Flow-based Deformation Guidance for Unpaired Multi-Contrast MRI Image-to-Image Translation

Toan Duc Bui *[1], Manh Nguyen[1,2], Ngan Le[3], and Khoa Luu[3]

[1] VinAI Research, Vietnam
[2] FPT University
[3] Department of Computer Science at University of Arkansas in Fayetteville
{v.toanbd1}@vinai.io

Abstract. Image synthesis from corrupted contrasts increases the diversity of diagnostic information available for many neurological diseases. Recently the image-to-image translation has experienced significant levels of interest within medical research, beginning with the successful use of the Generative Adversarial Network (GAN) to the introduction of cyclic constraint extended to multiple domains. However, in current approaches, there is no guarantee that the mapping between the two image domains would be unique or one-to-one. In this paper, we introduce a novel approach to **unpaired image-to-image translation** based on the **invertible architecture**. The invertible property of the flow-based architecture assures a cycle-consistency of image-to-image translation without additional loss functions. **We utilize the temporal information between consecutive slices to provide more constraints to the optimization for transforming one domain to another in unpaired volumetric medical images**. To capture temporal structures in the medical images, we explore the displacement between the consecutive slices using a deformation field. In our approach, the deformation field is used as a guidance to keep the translated slides realistic and consistent across the translation. The experimental results have shown that the synthesized images using our proposed approach are able to archive a competitive performance in terms of mean squared error, peak signal-to-noise ratio, and structural similarity index when compared with the existing deep learning-based methods on three standard datasets, i.e. HCP, MRBrainS13 and Brats2019.

Keywords: flow-based generator, image-to-image translation, cycleGAN

1 Introduction

In medical imaging, the task of obtaining diagnostic images from multiple modalities is necessary for accurate and comprehensive prediction of disease diagnosis. For example, T1-weighted (T1) brain images provide clear differentiate images of gray and white matter tissues, whereas T2-weighted (T2) images differentiate

* Corresponding author

fluid from cortical tissue. By leveraging the information provided by both of these image modalities, we can gain a more in-depth and completed picture of the diagnosis. However, obtaining separately both images is often costly, time-consuming, and maybe corrupted by noise and artifacts. Therefore, cross-modalities synthesis is a promising application to improve the clinical feasibility and utility of multi-contrast MRI. Image-to-image translation has recently gained attention in the medical imaging community, where the task is to estimate the corresponding image in the target domain from a given source domain image of the same subject. Generally, the image-to-image translation methods can be divided into two categories including: Generative Adversarial Networks (GANs) and Flow-based Generative Networks and summarized as follows:

Generative Adversarial Networks GANs are a class of latent variable generative models that clearly identify the generator as *deterministic mapping*. The deterministic mapping represents an image as a point in the latent space without regarding its feature ambiguity. Several different GAN-based models have been used to explore image-to-image translation in a literature study [2, 3, 14, 16]. For example, Zhu et al. [16] proposed a cycleGAN method for mapping between unpaired domains by using cycle-consistency dependence to constrain the optimal solutions provided by the generative network. Balakrishnan et al. [2] proposed a RecycleGAN to explore the temporal information by learning a prediction of the next frame for video generation. Chen et al. [3] proposed a 3D cycleGAN network to learn the mapping between CT and MRI. The drawback of 3D cycleGAN is it is memory consumption and loses the global information due to working on small patch sizes.

Flow-based Generative Networks are a class of latent variable generative models that clearly identify the generator as an *invertible mapping*. The invertible mapping provides a distributional estimation of features in the latent space. Recently, many efforts making use of flow-based generative networks have been proposed to transfer between two unpaired data [4, 5, 7, 10, 12]. For example, Grover et al. [5] introduced a flow to flow (alignflow) network for unpaired image-to-image translation. Sun et al. [12] introduced a conditional dual flow-based invertible network to transfer between positron emission tomography (PET) imaging and magnetic resonance imaging (MRI) images. By using invertible properties, the flow-based methods can ensure exact cycle consistency in translation from a source domain to the target and returning to the source domain without any further loss functions.

