Generative adversarial networks with adaptive normalization for synthesizing T2-weighted magnetic resonance images from diffusion-weighted images

Yanyan Mao[1,2], Chao Chen[2]*, Zhenjie Wang[1],
Dapeng Cheng[2,3], Panlu You[2], Xingdan Huang[4],
Baosheng Zhang[2] and Feng Zhao[2,3]

[1]College of Oceanography and Space Informatics, China University of Petroleum, Qingdao, China, [2]School of Computer Science and Technology, Shandong Business and Technology University, Yantai, China, [3]Shandong Co-Innovation Center of Future Intelligent Computing, Yantai, China, [4]School of Statistics, Shandong Business and Technology University, Yantai, China

Recently, attention has been drawn toward brain imaging technology in the medical field, among which MRI plays a vital role in clinical diagnosis and lesion analysis of brain diseases. Different sequences of MR images provide more comprehensive information and help doctors to make accurate clinical diagnoses. However, their costs are particularly high. For many image-to-image synthesis methods in the medical field, supervised learning-based methods require labeled datasets, which are often difficult to obtain. Therefore, we propose an unsupervised learning-based generative adversarial network with adaptive normalization (AN-GAN) for synthesizing T2-weighted MR images from rapidly scanned diffusion-weighted imaging (DWI) MR images. In contrast to the existing methods, deep semantic information is extracted from the high-frequency information of original sequence images, which are then added to the feature map in deconvolution layers as a modality mask vector. This image fusion operation results in better feature maps and guides the training of GANs. Furthermore, to better preserve semantic information against common normalization layers, we introduce AN, a conditional normalization layer that modulates the activations using the fused feature map. Experimental results show that our method of synthesizing T2 images has a better perceptual quality and better detail than the other state-of-the-art methods.

KEYWORDS

magnetic resonance imaging (MRI), images synthesis, generative adversarial network (GAN), image fusion, adaptive normalization

Introduction

Magnetic resonance imaging (MRI) has been used worldwide for the diagnosis of various conditions throughout the body, among which the brain and spine are the most effective. The magnetic field strength of MRI scanners has evolved from less than 0.5 T in the 1980s to the extensively used 1.5 and 3 T, and even 7 T (Van der Kolk et al., 2013; Lian et al., 2018). Different sequences of MR images from MRI scanners help doctors make more accurate decision. While doctors desire different sequences of images, certain conditions (e.g., medical conditions, patients' physical conditions, costs) result in sacrificing some imaging sequences to complete the scan (Shen et al., 2017). For example, diffusion-weighted imaging (DWI) in MR without contrast has a shorter scanning time but lower spatial resolution, making it difficult to identify small lesions. However, we can observe the lesion under higher-field strength T2-weighted MR images (T2).

Interpolation-based methods (e.g., nearest-neighbor, bilinear; Kim et al., 2010; Tam et al., 2010) are simple and rapid, but they blur sharp edges and fine details (Yang et al., 2014). Recent studies have shown promising results using learning-based methods for synthesizing MR images at high-field strengths, such as sparse learning (Zhang et al., 2012) and random forests (Alexander et al., 2014). For example, Xiang et al. (2018) proposed a deep embedding convolutional neural network (DECNN) to synthesize CT images from T1-weighted MR images, Qu et al. (2020) introduced a deep learning network that leverages wavelet domain to synthesize 7 T MRI from 3 T MRI. Isola et al. (2017) proposed that conditional generative adversarial networks (GANs) not only learn image-to-image mappings, but also learn a loss function for training the mappings. GANs tend to synthesize higher-quality images while DWI MRI and T2 MRI differ in both resolution and contrast (Andrew et al., 2019).

This article aims to address the problem of synthesizing T2 MRI from DWI MRI. When scanning a patient's cranial MRI, the brain tissue signals collected under different pulse sequences are quite different. The large variability in these acquired brain tissue signals makes this synthesis problem challenging to solve. Recently, convolutional neural network (CNN) has become a common method for image prediction; furthermore, many studies are constantly improving this model (Liao et al., 2013; Xu et al., 2016; Huang et al., 2021). Briefly, for a given objective, CNN can automatically learn to minimize the loss function. If it takes a simple approach and asks the CNN to minimize Euclidean distance between predicted and ground truth pixels, it will produce muzzy results (Pathak et al., 2016; Zhang et al., 2016). Therefore, it would be highly desirable if we can instead specify only a high-level goal, such as "make the output indistinguishable from ground truth," to synthesize realistic images and then automatically learn a loss function that satisfies this goal, which is also the research direction of

GANs (Goodfellow et al., 2014; Denton et al., 2015). In this work, examples of the DWI image and their corresponding T2 image are shown in **Figure 1**. These images were taken from a cerebral MRI of the same patients. In the DWI MR image, the blue arrows point to the "cerebrospinal fluid (CSF)" with a low-intensity value, and the yellow arrows point to the white matter with a high-intensity value. However, in the T2 MR image, the "CSF" appears to be bright, while the "white matter" appears to be dark. In general, the mapping between these two sequences of DWI and T2 is highly complex. Inspired by this, we incorporate some prior knowledge, such as the difference of signals between different tissues in MRI, to guide the synthesis of the generator and make the generator become more powerful. Therefore, as shown in the generator design method of GANs in **Figure 2**, we introduce high-level semantic information for splicing with features extracted from the source domain and receive the input of this information in the middle layer to generate T2 MRI with richer details.

In this article, we introduce GANs to solve the synthesis problem of DWI MRI to T2 MRI and use an adaptive normalization (AN) before activation function to make it suitable for image synthesis after adding high-level semantic information. Similar to the batch normalization (Ioffe and Szegedy, 2015), the activation is normalized in a channel-wise manner and then modulated with learned scale and shift. As shown in **Figure 3**, there are two generators $G : DWI \rightarrow T2$ and $R : T2 \rightarrow DWI$, and associated adversarial discriminators D_G and D_R. The forward generator G learns the mapping from DWI (source domain) to T2 MRI (target domain), and the reverse generator R learns the mapping from T2 to DWI MRI. We add a cycle consistency loss function (Zhou et al., 2016) that encourage the two cycle-generated behaviors $G(R(T2)) \approx T2$ and $R(G(DWI)) \approx DWI$. Finally, through the experiments we find that the images synthesized by the AN-GAN framework after using the AN method and adding high-level semantic information are effective for medical image synthesis. Some classical image synthesis methods such as pix2pix and cycle-GAN achieved better perceptual appearance; however, there could be excessive deformation in the generated images, and this may affect their clinical applications. However, the images generated by our proposed AN-GAN framework have more details.

