG (Menon et al., 1998). In contrast, EEG is easier to implement. Thus, if we can realize the conversion from EEG to fMRI, we can obtain both advantages of EEG and fMRI by implementing only one of them. However, to our best knowledge, few studies have explored the conversion of EEG to fMRI modality data.

In this work, we propose a brain imaging modality transfer framework, which uses 2D sensor cap images of EEG and T1-weighted axial images from fMRI to infer the target fMRI modality data, providing comprehensive reference information for medical diagnosis. Specifically, we focus on generating fMRI modality data from EEG modality data, which requires learning the mapping relationship between these two different modality data. However, it is well-known that converting between different modality data has always been a challenge. Fortunately, the emergence of the generative adversarial network (GAN) (Goodfellow et al., 2014) improves this aspect. GAN has achieved advanced performance in learning the mapping between different modalities. Meanwhile, with the development of deep learning, GAN has become one of the hot research directions in the field of medical imaging.

Therefore, we leverage GAN to achieve a brain imaging modality transfer framework named BMT-GAN. This framework introduces the combination of cycle-consistency loss and adversarial loss used in the CycleGAN framework proposed by Wolterink et al. (2017). which avoids the inability of traditional GAN to achieve pairing problems between input and output. Besides, the framework introduces non-adversarial loss, which reduces the perception and style differences between input and output modality images, and enhances the global consistency between input and output images. A qualitative and quantitative comparison with the existing GAN-based medical image modality transfer methods proves that BMT-GAN can effectively capture the overall tissue contrast and local anatomical details of fMRI images.

The contributions of this article are summarized as follows:

(1) We proposed a new framework named BMT-GAN to estimate two-dimensional fMRI images using two-dimensional EEG images;

(2) BMT-GAN framework combines the cycle-consistent loss, adversarial loss, and non-adversarial loss to achieve excellent brain imaging modality transfer performance;

(3) The proposed approach can be easily extended to other medical data translation tasks to benefit the medical imaging field.
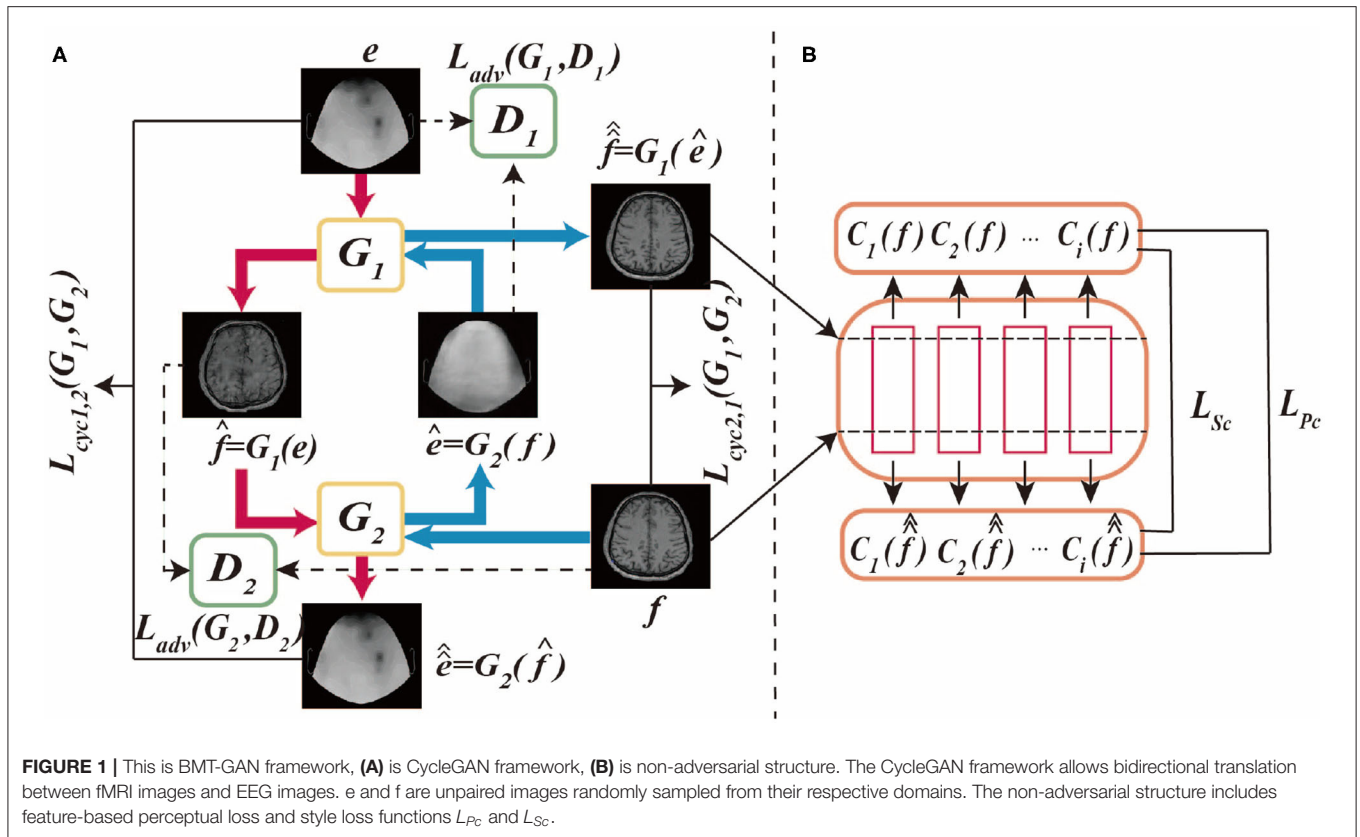
2. METHODS

The proposed BMT-GAN framework utilizes the dual thought (Farnia and Tse, 2018) of Welander et al. (2018), which not only adopts the cycle-consistency loss and adversarial loss of CycleGAN model but also introduces a new non-adversarial loss combination. Whatsmore, **Figures 1A,B** shows the BMT-GAN framework structure of two parts: CycleGAN and non-adversarial structure.

2.1. CycleGAN

CycleGAN is an unsupervised framework that allows image-to-image translation without matching training data samples in two domains. The network framework structure and data flow diagram are shown in **Figure 1A**.

As shown in **Figure 1A**, the differences between CycleGAN framework and traditional GAN are: (1) CycleGAN consists of two generators (G_1, G_2) and two discriminators (D_1, D_2); (2) CycleGAN instead of learning a single mapping, it learns two mapping functions between two domains: $G_1 : I_{EEG} \longrightarrow$

FIGURE 1 | This is BMT-GAN framework, **(A)** is CycleGAN framework, **(B)** is non-adversarial structure. The CycleGAN framework allows bidirectional translation between fMRI images and EEG images. e and f are unpaired images randomly sampled from their respective domains. The non-adversarial structure includes feature-based perceptual loss and style loss functions L_{Pc} and L_{Sc}.

I_{fMRI} and $G_2 : I_{fMRI} \longrightarrow I_{EEG}$, ensuring that the input and output images can be matched after translation. Where, I_{EEG} is the sample space of the given EEG image training data, $e \in I_{EEG}$, I_{fMRI} is the sample space of the given fMRI image training data, $f \in I_{fMRI}$.