Limitations of Existing Methods and Our Contributions The primary drawback of the cycleGAN model is that it can not perform one-to-one mapping for accurate and unique unpaired image translation, generates biased image translations of the inverse mapping [11]. Different from the GANs-based method, the flow-based method guarantees precise cycle consistency in mapping data points from a source domain to the target and returning to the source domain. However, the flow-based methods do not take into account the temporal information between consecutive slices. To address this problem, we propose a new method by inheriting the merits of the flow-based method and exploiting

temporal information between consecutive slices. Our approach provides more constraints to the optimization for transforming one domain to another domain. To capture temporal information, we employ a deformation field between consecutive slices by training a convolutional neural network. In our proposed approach, the deformation field plays a role of guidance to keep slices realistic and consistent across translation.

2 Related work

2.1 Cycle-Consistent Adversarial Networks (cycleGAN)

Let $\{x_i\}_{i=1}^{N}$ and $\{y_i\}_{i=1}^{M}$ be unpaired data samples for two domains, i.e. the source domain X and the target domain Y, respectively. Denote D and G as a discriminator network and a generator network. The cycleGAN model [16] solves unpaired image-to-image translation between these two domains by estimating two independent mapping functions $G_{X\to Y} : X \to Y$ and $G_{Y\to X} : Y \to X$. The two mapping functions $G_{X\to Y}$ and $G_{Y\to X}$ performed by neural networks are trained to fool the discriminator D_X and D_Y respectively. The discriminator D_X, and D_Y encourage the transferred images and the real images to be similar. Hence, the cycleGAN loss is defined as:

$$\begin{aligned}\mathcal{L}_{cycleGAN}(G_{X\to Y}, G_{Y\to X}, D_X, D_Y) = \mathcal{L}_{GAN}(G_{X\to Y}, D_Y) + \mathcal{L}_{GAN}(G_{Y\to X}, D_X) \\ + \lambda\mathcal{L}_{cycle}(G_{X\to Y}, G_{Y\to X}) + \beta\mathcal{L}_{identity}(G_{X\to Y}, G_{Y\to X}) \quad (1)\end{aligned}$$

where $\mathcal{L}_{GAN}$ is a GAN loss for the D network [16]. $\mathcal{L}_{cycle}$ is a cycle consistency loss that guarantees the transferred image from a time-point is able to bring back to the original image after appearance translation by the generator network G. For example, the cycle consistency loss of the data translated from $X \to Y$ via G_X and mapped back to the original domain X via G_Y is defined as:

$$\mathcal{L}_{cycle}(G_{X\to Y}, G_{Y\to X}) = \|G_{Y\to X}(G_{X\to Y}(x)) - x\|_1 \quad (2)$$

The identity loss $\mathcal{L}_{identity}$ is to regularize the generator to be near an identity mapping when real samples of the target domain are given as the input to the generator. The λ and β control the contribution of the two objective functions.

2.2 Flow-based Generative Models

Flow-based Generative Models are a class of latent variable generative models that clearly identify the generator as an invertible mapping $h : Z \to X$ between a set of latent variables Z and a set of observed variables X. Let p_X and p_Z indicate the marginal densities given by the model over X and Z, respectively. Using the change-of-variables formula, these marginal densities are defined as

$$p_X(x) = p_Z(z)\left|\det \frac{\partial h^{-1}}{\partial X}\right|_{X=x} \quad (3)$$

where $z = h^{-1}(x)$ because of the invertibility constraints. In particular, we use a multivariate Gaussian distribution $p_Z(z) = \mathcal{N}(\mu, 0, \mathbf{I})$. Unlike adversarial training, flow models trained with maximum likelihood estimation (MLE) explicitly require a prior $p_Z(z)$ with a tractable density to evaluate model likelihoods using the change-of-variables formula in the equation (3).