Related work

Recent studies on image synthesis based on learning include deep generative models. Many experts and scholars have carried out research on conditional image synthesis in order to improve the quality of synthesized images and achieved gratifying results. They have also made innovations in normalization methods.

Recent deep generative models include GANs (Goodfellow et al., 2014) and variational autoencoder (VAE;
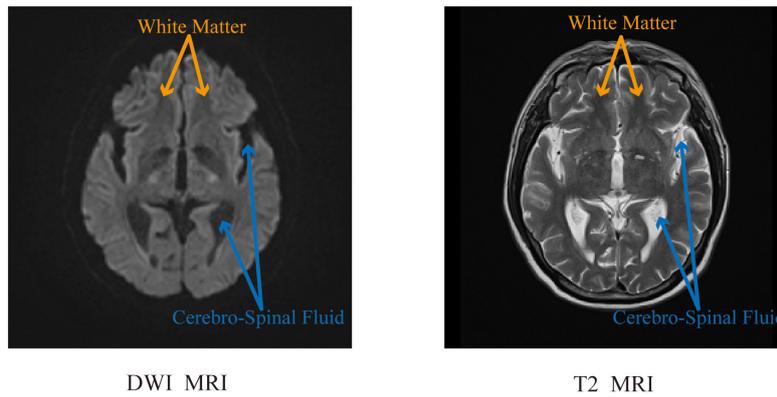
FIGURE 1
DWI MRI and T2 MRI.
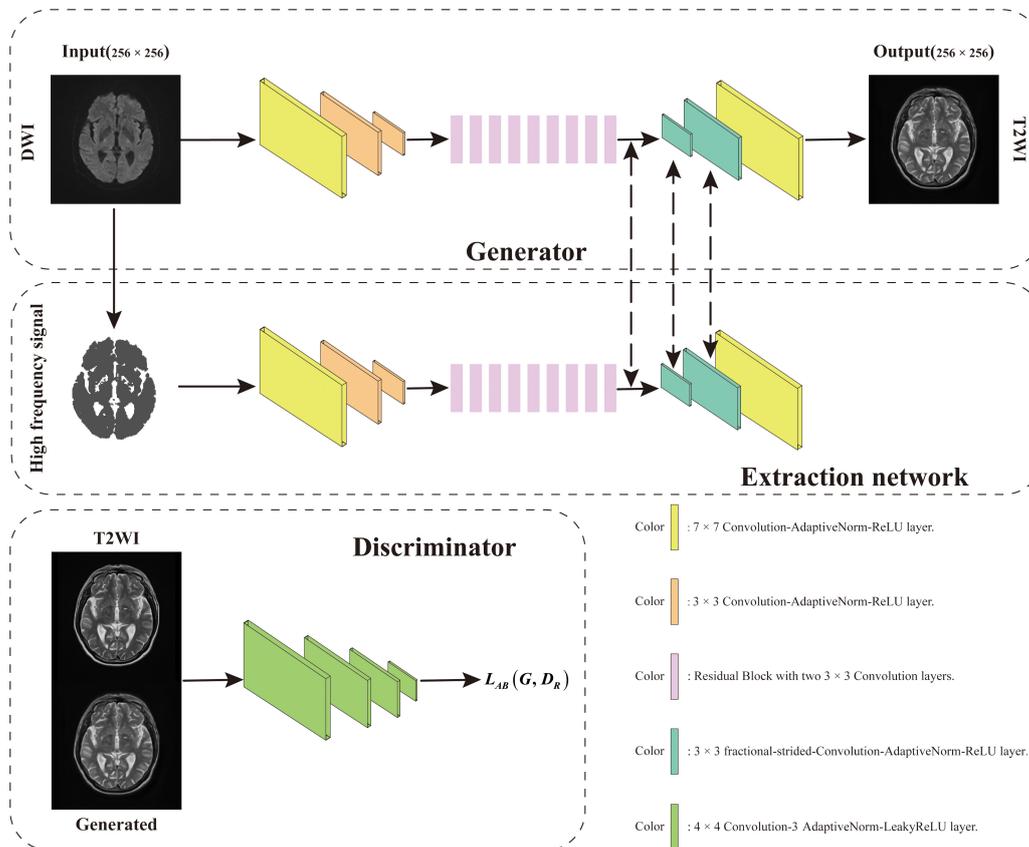
FIGURE 2
AN-GAN framework design consists of three parts, namely, generator, extraction network, and discriminator.

Kingma and Welling, 2014). The network we proposed is built on GANs and adds a module to extract high-level semantic information. The traditional GANs have a generator and a discriminator where the goal of the generator is to generate realistic images so that the discriminator cannot distinguish between synthetic and real images. The adversarial losses in

GANs have achieved many results on image synthesis problems, such as image generation (Denton et al., 2015), image generation (Zhu et al., 2016), and representation learning (Mathieu et al., 2016). Recent studies include person image synthesis and editing (Zhang et al., 2021), image inpainting (Liu et al., 2021), and image attribute editing (Wang et al., 2021). We refer to