The framework of CycleGAN consists of two major branches $Cyc_{1,2}$ and $Cyc_{2,1}$. In the generation network of the $Cyc_{1,2}$ branch, the generator G_1 takes the EEG images in the source domain I_{EEG} as input, and outputs the synthetic translation $\hat{f} = G_1(e)$. To prevent a group of different images from being mapped to a single image in the target domain, the model requires the two generators to keep the loop consistent with each other. Therefore, using $\hat{f}$ as the input of G_2, the output reconstruction $\hat{\hat{e}} = G_2(\hat{f})$ is obtained. In the $Cyc_{2,1}$ branch, the discriminator uses the fMRI images of the target domain I_{fMRI} as the input of G_2, and the synthesized translation $\hat{e} = G_2(f)$. In the same way, using $\hat{e}$ as the input of G_1 to get the output reconstruction $\hat{\hat{f}} = G_1(\hat{e})$. Therefore, the cycle-consistent loss of the two branches can be formulated as:

$$
\begin{aligned}
L_{Cyc_{1,2}}(G_1, G_2) &= E_{e \in I_{EEG}}[||G_2(G_1(e)) - e||_1] \\
L_{Cyc_{2,1}}(G_2, G_1) &= E_{f \in I_{fMRI}}[||G_1(G_2(f)) - f||_1]
\end{aligned}
\tag{1}
$$

Also, each generator uses a corresponding discriminator for adversarial training. In the discriminator of the $Cyc_{1,2}$ branch, the fMRI image in the target domain I_{fMRI} and the synthetic

translation $\hat{f}$ are used as the D_1 input of the discriminator. In the $Cyc_{2,1}$ branch, the EEG image in the source domain I_{EEG} and the synthetic translation $\hat{e}$ from the source domain are used as the input of D_2. Among them, D_1 and D_2 act as binary classifiers to distinguish the converted image from the target domain image as accurately as possible. Besides, the generator synthesizes high-quality images through iterative training to obfuscate the discriminator. Therefore, the model's adversarial loss function can be formulated as:

$$
\begin{aligned}
L_{Adv_{1,2}}(G_1, D_1) &= E_{f \in I_{fMRI}}[logD_1(f)] \\
&\quad + E_{e \in I_{EEG}}[log(1 - D_1(G_1(e)))] \\
L_{Adv_{2,1}}(G_2, D_2) &= E_{e \in I_{EEG}}[logD_2(e)] \\
&\quad + E_{f \in I_{fMRI}}[log(1 - D_2(G_2(f)))]
\end{aligned}
\tag{2}
$$

2.2. Non-adversarial Structure

In order to ensure that the input and output images can be paired after image translation, the loss functions in Equations (1) and (2) are introduced. However, during the training process, due to the large differences in the anatomical details of the 2D EEG image and the 2D fMRI image, the pixel-level loss becomes ineffective and the output translation image lacks clarity and fine structure (Johnson et al., 2016; Wang C. et al., 2018). For example, Jin et al. (2019) used a combination of adversarial loss and the traditional L1 distance loss. This loss combination has achieved good results in the translation task of paired images. Armanious et al. (2020)

used a combination of perceptual loss and style loss in addition to cycle-consistent and adversarial loss in the MedGAN framework, and also achieved success.

However, for unpaired image translation tasks, it is not feasible to use the above two-loss combinations due to the huge difference between the source domain image and the target domain image. Therefore, we introduce a feature-based loss function between input reference and output reconstruction as an additional constraint to improve output quality. The process is shown in **Figure 1B**. The first proposed loss function is the cycle-perceptual loss, L_{Pc}. This is achieved by extracting intermediate feature maps using a pre-trained feature extractor network, for both the input and the cycle-reconstructed images. This process is formulated as a non-adversarial perceptual loss function L_{Pc}:

$$L_{Pc} = \sum_{i=0}^{N} \omega_{pc,i}(||C_i(e) - C_i(\hat{e})||_1 + ||C_i(f) - C_i(\hat{\hat{f}})||_1) \quad (3)$$

Where C_i is the feature map extracted by the *ith* layer of the feature extractor network. N is the total number of layers, and $\omega_{pc,i}$ is the weight of each layer.

Secondly, a style-based loss function between input reference and output reconstruction is proposed to compensate for the difference in style representation between the reconstructed image and its corresponding target image. By calculating the relevance of the feature representation in the spatial range, the style distribution can be captured. To calculate the style loss, we must first calculate $G_{r_i}(e)$:

$$G_{r_i}(e)_{m,n} = \frac{1}{h_i, w_i, d_i} \sum_{h=1}^{h_i} \sum_{w=1}^{w_i} C_i(e)_{h,w,m} C_i(e)_{h,w,m} \quad (4)$$

Where $G_{r_i}(e)$ is the Gram matrix of each convolution block, the shape is $d_i d_i$, which is used to express the feature association. Its elements are obtained by inner product calculation on the height and width of the feature graph. h_i, w_i, d_i the height, width, and depth of the feature space. The non-adversarial style loss function is formulated as L_{Sc}:

$$L_{Sc} = \sum_{i=0}^{N} \omega_{sc,i} \frac{1}{4d_i^2}(||G_{r_i}(e) - G_{r_i}(\hat{e})||_F^2 + ||G_{r_i}(f) - G_{r_i}(\hat{\hat{f}})||_F^2) \quad (5)$$

$\omega_{sc,i}$ is the ith weight of the given Gram matrix.

Therefore, the ultimate minimum and maximum optimization task of the BMT-GAN framework is formulated as follows:

$$\min_{G_1,G_2} \max_{D_1,D_2} L = L_{Cyc_{1,2}}(G_1, G_2) + L_{Cyc_{2,1}}(G_2, G_1) + L_{Adv_{1,2}}(G_1, D_1)$$

$$+ L_{Adv_{2,1}}(G_2, D_2) + \omega_{pc}L_{Pc} + \omega_{sc}L_{Sc}$$

$$(6)$$

2.3. Generator and Discriminator

The generator network is composed of three components: encoder, converter, and decoder, while the discriminator network is composed of multiple convolutional layers, the flow diagram as shown in **Figures 2A,B**.

2.3.1. Generator Network

(1) The encoder is composed of three convolutional layers. The initial size of the input image is $256 \times 256 \times 1$. The first step is to extract features from the image through the convolutional layer. The number of features extracted from the convolutional layer can also be regarded as the number of different filters used to extract different features. The convolutional layer gradually extracts more advanced features in turn. After passing through the encoder, the input image changes from (256,256,1) to (64,64,128).

(2) The converter is composed of nine residual blocks. Different features of the image are combined by different channels of the output image of the encoder. According to these features, the feature vector of the image is converted from the source domain to the target domain. However, since the images in the EEG domain and fMRI domain do not have similar features, the feature vectors of images in the EEG domain should be converted into feature vectors in the fMRI domain through the sharing function of the generator during the conversion process. The residual block consists of two convolutional layers and the input residuals are added to the output to ensure that the input properties of the previous layer can also be applied to the later layers so that their output will not be different from the original input. Otherwise, the features of the original image will not be retained in the output.