Based on flow-based method [4], Grover et al. [5] proposed an alignflow method for unpaired image-to-image translation. In the method, the mapping between two domains $X \rightarrow Y$ can be represented through a shared feature space of latent variables Z by the composition of two invertible mapping [5]:

$$G_{X\rightarrow Y} = G_{Z\rightarrow Y} \circ G_{X\rightarrow Z}, \qquad G_{Y\rightarrow X} = G_{Z\rightarrow X} \circ G_{Y\rightarrow Z} \tag{4}$$

where $G_{X\rightarrow Z} = G^{-1}_{Z\rightarrow X}$ and $G_{Y\rightarrow Z} = G^{-1}_{Z\rightarrow Y}$. Due to the fact that composition of invertible mappings is invertible, both $G_{X\rightarrow Y}$ and $G_{Y\rightarrow X}$ are invertible [5]. On the other hand, we can obtain $G^{-1}_{X\rightarrow Y} = G_{Y\rightarrow X}$. Thus the equation (2) can rewrite as

$$\begin{aligned} \mathcal{L}_{cycle}(G_{X\rightarrow Y}, G_{Y\rightarrow X}) &= \|G_{Y\rightarrow X}(G_{X\rightarrow Y}(x)) - x\|_1 \\ &= \left\|G^{-1}_{X\rightarrow Y}(G_{X\rightarrow Y}(x)) - x\right\|_1 = 0 \end{aligned} \tag{5}$$

where $G^{-1}_{X\rightarrow Y} G_{X\rightarrow Y}$ results in an identical matrix.

Equation. 5 implies that the flow-based methods can guarantee precise cycle consistency in mapping from a source domain to the target and returning to the source domain without additional loss functions. Hence, the alignflow objective loss is defined as:

$$\begin{aligned} \mathcal{L}_{flow}(G_{X\rightarrow Y}, G_{Y\rightarrow X}, D_X, D_Y) &= \mathcal{L}_{GAN}(G_{X\rightarrow Y}, D_Y) + \mathcal{L}_{GAN}(G_{Y\rightarrow X}, D_X) \\ &\quad - \lambda_X \mathcal{L}_{MLE}(G_{Z\rightarrow X}) - \lambda_Y \mathcal{L}_{MLE}(G_{Z\rightarrow Y}) \end{aligned} \tag{6}$$

where $\lambda_Y, \lambda_Y \geq 0$ are hyperparameters that control the importance of the MLE terms for domains X and Y respectively.

Fig. 1 illustrates the difference between cycleGAN and alignflow methods. Unlikes cycleGAN, the alignflow method is the full invertible architecture that guarantees the cycle-consistency translations between two unpaired domains without an additional $\mathcal{L}_{cycle}$ function.

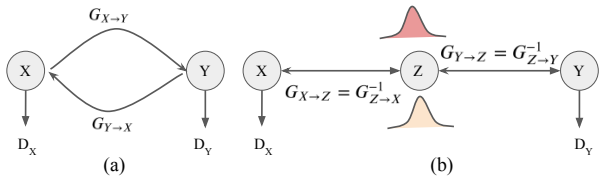

Fig. 1. A comparison between (a) cycleGAN and (b) alignflow generative model. Double-headed arrows denotes an invertible mapping

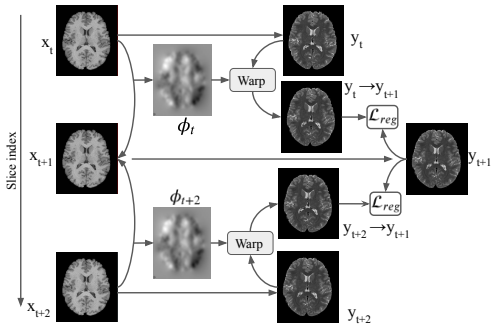

Fig. 2. Deformation Guided Temporal Constraints for domain Y

3 Proposed method

Our motivation is to learn a mapping between unpaired images from different domains by leveraging the temporal information between consecutive slices. We use the temporal information to constrain the mapping between two domains which should be consistent. Our method is an extension of alignflow [5] method with making use of temporal information between consecutive slides.