FIGURE 3
(A) Two generator mapping models $G: DWI \rightarrow T2$ and $R: T2 \rightarrow DWI$, and two associated adversarial discriminators D_G and D_R. D_G distinguishes between the images generated by G and T2, and vice versa for D_R and R. **(B)** Two consistency loss $c_{G,R}: DWI \rightarrow G(DWI) \rightarrow R(G(DWI)) \approx DWI$ and $c_{R,G}: T2 \rightarrow R(T2) \rightarrow G(R(T2)) \rightarrow T2$ to further regularize the mappings.
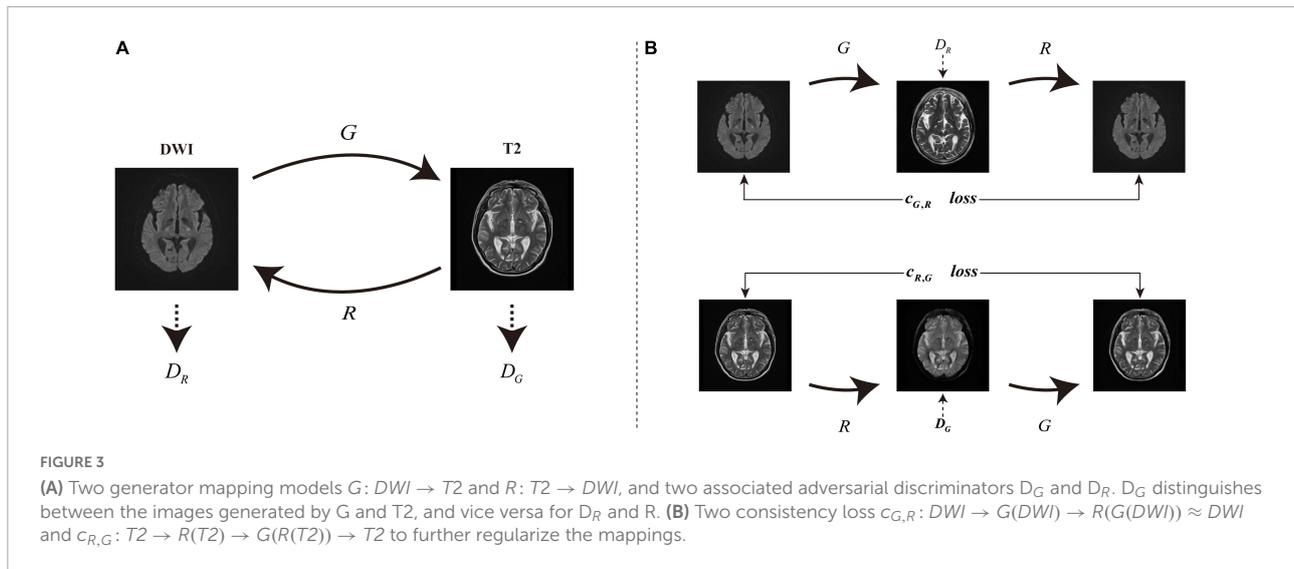
cycle-GAN framework proposed by Zhu et al. (2017) to train two generators and two discriminators, as shown in **Figure 3**, where one generator learns the mapping from the source domain to the target domain and the other generator learns the mapping from the target domain to the source domain, and the two generators and the corresponding discriminators have the same structure.

The input of many problems adopts the idea of conditional image synthesis, such as many models of text-to-image synthesis (Mescheder et al., 2018; Brock et al., 2019). Recent studies include converting the semantic layouts that construct from text to images through an image generator (Hong et al., 2018) and using a single-text condition translate image styles (Kwon and Ye, 2022). Another widely used form is image-to-image synthesis based on conditional GANs, where both input and output are images. Image-to-image translation can be traced back to the work of Hertzmann et al. (2001) on image analogies, who used non-parametric models (Efros and Leung, 1999) to create new images from a single paradigm. Recent studies have enhanced the expressiveness of the generator by providing an example style map to control the style of output image (Huang et al., 2018) and extracting information from semantic layout and scene attributes as condition variables (Karacan et al., 2016). Mirza and Osindero (2014) proposed conditional generative adversarial nets that use the given labels to generate specific images in the testing phase. The "pix2pix" framework was proposed by Isola et al. (2017), who used conditional GANs to learn image-to-image mappings. Zhu et al. (2017) proposed a cycle-GAN framework on this basis using unpaired data for training. Choi et al. (2018) proposed StarGAN that implements the transfer of multiple domains using one model. Compared with earlier non-parametric-based methods such as composing realistic pictures from simple freehand sketches annotated with text labels (Chen et al., 2009), learning-based methods are

generally faster during testing. In this article, DWI images and high-level semantic information extracted from DWI images are used as training sets, and the proposed AN method updates affine parameters to synthesize more detailed T2 MR images.

The normalization layer is an important part of the deep learning network now, including unconditional normalization and conditional normalization, which can be found in various classifiers. Currently popular unconditional normalization layers include instance normalization (Ulyanov et al., 2016), layer normalization (Ba et al., 2016), group normalization (Wu and He, 2018), and weight normalization (Salimans and Kingma, 2016). Conditional normalization includes conditional batch normalization (Dumoulin et al., 2017) and adaptive instance normalization (Huang and Belongie, 2017). Different from the earlier normalization techniques, conditional normalization layers require external data and generally operate as follows. First, layer activations are normalized to zero mean and unit deviation. Then the normalized activations are denormalized by modulating the activation using a learned affine transformation whose parameters are inferred from external data. In the style transfer task, these affine parameters are used to control the global style, and the spatial coordinates are consistent. However, our proposed AN applies spatially varying affine transformations and is suitable for synthesizing medical images by generators that incorporate high-level semantic information of target domain images.

Adaptive image synthesis

The goal of this article is to learn the mapping from set A to set B, that is, the mapping function from DWI to T2 MR. Training samples $\{x_1, x_2, \ldots, x_N\}$, $x_i \in A$, $\{y_1, y_2, \ldots, y_N\}$, $y_i \in B$, where N means the number of

training samples. We simultaneously train two generators $G : A \rightarrow B$ and $R : B \rightarrow A$, and two corresponding discriminators D_G and D_R. This training procedure is shown in **Figure 3**. The discriminator D_G distinguishes between the image $\{y\}$ of the target domain and the image $\{G(x)\}$ generated by the source domain, while the discriminator D_R distinguishes the image $\{x\}$ of the source domain and the image $\{R(y)\}$ generated by the target domain. Our training objective consists of two aspects, namely, adversarial loss function and consistency loss function. The former uses the data distribution of the images generated in the source domain to match the data distribution of the target domain images, and the latter prevents conflicts between the learned generators G and R. We also use the extracted high-level semantic information as a condition to guide the generator to synthesize images and use an AN method to make it suitable for image synthesis after stitching high-level semantic information.

Here, we give the design principle of cycle-GAN. We assume that there is some potential relationship between the domains. For example, they are the presentation of two different signals of brain tissues and organs. We can use supervision at the sets of level (there is one set of images in domain A and a different set in domain B) in the absence of paired data. Therefore, we design two mappings, i.e., $G : A \rightarrow B$ and $R : B \rightarrow A$. Meanwhile, G and R should be inverse to each other, and both mappings are bijections. The goal of the mapping G is that the output $\hat{y} = G(x)$, $x \in A$, is indistinguishable from images $y \in B$. The goal of the mapping R is that the output $\hat{x} = R(y)$, $y \in B$, is indistinguishable from images $x \in A$. This article realizes this assumption by training the mappings G and R simultaneously. Then the consistency loss is introduced to enforce guarantee $G(R(B)) \approx B$. Finally, this loss is combined with the adversarial loss in domain A and domain B to achieve the goal of image-to-image conversion.