(3) The role of a decoder is to reconstruct low-level features from feature vectors, which can be accomplished by using a deconvolution layer. Finally, the low-level functions are converted to images in the target domain.

2.3.2. Discriminator Network

The discriminator takes an image as input and predict whether the image is the original image or the output of the generator. The discriminator is composed of multiple convolutional layers. After extracting features from the image, the discriminator can determine whether these features belong to a specific category. The last layer of the discriminator network is the convolutional layer used to generate one-dimensional output.
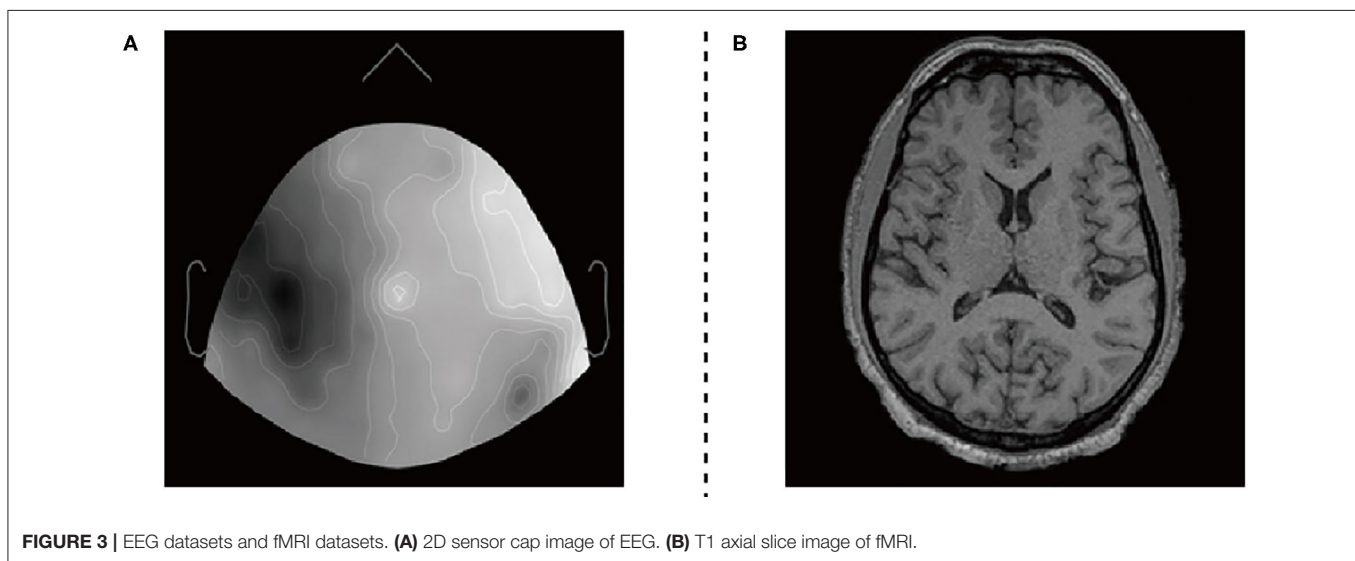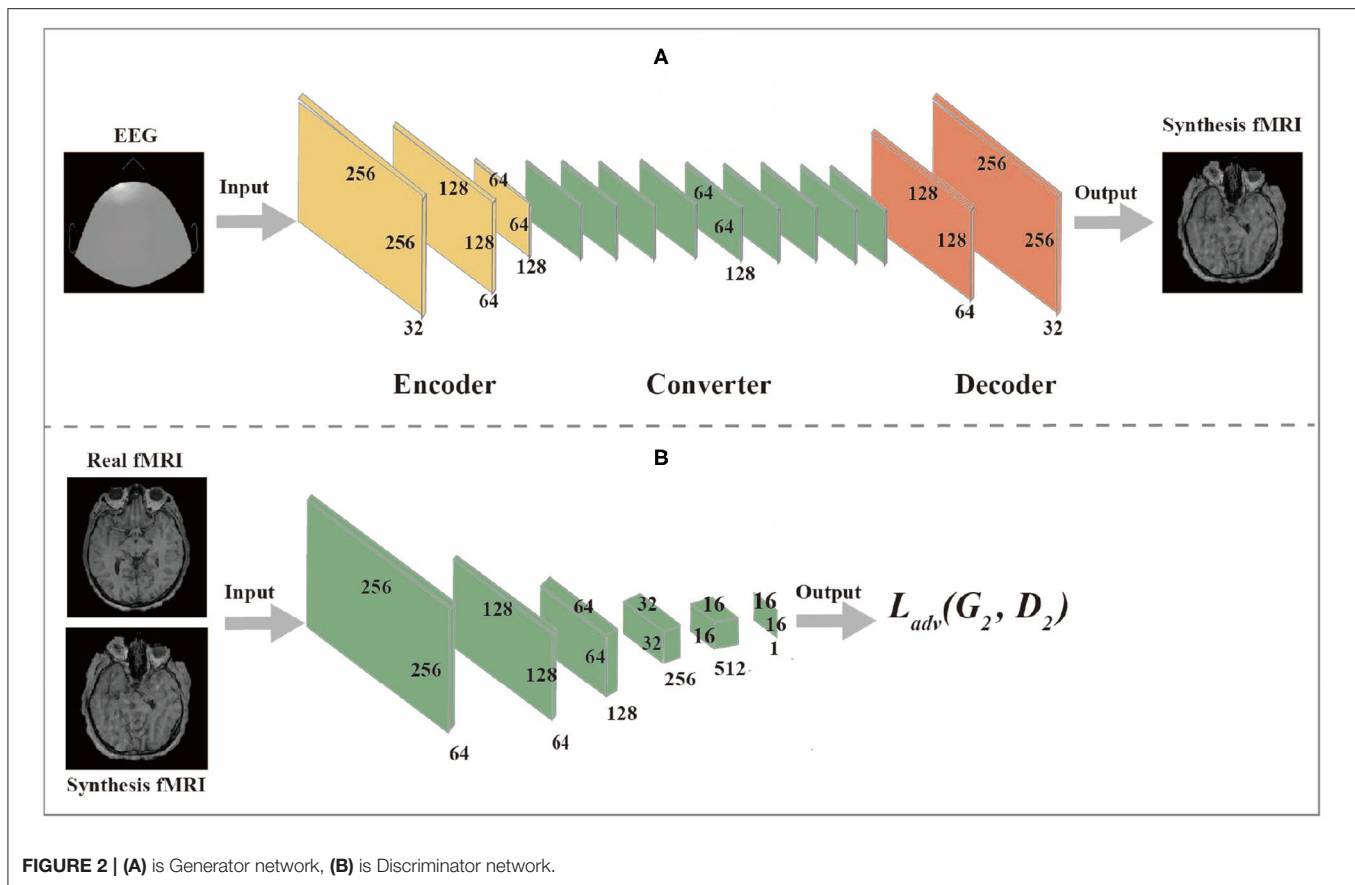
3. EXPERIMENT AND RESULTS