3.1 Deformation Guided Temporal Constraints

To obtain the displacement between consecutive slices, we use an unsupervised registration network [1] to learn a deformation field ϕ of a slice x_t and its consecutive slices x_k. The deformation field ϕ can be obtained using a convolutional neural network (CNN) [1] by minimizing the loss function

$$\mathcal{L}(\phi) = \|x_t - \big(x_k \circ \phi(x_t, x_k)\big)\|_2 + \|\nabla\phi\|_2 \tag{7}$$

where $\circ$ denotes the spatial transformation operation. The first term ensures that the distance between the next slice x_t and the warped current slice $x_k \circ \phi(.)$ to be close. The second term imposes regularization on $\phi(.)$.

To guarantee the consistency of the image translation, the $\mathcal{L}_1$ loss is used to measure the difference between the warping of fake images on consecutive slice t^{th} and the translation of reference slice k^{th}. We define the temporal consistency loss function for mapping $X \to Y$ and $Y \to X$ as:

$$\begin{aligned} \mathcal{L}_{reg}(X, G_{X\to Y}) &= \sum_{k=0,k\neq t}^{n} \|G_{X\to Y}(x_t) - G_{X\to Y}(x_k) \circ \phi(x_t, x_k)\|_1 \\ \mathcal{L}_{reg}(Y, G_{Y\to X}) &= \sum_{k=0,k\neq t}^{n} \|G_{Y\to X}(y_t) - G_{Y\to X}(y_k) \circ \phi(y_t, y_k)\|_1 \end{aligned} \tag{8}$$

Fig. 2 illustrates an example for image-to-image translation from domain $X \to Y$ using temporal constraints. Let x_t, x_{t+1}, x_{t+2} be consecutive slices of real images

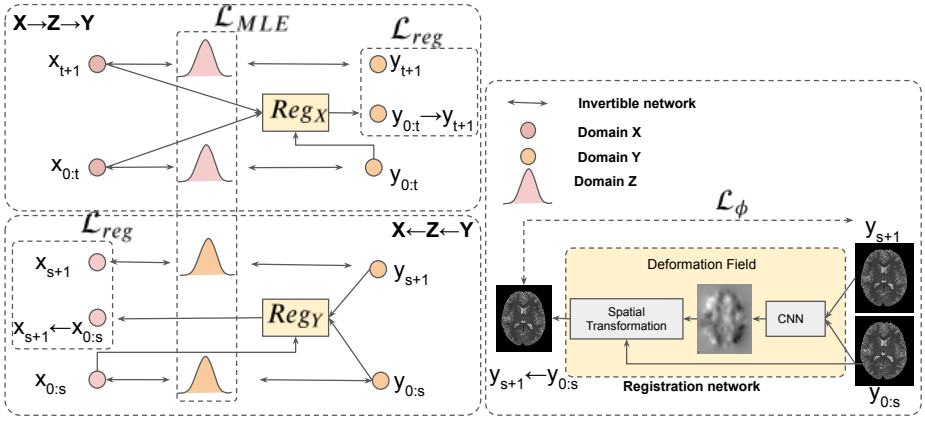

Fig. 3. Our flow-based deformation guidance approach for unpaired image-to-image translation.

in the source domain X. A mapping function $G_{X\to Y}$ generates the fake image y_t, y_{t+1}, y_{t+2} on target domain Y. On the source domain, we can learn displacement fields $\phi_t(.), \phi_{t+2}(.)$ between (x_t, x_{t+1}) and (x_{t+2}, x_{t+1}). To constrain the consistency of the mapping from $X \to Y$, we minimize the distance (i) between the warped fake image $y_t \bigcirc \phi_t(.)$ and y_{t+1} for mapping from t^{th} slice and $(t+1)^{th}$ slice, and (ii) between the warped fake image $y_{t+2} \bigcirc \phi_{t+2}(.)$ and y_{t+1} for mapping from $(t+2)^{th}$ slice and $(t+1)^{th}$ slice.