Adversarial loss

The mapping functions of both generators use adversarial losses, first proposed by Goodfellow et al. (2014). Let y be the adversarial loss function composed of the generator $G : A \rightarrow B$ and the corresponding discriminator D_G as follows:

$$L_{GD_G}(G, D_G, A, B) = E_{x \sim P_x}\left[\log\left(1 - D_G(\hat{y})\right)\right] + \quad (1)$$
$$E_{y \sim P_y}\left[logD_G(y)\right]$$

where $\hat{y} = G(x)$, P_x and P_y are the distributions of the source domain and the ground truth image. The goal of the generator G is to synthesize images $\hat{y}$ that look similar to the real image y, which is in the set B, while the goal of the discriminator D_G is to distinguish $\hat{y}$ from y. Similarly, the adversarial loss function composed by the generator $R : B \rightarrow A$ and the corresponding

discriminator D_R can be written as follows:

$$L_{RD_R}(R, D_R, B, A) = E_{y \sim P_y}\left[\log\left(1 - D_R(\hat{x})\right)\right] + \quad (2)$$
$$E_{x \sim P_x}\left[logD_R(x)\right]$$

where $\hat{x} = R(y)$. The goal of the generator R is to synthesize images $\hat{x}$ that look similar to the real image x, which is in the set B, while the goal of the discriminator D_G is to distinguish $\hat{x}$ from x. Finally, our goal is to make the image synthesized by the generator closer to the real image, against the discriminator that distinguishes the generated image from the real image, which can be expressed as follows:

$$\min_G \max_{D_G} L_{GD_G}(G, D_G, A, B), \quad \min_R \max_{D_R} L_{RD_R} \quad (3)$$
$$(R, D_R, B, A)$$

Consistency loss

Our model uses l_1 regularization as a pixel-level constraint to penalize network in order to avoid the blurring effect of the generated images when using l_2 regularization. Our discriminator adopts the structure of "70 × 70" PatchGAN. This discriminator divides image into $N \times N$ patches equally, penalizes structure at the scale of patches, and then classifies the true and false of each patch., As shown in **Figure 3**, we think the mapping function is cycle-consistent which can reduce unnecessary mapping between set A and set B. Therefore, we use a consistency loss to motivate the two behaviors $x \rightarrow \hat{y} \rightarrow R(\hat{y}) \approx x$ and $y \rightarrow \hat{x} \rightarrow G(\hat{x}) \approx y$ as follows

$$L_{adv}(G, R) = E_{y \sim P_y}\left[G(\hat{x}) - y_1\right] + E_{x \sim P_x}\left[R(\hat{y}) - x_1\right] \quad (4)$$

where $||*||_{;1}$ means l_1 -norm. Our total loss function is written as follows:

$$L = L_{GD_G}(G, D_G, A, B) + L_{RD_R}(R, D_R, B, A) + \quad (5)$$
$$\lambda L_{adv}(G, R)$$

where $\lambda = 10$.

Adaptive normalization

We propose a new conditional normalization method to address the problem regarding the fusion of introduced high-level semantic information. Let $m \in Q^{H \times W}$ be a mask, where H is the image height and W is the image width. The high-level semantic information extracted from the residual network is fused to the source domain information. Then they are characterized on m. First, let l^i be the activations for a batch of N samples in the ith layer in a deep convolutional network. Let C^i be the number of channels in this layer. Let H^i and W^i be the height and width of the activation map in this
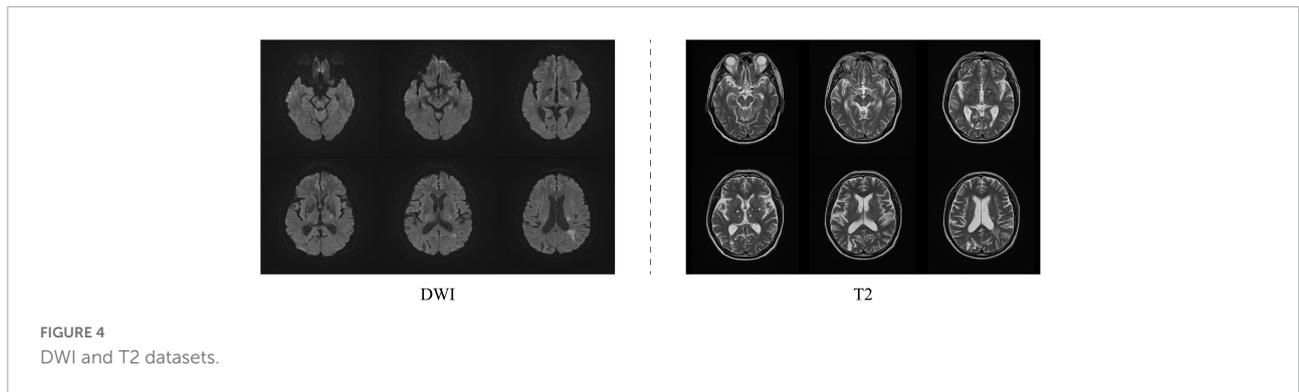
FIGURE 4
DWI and T2 datasets.

layer. Similar to the batch normalization, the pixel values after the convolution operation are normalized in a channel-wise manner, then modulate with the learned scaling parameter γ and shift parameter μ. The value in space ($n \in N, c \in C^i, a \in H^i, b \in W^i$) can be expressed as follows:

$$\gamma_{c,a,b}^i (m) \frac{h_{n,c,a,b}^i - \mu_c^i}{\sigma_c^i} + \beta_{c,a,b}^i (m) \qquad (6)$$

$$\mu_c^i = \frac{1}{NH^iW^i} \sum_{n,a,b} h_{n,c,a,b}^i, \qquad (7)$$

$$\sigma_c^i = \sqrt{\frac{1}{NH^iW^i} \sum_{n,a,b} \left(\left(h_{n,c,a,b}^i\right)^2 - \left(\mu_c^i\right)^2 \right)}$$

where $h_{n,c,a,b}^i$ is the activation before normalization. μ_c^i and σ_c^i are the mean and standard deviation of the activations in channel c. $\gamma_{c,a,b}^i (m)$ and $\beta_{c,a,b}^i (m)$ are affine parameters learned in the normalization layer, which depend on the high-level semantic information and vary with respect to the location (a, b). The value of mask m in the activation map at the site (n, a, b) is updated by using the scaling parameter $\gamma_{c,a,b}^i (m)$ and the translation parameter $\beta_{c,a,b}^i (m)$.