We compared the BMT-GAN framework with several state-of-the-art brain imaging modality translation approaches. In this section, we will describe the dataset, model performance evaluation, evaluation metrics, qualitative, and quantitative results to demonstrate the effectiveness of the brain imaging modality migration approach.
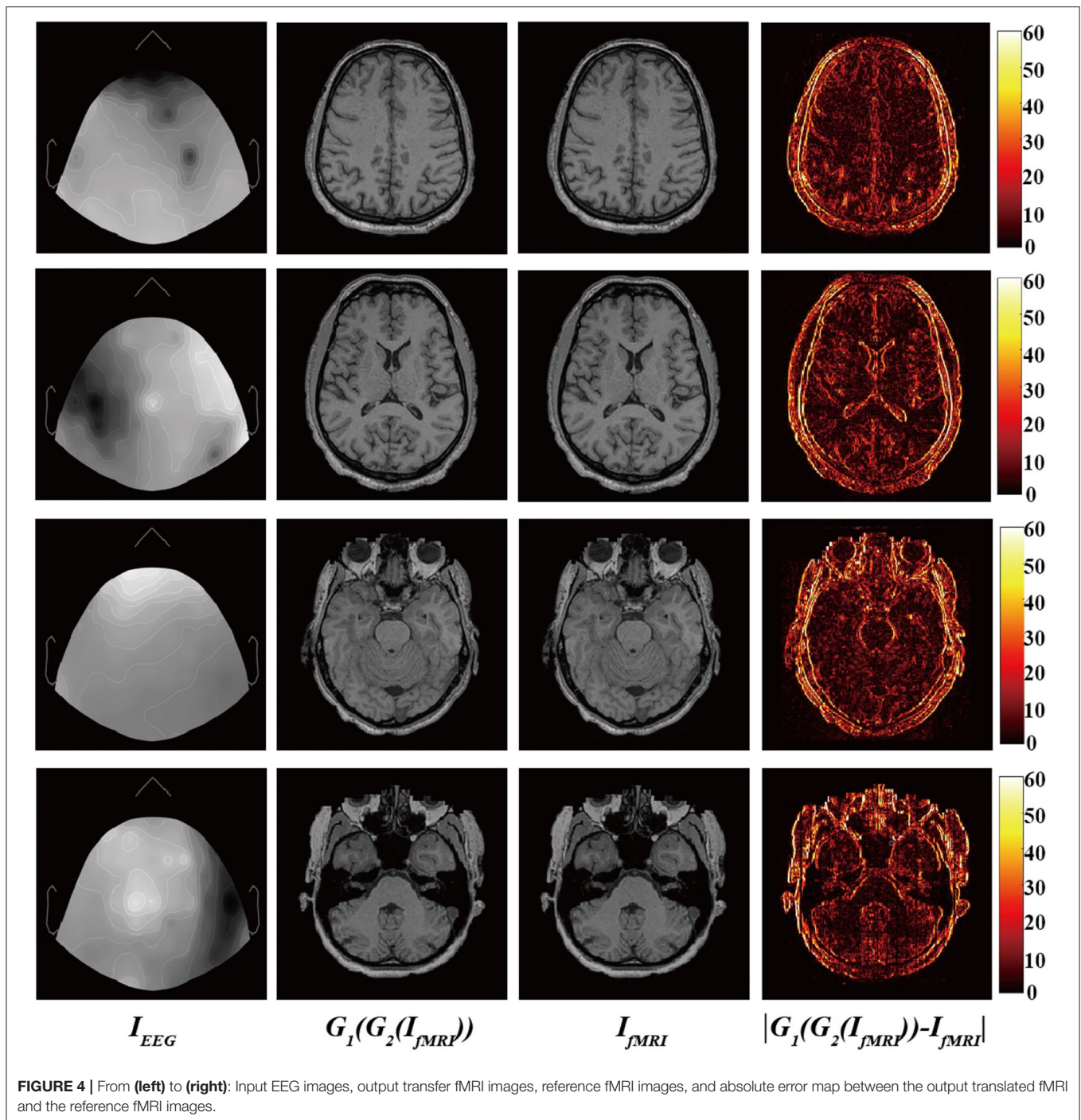
3.1. Datasets

In our work, we applied the two-dimensional sensor cap image of EEG and the two-dimensional axial slice of the high-resolution T1-weighted image of fMRI. The utilized datasets are illustrated in **Figure 3**.

The EEG and fMRI imaging data from 17 adult volunteers. They were acquired in a Siemens Avanto 1.5T clinical scanner using a self-shielded gradient set with maximum gradient amplitude of $40 m^T m^{-1}$, derived from Deligianni et al. (2014).

FIGURE 2 | **(A)** is Generator network, **(B)** is Discriminator network.

FIGURE 3 | EEG datasets and fMRI datasets. **(A)** 2D sensor cap image of EEG. **(B)** T1 axial slice image of fMRI.

The fMRI imaging data acquisition was based on a $T2*$-weighted gradient-echo EPI sequence with 300 volumes, and the effective voxel size is $3.3 \times 3.3 \times 4.0mm^3$. Scalp EEG was recorded during the MRI scan using a 64-channel MR-compatible electrode cap (BrainCap MR, Gilching, Germany) at a native frequency of 1,000 Hz.

The two types of imaging data were visualized by using Brainstorm visualization software. Each volunteer involved 35 or more EEG and fMRI two-dimensional axial slices. They have pixels of the same size as $256 \times 256 \times 1$. We divide the datasets into two training sets and two test sets. The training set includes EEG images and fMRI images of 12 subjects, and the test set

FIGURE 4 | From **(left)** to **(right)**: Input EEG images, output transfer fMRI images, reference fMRI images, and absolute error map between the output translated fMRI and the reference fMRI images.

includes EEG images and fMRI images of the remaining subjects. Secondly, we used the Intel(R) Core(TM) i7-9700 CPU processor for $2e^5$ iterations. The entire training process takes about 40 h.

Finally, **Figure 4** shows the input EEG image, the output fMRI image after modality transfer, the target fMRI image, and the absolute difference map between the output fMRI and the reference fMRI image.

It is worth mentioning that we narrowed the original difference range to [0,60] to make the absolute difference clearer and easier to analyze. Observed from local areas, there are still differences in some local areas. The biggest difference appears in the area of bone structure, while the smallest difference appears in the soft tissue area of brain tissue. This part may be caused by the misregistration between EEG and fMRI images and the

large difference in image structure. From a holistic view, BMT-GAN has successfully learned different anatomical structures in fMRI images (such as bone, gyrus, and soft tissue of brain tissue) in EEG images, and learned to distinguish different anatomical structures with similar pixel intensity. In addition, BMT-GAN also implements the translation of EEG images to fMRI images and ensures the global consistency of the output fMRI images to the reference fMRI images.

3.2. State-of-the-Art Works

We compared BMT-GAN with the following methods:

(1) UNIT (Wang T.-C. et al., 2018): A method of using latent shared space for image transfer.