3.2 Network diagram

Fig. 3 illustrates the proposed network diagram for unpaired image-to-image translation. Our proposed network architecture inherits the advantages of invertible property of alignflow [5]. During training, we add two additional networks Reg_X and Reg_Y for each domain to learn the deformation field $\phi(.)$. These additional networks only use in training time, without increasing the model complexity and inference time comparison with the baseline flow-based method. The temporal constraint via $\mathcal{L}_{reg}(.)$ losses ensures the mapping of consecutive slices on the source domain should be consistent on the target domain. Finally, our objective function is defined as:

$$\begin{aligned}\mathcal{L}_{flow\_reg}(G_{X\to Y}, G_{Y\to X}, D_X, D_Y, \phi) &= \mathcal{L}_{flow}(G_{X\to Y}, G_{Y\to X}, D_X, D_Y) \\ &+ \lambda_1 \mathcal{L}_{reg}(X, G_{X\to Y}) + \lambda_2 \mathcal{L}_{reg}(Y, G_{Y\to X}) + \beta_1 \mathcal{L}_X(\phi) + \beta_2 \mathcal{L}_Y(\phi) \\ &+ \gamma_1 \mathcal{L}_{TV}(X) + \gamma_2 \mathcal{L}_{TV}(Y)\end{aligned} \quad (9)$$

where λ_1, λ_2, β_1, and β_2 control the relative importance of the temporal consistence losses and the two registration losses. $\mathcal{L}_{TV}$ denotes total variation (TV) loss to impose spatial smoothness by measuring the horizontal and vertical gradient of generated images [15]. These TV losses are weighted by γ_1, γ_2.

4 Experimental results

4.1 Datasets and Training

We used common medical datasets to measure the robustness of our method against the existing methods: cycleGAN [16], recycleGAN [2], cycleflow [11] and alignflow [5]. cycleGAN [16] is an unpaired image-to-image translation that works on single slice level. RecycleGAN [2] built upon the cycleGAN and add a temporal predictor that is trained to predict future slice in a set of previous consecutive slices. cycleflow [11] is a flow-based method, but ignores the shared latent space Z (directly map from $X \rightarrow Y$, instead of $X \rightarrow Z \rightarrow Y$ as the alignflow method). The synthetic image from each method was quantitatively compared with the real paired image using the following performance metrics: mean squared error (MSE), peak signal-to-noise ratio (PSNR), and structural similarity index (SSIM).

Human Connectome Project (HCP) is provided by the Human Connectome project [13]. We used T1 as the source domain and T2 as the target domain. We extract the axial view of T1/T2 images into 2D images. We split the 2D images into 1150 images for training set and 500 images for testing set.

MRBrainS13: [8] contains 15 subjects for training and validation and 6 subjects for testing. For each subject, two modalities are available that include T1-weighted, and T2-FLAIR with an image size of $48 \times 240 \times 240$. We extract the dataset into 2D images with 450 images for training and 150 images for testing

Brats2019: [9] includes 210 HGG scans and 75 LGG scans. Each scan has a dimension of $240 \times 240 \times 155$. For each scan, we extract it to 2D images and use 770 images for training and 250 images for testing.

Training All networks were implemented using the Pytorch framework and trained on the 12GB GPU. The input image is resized to 128×128 and normalized to $[-1, 1]$. We used axial slices (10 slices around the middle slice) from the each subject. The Adam optimizer with a batch size of two was used to train the network. The initialization learning rate was set as 0.0002 and was decreased ten times every 20 epochs. We trained each model for 100 epochs. The balance weights were set as $\lambda_X = \lambda_Y = 1e^{-5}, \lambda = \lambda_1 = \lambda_2 = 10, \beta_1 = \beta_2 = 1, \gamma_1 = \gamma_2 = 1$. The discriminator network is a 70×70 PatchGAN [6]. For alignflow network [5], we set the number of scale was 1, number of block was 3. We use two consecutive slices (before and later slices) to learn the temporal constraint.