Experiments

To verify the effectiveness of proposed AN-GAN, we evaluate our method through experiments on MR image translation, including descriptions of experimental data, experimental settings, and evaluation metrics. In the ablation experiments, the improvements we explore benefit from two points in the proposed framework, namely, high-level semantic information extracted from the source domain and the proposed new normalization method. To verify the effectiveness of these two modules, we will remove the high-level semantic information module in the AN-GAN framework, use a batch normalization method, and compare these two experimental results with the proposed method. Finally, we will compare with existing GANs models CGAN, pix2pix,

cycle-GAN, and StarGAN. The experimental data included two MRI sequences, namely, DWI and T2. **Figure 4** shows an example of these data.
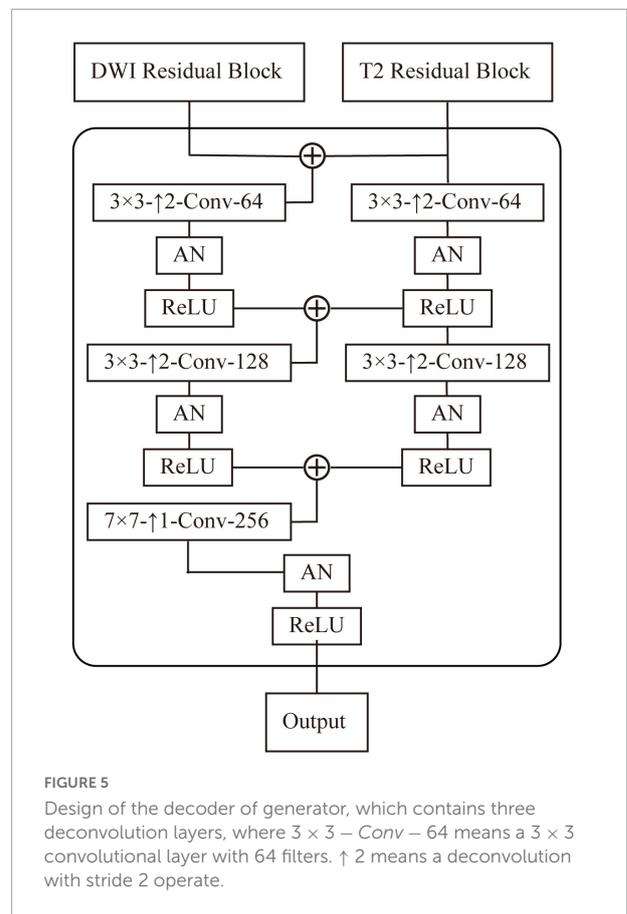

FIGURE 5
Design of the decoder of generator, which contains three deconvolution layers, where $3 \times 3 - Conv - 64$ means a 3×3 convolutional layer with 64 filters. ↑ 2 means a deconvolution with stride 2 operate.

TABLE 1 Quantitative assessment of MRI image conversion.

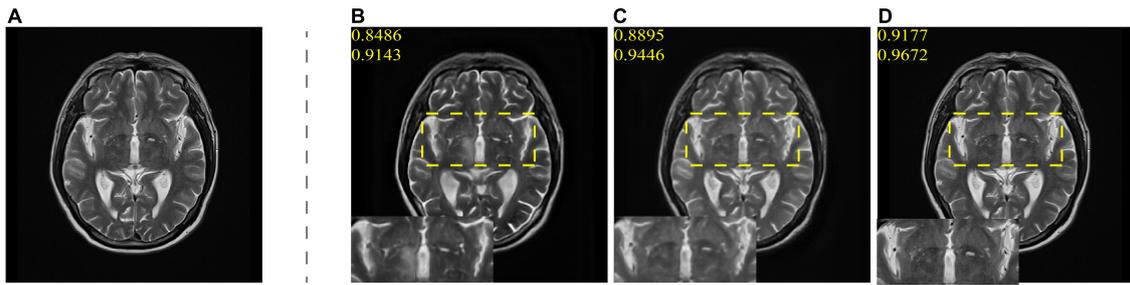| Method	MSE	PSNR	SSIM	FSIM
AN-GAN (Without AN)	111.22 ± 5	27.67 ± 1	0.85 ± 0.2	0.91 ± 0.2
AN-GAN (Without mask)	96.86 ± 5	28.27 ± 1	0.89 ± 0.2	0.94 ± 0.2
AN-GAN	86.40 ± 5	29.30 ± 1	0.92 ± 0.1	0.97 ± 0.2

FIGURE 6
(A) is the real T2 image; **(B)** is the T2 image generated with mask information; **(C)** is the T2 image generated with AN; and **(D)** is the T2 images generated with mask information and AN. The numbers in yellow represent the SSIM and FSIM scores of the generated images.
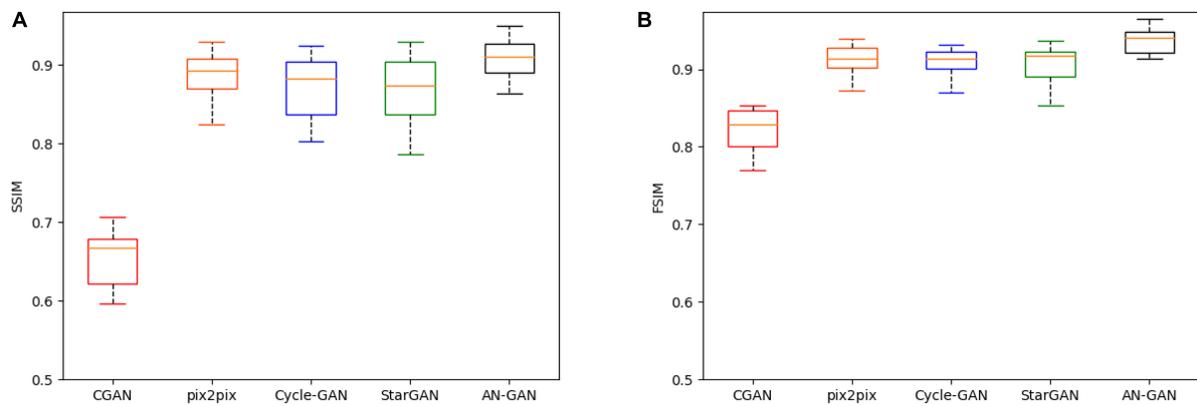