(2) CGAN (Mirza and Osindero, 2014; Bayramoglu et al., 2017): A method to generate the specified image. It adds additional information, such as labels, to the input of the original GAN generator and discriminator.
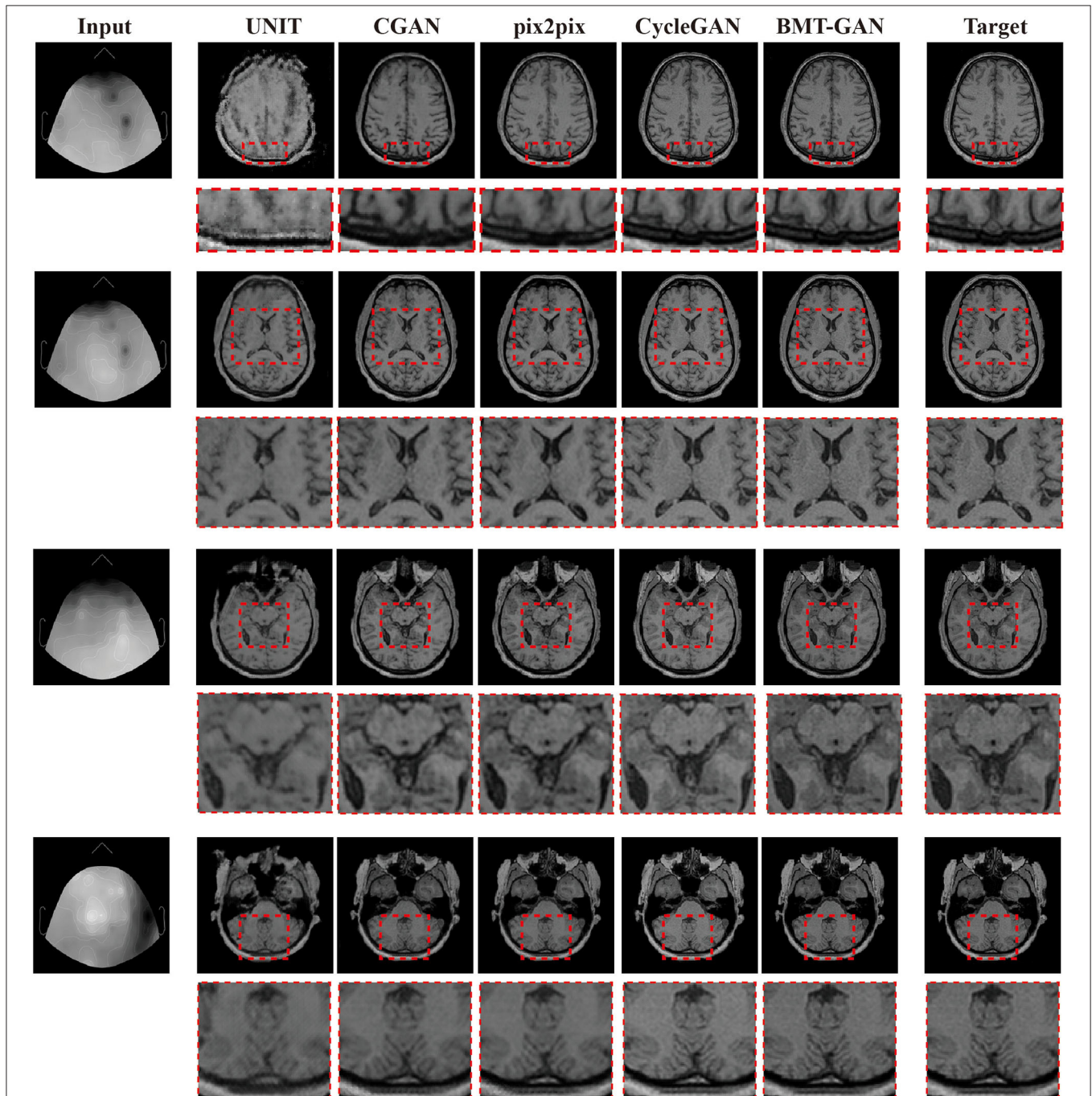

FIGURE 5 | Qualitative comparisons between the BMT-GAN framework and other brain imaging modality transfer technologies. Describes the task of converting EEG to fMRI.

TABLE 1 | Quantitative comparison of brain imaging modality transfer technologies.

Method	MSE	PSNR	SSIM	VIF	IFC
UNIT	1.0230e+03	18.0322	0.5705	0.0465	0.2733
CGAN	340.4902	22.8098	0.6783	0.1614	0.9207
pix2pix	241.0907	24.3090	0.8511	0.3392	2.3423
CycleGAN	143.9875	26.5476	0.8092	0.3323	2.1211
BMT-GAN	128.6233	27.0376	0.8627	0.3575	2.4794

(3) pix2pix (Choi and Lee, 2018; Olut et al., 2018): A method uses a combination of L1 distance and antagonistic loss to train paired data.

(4) CycleGAN (Dar et al., 2019): This method learns two mappings, and introduces two Cycle-consistent loss and adversarial loss to regularize the mapping.

3.3. Evaluation Metrics

We use three commonly used image quality evaluation metrics: Mean-square error (MSE) (Chai and Draxler, 2014), Peak signal-to-noise ratio (PSNR) (De Boer et al., 2003), and Structural similarity (SSIM) (Wang et al., 2004). Their definitions are as follows:

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |(I_{fMRI}(i) - G_1(I_{EEG}(i)))| \qquad (7)$$

Where it is assumed that the number of 2D slices referred to in the fMRI image is N, i is the index of the image slices, and MAE is the measurement of the average distance between each pixel of the composite image and the real image. The PSNR formula is as follows:

$$MSE = \frac{1}{N} \sum_{i=1}^{N} (I_{fMRI}(i) - G_1(I_{EEG}(i)))^2 \qquad (8)$$

$$PSNR = 10 * log_{10} \left(\frac{MAX_{I_{fMRI}}^2}{MSE} \right) \qquad (9)$$

Where MAX is the maximum intensity value of the original image and the composite image. Since each pixel of the resulting composite image and the original image is represented by an 8-bit binary, $MAX = 255$. Both PSNR and MAE can be used to compare the performance of different models over a unified dataset, but they are limited only to image alignment. Therefore, we also need the structural similarity (SSIM) index, which can not only measure the pixel brightness difference of the image but also evaluate the contrast and structure of the image. Its formula is defined as:

$$\begin{cases} L(x,y)_i = \frac{2\mu_{i,x}\mu_{i,y}+C_1}{\mu_{i,x}^2+\mu_{i,y}^2+C_1} \\ C(x,y)_i = \frac{2\sigma_{i,x}\sigma_{i,y}+C_2}{\sigma_{i,x}^2+\sigma_{i,y}^2+C_2} \\ S(x,y)_i = \frac{\sigma_{i,xy}+C_3}{\sigma_{i,x}\sigma_{i,y}+C_3} \end{cases} \qquad (10)$$

$$SSIM(x,y)_i = [L(x,y)_i^\alpha * C(x,y)_i^\beta * S(x,y)_i^\gamma] \qquad (11)$$