4.2 Performance Evaluation

Qualitative evaluation Fig. 4 illustrates the image translation on different datasets. The proposed methods (in the last column) provided a better synthetic image, resulting in better MSE, SSIM and PSNR scores. For example, the proposed synthetic T2 image provides a high qualitatively difference along the tumor boundary (indicated by the red arrows in the fifth row) than in existing methods using the available source T1 image as input.

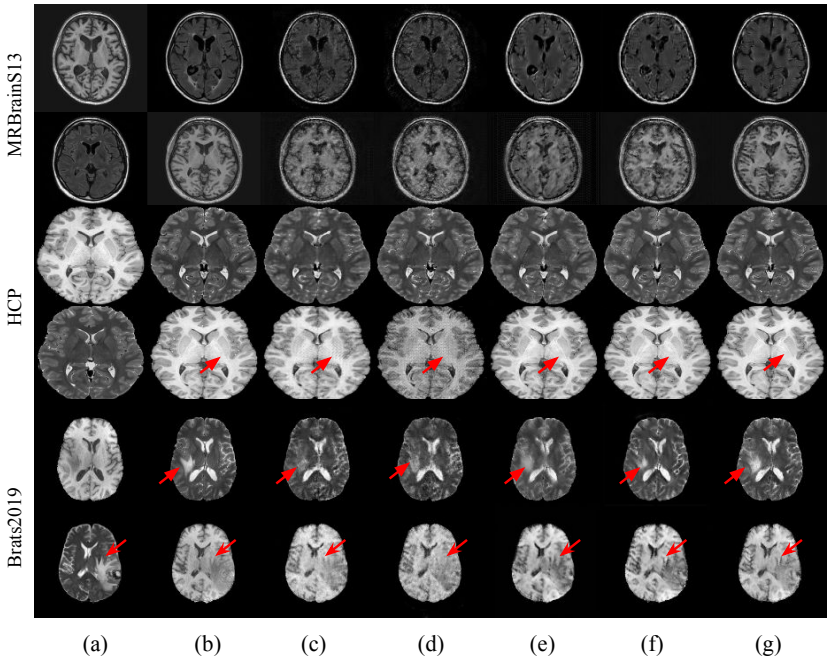

Fig. 4. A visualization of synthetic images on different datasets generated by (a) source image, (b) target image, (c) cycleGAN, (d) recycleGAN, (e) cycleflow, (f) alignflow, and (g) our method. Our method provides a good boundary on the tumor regions (red arrows in the fifth row) compared with the existing methods

Table 1. Comparison between the proposed method against other image-to-image translation methods on **HCP, MRBrainS13, Brats19** datasets.

	Method	**MSE** ↓		**PSNR** ↑		**SSIM** ↑	
		T1 → T2	T2 → T1	T1 → T2	T2 → T1	T1 → T2	T2 → T1
HCP	cycleGAN [16]	0.0193	0.0167	23.2	24.4	0.783	0.793
	recycleGAN [2]	0.0212	0.0182	22.8	24.0	0.773	0.797
	cycleflow [11]	0.0213	0.0189	22.8	23.8	0.771	0.785
	alignflow (baseline) [5]	0.0200	0.0158	23.1	24.6	0.785	0.811
	our method	**0.0179**	**0.0143**	**23.5**	**25.1**	**0.80**	**0.820**
MRBrainS13	cycleGAN [16]	0.0139	**0.0235**	24.7	22.4	0.793	0.704
	recycleGAN [2]	0.0154	0.0250	24.3	22.1	0.761	0.714
	cycleflow [11]	0.0158	0.0406	24.2	20.0	0.790	0.506
	alignflow (baseline) [5]	0.0165	0.0254	24.0	22.0	0.781	0.728
	our method	**0.0128**	0.0236	**25.1**	**22.4**	**0.819**	**0.741**
Brats2019	cycleGAN [16]	**0.0178**	0.0281	**24.1**	22.7	0.833	0.797
	recycleGAN [2]	0.0190	0.0272	23.8	22.6	0.824	0.785
	cyclelow [11]	0.0251	0.0304	22.7	21.8	0.800	0.788
	alignflow (baseline) [5]	0.022	0.0306	23.4	21.8	0.830	0.784
	our method	0.0188	**0.0258**	23.9	**22.8**	**0.842**	**0.808**