FIGURE 7
Boxplots of SSIM **(A)** and FSIM **(B)** scores for synthetic images of three network models (CGAN, pix2pix, Cycle-GAN, StarGAN, and AN-GAN) on the test set.
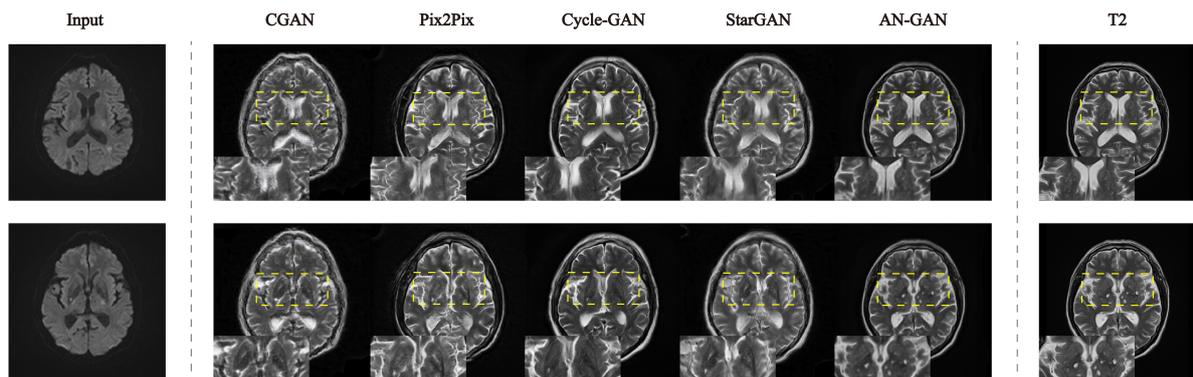

FIGURE 8
Experimental results of CGAN, pix2pix, Cycle-GAN, StarGAN, and AN-GAN in the test set. Synthesis of images from DWI to T2.

Datasets

The study was approved by the Institutional Review Board of Yantai Yuhuangding Hospital and the Ethics Committee of Shandong Technology and Business University, and informed consent from the patients was waived. The experimental data used in this article are collected from 20 adult volunteers at Yantai Yuhuangding Hospital. They were scanned under a DISCOVERY MR750w MRI scanner, using a self-shielding gradient set with a maximum gradient amplitude of $40\mathrm{m^T m^{-1}}$.

The size of each DWI MR image is 256 × 256 × 16, the size of each T2 MR image is 256 × 256 × 12, and the voxel size is 1 mm × 1 mm × 1 mm.

Experimental settings

We conduct all experiments using Windows 11 with NVIDIA RTX3060 GPU and 12th Gen Intel(R) Core (TM) i5-12400F CPU, and the environment is Python3.7 and PyTorch1.8.0. Based on the scanning direction of the 3D medical images, we slice the 3D medical images scanned by all MRI equipment into 2D images and use the 2D slices to train our proposed model. When conducting experiments, we crop the size of the input image to 256 × 256, where the parameter λ of Eq. 5 is set to 10, the initial learning rate is set to 0.0001, and the batch size is set to 2.

The AN-GAN framework is shown in **Figure 5**. We normalize each layer of the network using an AN approach. The discriminator uses 70 × 70 PatchGANs (Isola et al., 2017). It uses fewer parameters than the full-image discriminator and can process arbitrary images in a convolutional fashion. During the training process, the adversarial loss uses a least-squares loss (Mao et al., 2017). In the process of deconvolution, the high-level semantic information of the source domain is used to guide the synthesis of the generator, where the decoder design is shown in **Figure 5**.

In this article, we use mean squared error (MSE), peak signal to noise ratio (PSNE), structural similarity (SSIM) (Wang et al., 2004), and feature similarity index measure (FSIM) (Zhang et al., 2011) for an objective evaluation of image translation results. The real images of all target domains are used as reference datasets, and the SSIM and FSIM scores of the generated images are used as quantitative evaluation criteria.

Ablation study

To verify the effectiveness of adding AN methods and high-level semantic information, we compare AN-GAN with AN with AN-GAN using only batch normalization and AN-GAN adding only high-frequency information GANs are compared and quantitatively analyzed in our dataset. As shown in **Table 1**, AN-GAN using AN and high-level semantic information effectively improves image translation performance. The synthesized image is shown in **Figure 6**. From the CSF in the yellow box area in **Figures 6B,D**, it can be seen that only adding high-frequency information does not improve the quality of image translation, and the synthesized result is blurred. Among them, (**Figure 6C**) compared with (**Figure 6D**), after using high-frequency information and AN, the image structure synthesized by AN-GAN is more complete, the details are richer, and it is closer to the real image. SSIM and FSIIM scores are used

for quantitative evaluation. We find that using high-frequency information and AN, SSIM achieves 0.9177 and FSIM achieves 0.9762, which are higher than the scores of 0.8486 and 0.9143 for adding high-frequency information alone and 0.8895 and 0.9446 for using AN alone.

Comparison

We evaluate the feasibility and effectiveness of the AN-GAN framework that will be compared with classic models in the field of image translation as follows:

- CGAN: a method to generate a specific image. The proposal of CGAN enables GAN to use images and corresponding labels for training, and use the given labels to generate specific images in the testing phase.
- Pix2pix: a method using patch-level discriminators.
- Cycle-GAN: a method to learn two mappings simultaneously.
- StarGAN: this method implements the transfer of multiple domains using one model.

In this article, these five models are quantitatively evaluated on the test set. The results of SSIM and FSIM are shown in the boxplot in **Figure 7**. Compared with CGAN, pix2pix, Cycle-GAN, and StarGAN, the AN-GAN framework is more stable and the effect is better. As shown in **Figure 8**, the images generated by the CGAN framework have large deformations. Pix2pix, Cycle-GAN, and StarGAN frameworks all successfully implement image translation between the two domains with good results, but some brain tissues are not visually clear. The results of our proposed AN-GAN show clear details and distinct texture of soft tissue, which is superior to other methods.