Set α, β, γ to 1, you can get:

$$SSIM(x,y)_i = \frac{(2\mu_{i,x}\mu_{i,y} + C_1)(2\sigma_{i,x}\sigma_{i,y} + C_2)}{(\mu_{i,x}^2 + \mu_{i,y}^2 + C_1)(\sigma_{i,x}^2 + \sigma_{i,y}^2 + C_2)} \qquad (12)$$

Where, $\mu_{i,x}\mu_{i,y}$ are the mean value of the synthetic image and the real image, respectively, and $\sigma_{i,x}\sigma_{i,y}$ are the synthetic image and the real image respectively The variance of the image, $\sigma_{i,xy}$ is the covariance between the synthetic image and the real image, C_1, C_2, C_3 are constants to avoid system errors caused by a denominator of 0.
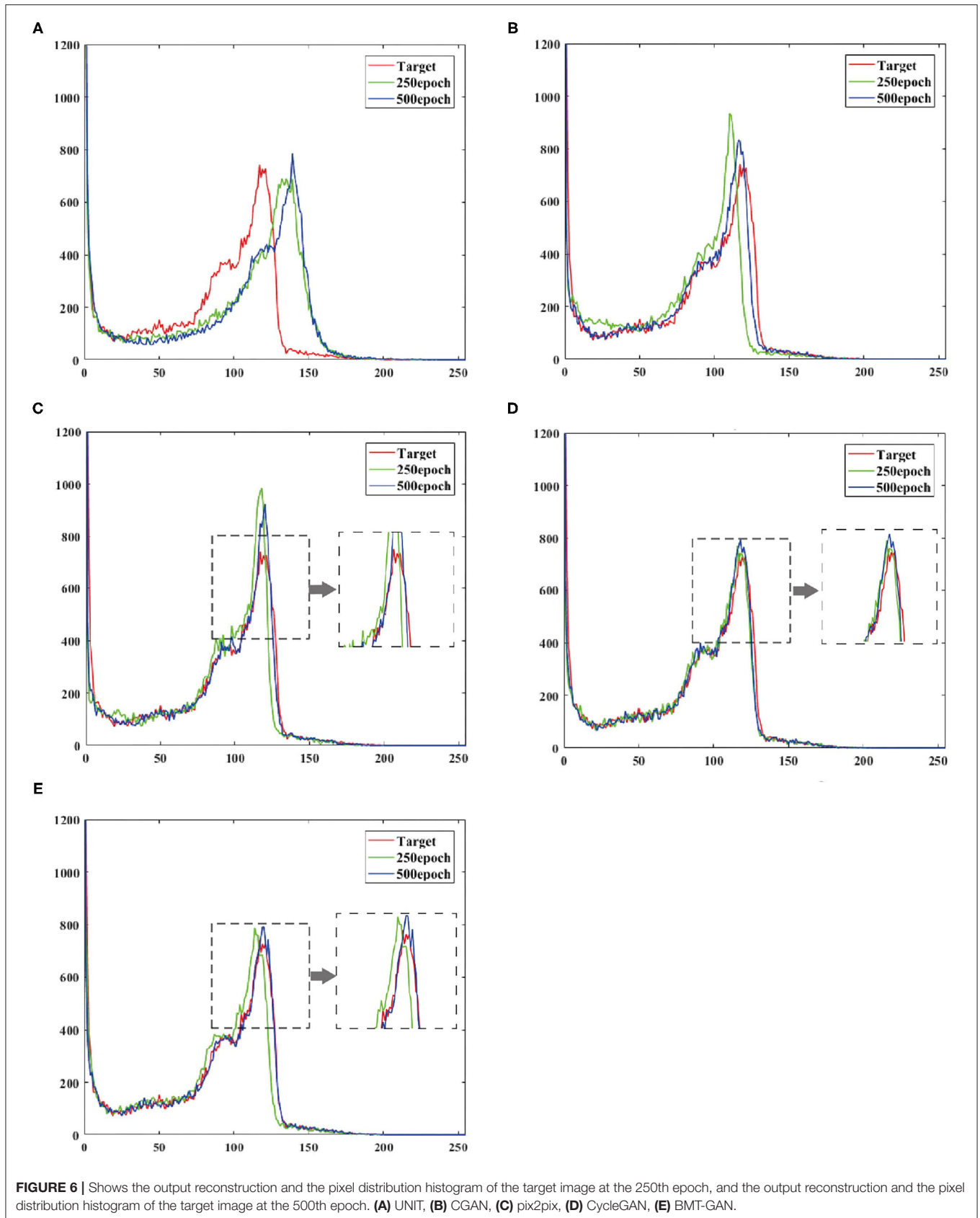
The above three evaluation metrics measure the quality of the image from a statistical perspective by calculating the pixel error between the output image and the target image. These three evaluation metrics are relatively simple and easy to implement, but they do not take into account the local visual factors of human eyes, so there is no way to grasp the local quality of the image. Recent research has shown that Visual Information Fidelity (VIF) (Sheikh and Bovik, 2006) and Information Fidelity Criterion (IFC) (Sheikh et al., 2006) are widely used to evaluate image quality. They extend the link between the image and the human eyes in terms of fidelity of information. Therefore, we introduce these two valuation metrics, which make image similarity measurement more effective.

3.4. Result Analysis

The proposed BMT-GAN framework in this paper is experimentally verified. Compared with the traditional pix2pix framework, CGAN framework in the supervised field, and the famous CycleGAN, UNIT frameworks in the unsupervised field to prove the superiority of our proposed framework. Through the final five evaluation metrics and convergence performance to compare the similarity of pixel distribution histogram indicators to prove the feasibility and efficiency of BMT-GAN.

3.4.1. Prediction Performance

The experiment results are obtained by using target fMRI and output fMRI images as main examples. During training, the training data are completely consistent, and the same input is used for verification. The prediction results are shown in **Figure 5**. The UNIT framework completes the image transformation between the two domains, but distortion occurs in some areas of the image resulting in the blurred output of the translated image. Qualitatively, pix2pix and CGAN frameworks have also successfully achieved image conversion between two domains. It shows better performance than UNIT but lacks

FIGURE 6 | Shows the output reconstruction and the pixel distribution histogram of the target image at the 250th epoch, and the output reconstruction and the pixel distribution histogram of the target image at the 500th epoch. **(A)** UNIT, **(B)** CGAN, **(C)** pix2pix, **(D)** CycleGAN, **(E)** BMT-GAN.

high resolution. The BMT-GAN framework and CycleGAN framework in this paper both have good results. However, the proposed BMT-GAN framework yields results that are closer to the ground truth, more closely connected between output, and input, twisted less and results not only on the style transformation is more apparent. The results show that the framework proposed in this paper has a good application value in the field of image translation.

In order to more accurately verify the performance of the proposed model. We introduce MSE, PSNR, SSIM, and other property price indicators to measure the output results through these numerical scoring standards. The specific results are shown in **Table 1**.