Conclusion

In various clinical scenarios, medical images are crucial for the diagnosis and treatment of diseases. Different sequences of MRI images provide doctors with different lesion information, which complement each other and help doctors make accurate decisions in clinical scenarios. However, the cost of MRI equipment is high. Using low-cost scanned MRI image sequences to synthesize other sequence MRI images can not only save the cost of patients but also provide doctors with more comprehensive lesion information. In contrast to the existing methods, this article proposes the AN-GAN framework, which adds high-frequency information to guide generator training and designs an AN method to make it suitable for generators incorporating the high-frequency information. Through experiments on our collected datasets, we demonstrate that AN-GAN outperforms other state-of-the-art methods.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by the Institutional Review Board of Yantai Yuhuangding Hospital and Ethics Committee of Shandong Technology and Business University. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

Author contributions

YM: conceptualization, methodology, and writing—review and editing. CC: conceptualization, software, writing—original draft, methodology, formal analysis, investigation, and validation. ZW: writing—review, editing, and validation. DC, PY, XH, BZ, and FZ: writing—review and editing. All authors contributed to the article and approved the submitted version.

Funding

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.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Alexander, D. C., Zikic, D., Zhang, J., Zhang, H., and Criminisi, A. (2014). Image quality transfer via random forest regression: Applications in diffusion MRI. *Med. Image Comput. Comput. Assist. Interv.* 17(Pt 3), 225–232. doi: 10.1007/978-3-319-10443-0_29

Andrew, B., Jeff, D., and Karen, S. (2019). Large scale GAN training for high fidelity natural image synthesis. *arXiv [Preprint]*

Ba, J., Kiros, J. R., and Hinton, G. E. (2016). Layer normalization. *ArXiv [Preprint]*

Brock, A., Donahue, J., and Simonyan, K. (2019). Large scale GAN training for high fidelity natural image synthesis. *ArXiv [Preprint]* doi: 10.48550/arXiv.1809.11096

Chen, T., Cheng, M.-M., Tan, P., Shamir, A., and Hu, S. (2009). Sketch2photo: Internet image montage. *ACM Trans. Graph.* 28:124. doi: 10.1145/1618452.1618470

Choi, Y., Choi, M., Kim, M., Ha, J. W., Kim, S., and Choo, J. (2018). "Stargan: Unified generative adversarial networks for multi-domain image-to-image translation," in *Proceedings of the 2018 IEEE/CVF conference on computer vision and pattern recognition*, Salt Lake City, UT, 8789–8797. doi: 10.1109/CVPR.2018.00916

Denton, E., Chintala, S., Szlam, A., and Fergus, R. (2015). Deep generative image models using a laplacian pyramid of adversarial networks. *ArXiv [Preprint]* doi: 10.48550/arXiv.1506.05751

Dumoulin, V., Shlens, J., and Kudlur, M. (2017). A learned representation for artistic style. *ArXiv [Preprint]* doi: 10.48550/arXiv.1610.07629

Efros, A. A., and Leung, T. K. (1999). "Texture synthesis by non-parametric sampling," in *Proceedings of the seventh IEEE international conference on computer vision*, Vol. 2, Kerkyra, 1033–1038. doi: 10.1109/ICCV.1999.790383

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2014). "Generative adversarial nets," in *Proceedings of the 27th international conference on neural information processing systems - volume 2*, Cambridge, MA, 2672–2680.

Hertzmann, A., Jacobs, C. E., Oliver, N., Curless, B., and Salesin, D. (2001). "Image analogies," in *Proceedings of the SIGGRAPH conference*, New York, NY, 327–340. doi: 10.1145/383259.383295

Hong, S., Yang, D., Choi, J., and Lee, H. (2018). "Inferring semantic layout for hierarchical text-to-image synthesis," in *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, Silver Spring, MD, 7986–7994.

Huang, X., and Belongie, S. (2017). "Arbitrary style transfer in real-time with adaptive instance normalization," in *Proceedings of the IEEE international conference on computer vision (ICCV)*, Montreal, 1510–1519. doi: 10.1109/ICCV.2017.167

Huang, X., Liu, M.-Y., Belongie, S., and Kautz, J. (2018). "Multimodal unsupervised image-to-image translation," in *Computer vision – ECCV 2018. ECCV 2018. Lecture notes in computer science*, Vol. 11207, eds V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss (Cham: Springer), 179–196. doi: 10.1007/978-3-030-01219-9_11

Huang, Y., Zheng, F., Wang, D., Huang, W., Scott, M. R., and Shao, L. (2021). "Brain image synthesis with unsupervised multivariate canonical CSC l4Net," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, (Nashville, TN: IEEE), 5877–5886. doi: 10.1109/CVPR46437.2021.00582

Ioffe, S., and Szegedy, C. (2015). "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proceedings of the 32nd international conference on international conference on machine learning - volume 37 (ICML'15)*, Lille, 448–456. doi: 10.1007/s11390-020-0679-8

Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A. (2017). "Image-to-image translation with conditional adversarial networks," in *Proceedings of the 2017 IEEE conference on computer vision and pattern recognition (CVPR)*, Honolulu, HI, 5967–5976.

Karacan, L., Akata, Z., Erdem, A., and Erdem, E. (2016). Learning to generate images of outdoor scenes from attributes and semantic layouts. *ArXiv [Preprint]* doi: 10.48550/arXiv.1612.00215

Kim, K., Habas, P. A., Rousseau, F., Glenn, O. A., Barkovich, A. J., and Studholme, C. (2010). Intersection based motion correction of multislice MRI for

3-D *in utero* fetal brain image formation. *IEEE Trans. Med. Imaging* 29, 146–158. doi: 10.1109/TMI.2009.2030679

Kingma, D. P., and Welling, M. (2014). "Auto-encoding variational bayes," in *Proceedings of the 2nd international conference on learning representations, ICLR 2014*, Banff. doi: 10.1093/bioinformatics/btaa169

Kwon, G., and Ye, J. (2022). CLIPstyler: Image style transfer with a single text condition. *ArXiv [Preprint]* doi: 10.48550/arXiv.2112.00374

Lian, C., Zhang, J., Liu, M., Zong, X., Hung, S. C., Lin, W., et al. (2018). Multi-channel multi-scale fully convolutional network for 3D perivascular spaces segmentation in 7T MR images. *Med. Image Anal.* 46, 106–117. doi: 10.1016/j.media.2018.02.009

Liao, S., Gao, Y., Oto, A., and Shen, D. (2013). Representation learning: A unified deep learning framework for automatic prostate MR segmentation. *Med. Image Comput. Comput. Assist. Interv.* 16(Pt 2), 254–261. doi: 10.1007/978-3-642-40763-5_32