Table 1 shows the MSE, PSNR, and other indicators values computed based on the predicted images with respect to the ground truth. Specifically, BMT-GAN again demonstrates the best performance among all the compared methods. BMT-GAN improves the state-of-the-art PSNR/SSIM performance from 26.5476/0.8092 given by CycleGAN to 27.0376/0.8627. However, as analyzed before, UNIT combined with traditional GAN networks and variational autoencoders to improve the output results, although very good results can be achieved, because two sets of GAN networks are trained at the same time, each

step of training requires parameter updates for the two sets of GAN networks, resulting in higher training costs Large, the convergence rate is slower.

3.4.2. Convergence Performance

To verify the effect of convergence speed on the model, we divided the training period of the model into 250 epochs and 500 epochs, and obtained the pixel distribution histogram of the reconstructed image and the target image at different epochs. Observe the distance between the pixel distribution histogram of different epochs and the target pixel distribution histogram. The size of the distance indicates how fast the model converges, as shown in **Figure 6**.

At the 250th epoch, we can see that the UNIT model has the slowest convergence speed, and the CycleGAN model has the fastest convergence speed. But because the histogram of the 500 epoch does not give clear results. Therefore, we use two commonly used evaluation indicators: Euclidean metric (EM) and cosine similarity (CS) to compare the similarity between the output reconstructed image of the five models and the pixel distribution histogram of the target image.
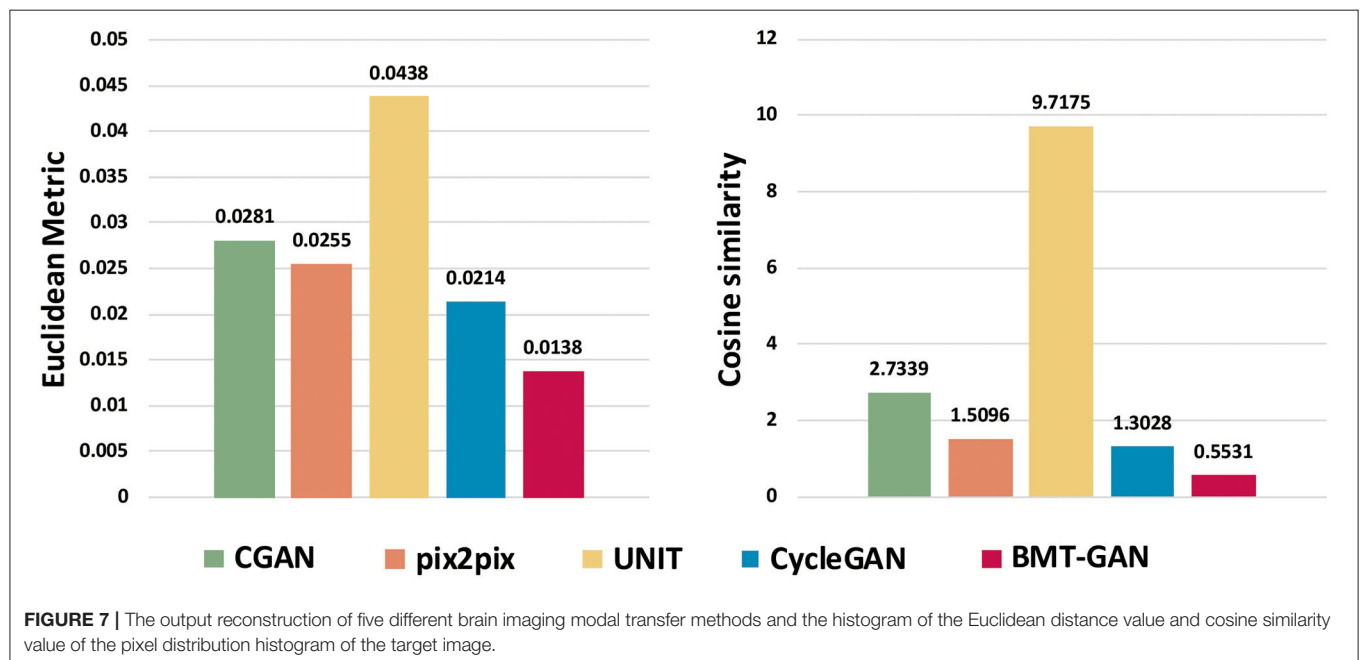
Table 2 and **Figure 7** show that the model proposed in this paper is closest to the target value. Compared with other models, the BMT-GAN framework proposed in this paper can more effectively improve the performance of image translation tasks and enhance the relationship between output and input.

4. CONCLUSION

This paper proposes a novel brain imaging modality transfer framework, namely BMT-GAN. The proposed BMT-GAN framework introduces the adversarial loss of discriminator network, dual-period consistent loss of Unpaired training data

TABLE 2 | Quantitative comparison of brain imaging modality transfer technologies.

Method	UNIT	CGAN	pix2pix	CycleGAN	BMT-GAN
EM	0.0438	0.0281	0.0255	0.0214	0.0138
CS	9.7175	2.7339	1.5096	1.3028	0.5531

FIGURE 7 | The output reconstruction of five different brain imaging modal transfer methods and the histogram of the Euclidean distance value and cosine similarity value of the pixel distribution histogram of the target image.

and the feature-based non-adversarial loss combination to realize the modality transformation from EEG image to fMRI image. Specifically, the traditional modality transfer methods of brain imaging are based on pixel loss, which usually leads to blurring results. In the field of medical images, small structures can significantly change the diagnostic information of images. Therefore, to capture the difference between the high-frequency components in the image, we introduce non-adversarial perception and style loss based on CycleGAN. This application can be said to be novel. It makes full use of the complementary information of different modality images and can synthesize fMRI images with better tissue contrast and anatomical details, which improves the diagnostic value of brain imaging technology in clinical medicine.

The qualitative and quantitative results show that the performance of BMT-GAN is superior to the existing methods of brain imaging modality transfer. Specifically, the output fMRI images were closer to the target, and the lowest MSE (128.6233), the highest PSNR (27.0367), and the highest SSIM (0.8627) were obtained. However, the deviation between EEG images and reference fMRI images may have a significant impact on quantitative evaluation. Although quantitative measurement is a standard to evaluate the performance of a method, in this case, the numerical difference in quantitative evaluation cannot correctly represent the quality difference. In future work, the accuracy of synthetic fMRI images based on medical expert perception studies will be evaluated.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

REFERENCES

Armanious, K., Jiang, C., Fischer, M., Küstner, T., Hepp, T., Nikolaou, K., et al. (2020). Medgan: Medical image translation using gans. *Comput. Med. Imaging Graph.* 79:101684. doi: 10.1016/j.compmedimag.2019.101684

Bayramoglu, N., Kaakinen, M., Eklund, L., and Heikkila, J. (2017). "Towards virtual h&e staining of hyperspectral lung histology images using conditional generative adversarial networks," in *Proceedings of the IEEE International Conference on Computer Vision Workshops* (Venice), 64–71. doi: 10.1109/ICCVW.2017.15

Chai, T., and Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)?-arguments against avoiding RMSE in the literature. *Geosci. Model Dev.* 7, 1247–1250. doi: 10.5194/gmd-7-1247-2014

Choi, H., and Lee, D. S. (2018). Generation of structural MR images from amyloid pet: application to mr-less quantification. *J. Nucl. Med.* 59, 1111–1117. doi: 10.2967/jnumed.117.199414

Cordier, N., Delingette, H., Lê, M., and Ayache, N. (2016). Extended modality propagation: Image synthesis of pathological cases. *IEEE Trans. Med. Imag.* 35, 2598–2608. doi: 10.1109/TMI.2016.2589760

Dar, S. U., Yurt, M., Karacan, L., Erdem, A., Erdem, E., and Çukur, T. (2019). Image synthesis in multi-contrast MRI with conditional generative adversarial networks. *IEEE Trans. Med. Imag.* 38, 2375–2388. doi: 10.1109/TMI.2019.2901750

De Boer, J. F., Cense, B., Park, B. H., Pierce, M. C., Tearney, G. J., and Bouma, B. E. (2003). Improved signal-to-noise ratio in spectral-domain compared with time-domain optical coherence tomography. *Opt. Lett.* 28, 2067–2069. doi: 10.1364/OL.28.002067

Deligianni, F., Centeno, M., Carmichael, D. W., and Clayden, J. D. (2014). Relating resting-state fMRI and EEG whole-brain connectomes across frequency bands. *Front. Neurosci.* 8:258. doi: 10.3389/fnins.2014.00258

Denton, E. L., Chintala, S., szlam, a., and Fergus, R. (2015). "Deep generative image models using a laplacian pyramid of adversarial networks," in *Advances in Neural Information Processing Systems* (Montreal, QC), 1486–1494.

Farnia, F., and Tse, D. (2018). "A convex duality framework for gans," in *Advances in Neural Information Processing Systems, Vol. 31* (Montreal, QC: Curran Associates, Inc.), 5248–5258.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2014). "Generative adversarial nets," in *Advances in Neural Information Processing Systems, Vol. 27* (Cambridge), 2672–2680.

Han, X. (2017). MR-based synthetic ct generation using a deep convolutional neural network method. *Med. Phys.* 44, 1408–1419. doi: 10.1002/mp.12155

Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A. (2017). "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Honolulu, HI). doi: 10.1109/CVPR.2017.632

Jin, C.-B., Kim, H., Liu, M., Han, I. H., Lee, J. I., Lee, J. H., et al. (2019). Dc2anet: Generating lumbar spine MR images from CT scan data based on semi-supervised learning. *Appl. Sci.* 9:2521. doi: 10.3390/app9122521

Johnson, J., Alahi, A., and Fei-Fei, L. (2016). "Perceptual losses for real-time style transfer and super-resolution," in *Computer Vision-ECCV 2016* (Amsterdam), 694–711. doi: 10.1007/978-3-319-46475-6_43

Li, R., Zhang, W., Suk, H.-I., Wang, L., Li, J., Shen, D., et al. (2014). "Deep learning based imaging data completion for improved brain disease diagnosis," in *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2014* (Boston, MA), 305–312. doi: 10.1007/978-3-319-10443-0_39

Menon, R. S., Gati, J. S., Goodyear, B. G., Luknowsky, D. C., and Thomas, C. G. (1998). Spatial and temporal resolution of functional magnetic resonance imaging. *Biochem. Cell Biol.* 76, 560–571. doi: 10.1139/o98-051

Mirza, M., and Osindero, S. (2014). Conditional generative adversarial nets. *arXiv [Preprint]. arXiv:1411.1784.*

Olut, S., Sahin, Y. H., Demir, U., and Unal, G. (2018). "Generative adversarial training for MRA image synthesis using multi-contrast MRI," in *International Workshop on Predictive Intelligence in Medicine* (Granada), 147–154. doi: 10.1007/978-3-030-00320-3_18

Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., Chen, X., et al. (2016). "Improved techniques for training gans," in *Advances in Neural Information Processing Systems, Vol. 29* (Barcelona), 2234–2242.

Sheikh, H. R., and Bovik, A. C. (2006). Image information and visual quality. *IEEE Trans. Image Process. A Publ. IEEE Signal Process. Soc.* 15, 430–444. doi: 10.1109/TIP.2005.859378

Sheikh, H. R., Member, IEEE, Bovik, A. C., and Fellow (2006). An information fidelity criterion for image quality assessment using natural scene statistics. *IEEE Trans. Image Process.* 14, 2117–2128. doi: 10.1109/TIP.2005.859389

Tang, H., Xu, D., Sebe, N., and Yan, Y. (2019). "Attention-guided generative adversarial networks for unsupervised image-to-image translation," in *2019 International Joint Conference on Neural Networks (IJCNN)* (Budapest), 1–8. doi: 10.1109/IJCNN.2019.8851881

Wang, C., Xu, C., Wang, C., and Tao, D. (2018). Perceptual adversarial networks for image-to-image transformation. *IEEE Trans. Image Process.* 27, 4066–4079. doi: 10.1109/TIP.2018.2836316

Wang, T.-C., Liu, M.-Y., Zhu, J.-Y., Tao, A., Kautz, J., and Catanzaro, B. (2018). "High-resolution image synthesis and semantic manipulation with conditional gans," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. doi: 10.1109/CVPR.2018.00917

Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* 13, 600–612. doi: 10.1109/TIP.2003.819861

Welander, P., Karlsson, S., and Eklund, A. (2018). Generative adversarial networks for image-to-image translation on multi-contrast MR images - A comparison of cyclegan and UNIT. *arXiv [Preprint]. arXiv:1806.07777.*

Wolterink, J. M., Dinkla, A. M., Savenije, M. H. F., Seevinck, P. R., van den Berg, C. A. T., and Isgum, I. (2017). "Deep MR to CT synthesis using unpaired data," in *Simulation and Synthesis in Medical Imaging*, eds S. A. Tsaftaris, A. Gooya, A. F. Frangi, J. L. Prince (Cham: Springer International Publishing), 14–23. doi: 10.1007/978-3-319-68127-6_2

Yang, Q., Li, N., Zhao, Z., Fan, X., Chang, E.-C., Xu, Y., et al. (2018). MRI image-to-image translation for cross-modality image registration and segmentation. *arXiv [Preprint]. arXiv:1801.06940.*

Ye, D. H., Zikic, D., Glocker, B., Criminisi, A., and Konukoglu, E. (2013). "Modality propagation: Coherent synthesis of subject-specific scans with data-driven regularization," in *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2013*, eds. K. Mori, I. Sakuma, Y. Sato, C. Barillot, and N. Navab (Berlin: Heidelberg), 606–613. doi: 10.1007/978-3-642-40811-3_76

Yi, X., Walia, E., and Babyn, P. (2019). Generative adversarial network in medical imaging: a review. *Med. Image Anal.* 58:101552. doi: 10.1016/j.media.2019.101552

Zhao, J. J., Mathieu, M., and LeCun, Y. (2016). Energy-based generative adversarial network. *arXiv [Preprint]. arXiv:1609.03126.*

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.