Liu, H., Wan, Z., Huang, W., Song, Y., Han, X., and Liao, J. (2021). "PD-GAN: Probabilistic diverse GAN for image inpainting," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, New Orleans, LA, 9367–9376. doi: 10.1109/CVPR46437.2021.00925

Mao, X., Li, Q., Xie, H., Lau, R. Y. K., Wang, Z., and Smolley, S. P. (2017). "Least squares generative adversarial networks," in *Proceedings of the IEEE international conference on computer vision (ICCV)*, Montreal, 2813–2821. doi: 10.1109/ICCV.2017.304

Mathieu, M., Zhao, J., Sprechmann, P., and Ramesh, A. (2016). "Disentangling factors of variation in deep representation using adversarial training," in *Proceedings of the 30th conference on neural information processing systems (NIPS 2016)*, Barcelona, 5047–5055. doi: 10.1109/TPAMI.2021.3077397

Mescheder, L., Geiger, A., and Nowozin, S. (2018). "Which training methods for GANs do actually converge?," in *Proceedings of the international conference on machine learning (ICML)*, Baltimore, MD, 3481–3490.

Mirza, M., and Osindero, S. (2014). Conditional generative adversarial nets. *ArXiv [Preprint]* doi: 10.48550/arXiv.1411.1784

Pathak, D., Krähenbühl, P., Donahue, J., Darrell, T., and Efros, A. A. (2016). "Context encoders: Feature learning by inpainting," in *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, (Seattle, WA: IEEE), 2536–2544. doi: 10.1109/CVPR.2016.278

Qu, L., Zhang, Y., Wang, S., Yap, P. T., and Shen, D. (2020). Synthesized 7T MRI from 3T MRI via deep learning in spatial and wavelet domains. *Med. Image Anal.* 62:101663. doi: 10.1016/j.media.2020.101663

Salimans, T., and Kingma, D. P. (2016). "Weight normalization: A simple reparameterization to accelerate training of deep neural networks," in *Proceedings of the 30th international conference on neural information processing systems*, Red Hook, NY, 901–909.

Shen, D., Wu, G., and Suk, H. I. (2017). Deep learning in medical image analysis. *Annu. Rev. Biomed. Eng.* 19, 221–248. doi: 10.1146/annurev-bioeng-071516-044442

Tam, W.-S., Kok, C.-W., and Siu, W.-C. (2010). Modified edge-directed interpolation for images. *J. Electron. Imaging* 19:013011. doi: 10.1117/1.3358372

Ulyanov, D., Vedaldi, A., and Lempitsky, V. (2016). Instance normalization: The missing ingredient for fast stylization. *ArXiv [Preprint]* doi: 10.48550/arXiv.1607.08022

Van der Kolk, A. G., Hendrikse, J., Zwanenburg, J. J., Visser, F., and Luijten, P. R. (2013). Clinical applications of 7 T MRI in the brain. *Eur. J. Radiol.* 82, 708–718. doi: 10.1016/j.ejrad.2011.07.007

Wang, T., Zhang, Y., Fan, Y., Wang, J., and Chen, Q. (2021). High-Fidelity GAN inversion for image attribute editing. *ArXiv [Preprint]* doi: 10.48550/arXiv.2109.06590

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

Wu, Y., and He, K. (2018). Group normalization. *Int. J. Comput. Vis.* 128, 742–755. doi: 10.1007/s11263-019-01198-w

Xiang, L., Wang, Q., Nie, D., Zhang, L., Jin, X., Qiao, Y., et al. (2018). Deep embedding convolutional neural network for synthesizing CT image from T1-Weighted MR image. *Med. Image Anal.* 47, 31–44. doi: 10.1016/j.media.2018.03.011

Xu, J., Xiang, L., Liu, Q., Gilmore, H., Wu, J., Tang, J., et al. (2016). Stacked Sparse Autoencoder (SSAE) for nuclei detection on breast cancer histopathology images. *IEEE Trans. Med. Imaging* 35, 119–130. doi: 10.1109/TMI.2015.2458702

Yang, C.-Y., Ma, C., and Yang, M.-H. (2014). "Single-image super-resolution: A benchmark," in *Computer vision – ECCV 2014. ECCV 2014. Lecture notes in computer science*, Vol. 8692, eds D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars (Cham: Springer), 372–386. doi: 10.1007/978-3-319-10593-2_25

Zhang, J., Li, K., Lai, Y.-K., and Yang, J. (2021). "PISE: Person image synthesis and editing with decoupled GAN," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, (Nashville, TN: IEEE), 7978–7986. doi: 10.1109/CVPR46437.2021.00789

Zhang, L., Zhang, L., Mou, X., and Zhang, D. (2011). FSIM: A feature similarity index for image quality assessment. *IEEE Trans. Image Process.* 20, 2378–2386. doi: 10.1109/TIP.2011.2109730

Zhang, R., Isola, P., and Efros, A. A. (2016). "Colorful image colorization," in *Computer vision – ECCV 2016. ECCV 2016. Lecture notes in computer science*, Vol. 9907, eds B. Leibe, J. Matas, N. Sebe, and M. Welling (Cham: Springer). doi: 10.1007/978-3-319-46487-9_40

Zhang, Y., Wu, G., Yap, P.-T., Feng, Q., Lian, J., Chen, W., et al. (2012). Hierarchical patch-based sparse representation—A new approach for resolution enhancement of 4D-CT lung data. *IEEE Trans. Med. Imaging* 31, 1993–2005. doi: 10.1109/TMI.2012.2202245

Zhou, T., Krahenbuhl, P., Aubry, M., Huang, Q., and Efros, A. A. (2016). "Learning dense correspondence via 3d-guided cycle consistency," in *Proceedings of the 2016 IEEE conference on computer vision and pattern recognition (CVPR)*, Las Vegas, NV, 117–126. doi: 10.1109/CVPR.2016.20

Zhu, J.-Y., Krähenbühl, P., Shechtman, E., and Efros, A. A. (2016). "Generative visual manipulation on the natural image manifold," in *Computer vision – ECCV 2016. ECCV 2016. Lecture notes in computer science*, Vol. 9909, eds B. Leibe, J. Matas, N. Sebe, and M. Welling (Cham: Springer), 597–613. doi: 10.1167/18.11.20

Zhu, J.-Y., Park, T., Isola, P., and Efros, A. A. (2017). "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision (ICCV)*, Montreal, 2242–2251. doi: 10.1109/ICCV.2017.244