Quantitative evaluation Tables 1 reports the MSE, PSNR and SSIM values of the proposed method and existing methods. From the table, it is clear that the flow-based method (such as cycleflow [11], alignflow [5] and our method) provides competitive results with GAN-based method (such as cycleGAN, recycleGAN). By adding temporal constraints, the proposed network outperforms the baseline method (alignflow) on all performance metrics. Different from recycleGAN, that exploits temporal information via future slice prediction from consecutive slices, the proposed method measures pixel-wise temporal consistency by directly warping the synthetic slices with the deformation field of the consecutive slices from the source, and thus achieves better performances. This indicates the effectiveness of the proposed method in the unpaired image to image translation for medical image.

5 Conclusion

We presented an effective method for image-to-image translation based on flow-based methods and deformation information that allows the proposed method to exploit the temporal information between consecutive slices to constrain the translation image. We show that the proposed method can provide a good translation image, yielding a better MSE, PSNR, and SSIM on various MRI datasets. Although our network is a fully invertible property, it requires more memory resource than GAN-based methods (such as cycleGAN, recycleGAN,...).

References

1. Balakrishnan, G., Zhao, A., Sabuncu, M.R., Guttag, J., Dalca, A.V.: An unsupervised learning model for deformable medical image registration. In: Proceedings of the CVPR. pp. 9252–9260 (2018)
2. Bansal, A., Ma, S., Ramanan, D., Sheikh, Y.: Recycle-gan: Unsupervised video retargeting. In: Proceedings of the European conference on computer vision (ECCV). pp. 119–135 (2018)
3. Chen, X., Lian, C., Wang, L., Deng, H., Fung, S.H., Nie, D., Thung, K.H., Yap, P.T., Gateno, J., Xia, J.J., et al.: One-shot generative adversarial learning for mri segmentation of craniomaxillofacial bony structures. IEEE transactions on medical imaging (2019)
4. Dinh, L., Sohl-Dickstein, J., Bengio, S.: Density estimation using real nvp. arXiv preprint arXiv:1605.08803 (2016)
5. Grover, A., Chute, C., Shu, R., Cao, Z., Ermon, S.: Alignflow: Cycle consistent learning from multiple domains via normalizing flows. arXiv preprint arXiv:1905.12892 (2019)
6. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1125–1134 (2017)
7. Kingma, D.P., Dhariwal, P.: Glow: Generative flow with invertible 1x1 convolutions. In: Advances in Neural Information Processing Systems. pp. 10215–10224 (2018)

8. Mendrik, A.M., Vincken, K.L., Kuijf, H.J., Breeuwer, M., Bouvy, W.H., De Bresser, J., Alansary, A., De Bruijne, M., Carass, A., El-Baz, A., et al.: Mrbrains challenge: online evaluation framework for brain image segmentation in 3t mri scans. Computational intelligence and neuroscience **2015** (2015)
9. Menze, B.H., Jakab, A., Bauer, et al.: The multimodal brain tumor image segmentation benchmark (brats). TMI **34**(10), 1993–2024 (2015)
10. van der Ouderaa, T.F., Worrall, D.E.: Reversible gans for memory-efficient image-to-image translation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 4720–4728 (2019)
11. Shen, Z., Zhou, S.K., Chen, Y., Georgescu, B., Liu, X., Huang, T.: One-to-one mapping for unpaired image-to-image translation. In: The IEEE Winter Conference on Applications of Computer Vision. pp. 1170–1179 (2020)
12. Sun, H., Mehta, R., Zhou, H.H., Huang, Z., Johnson, S.C., Prabhakaran, V., Singh, V.: Dual-glow: Conditional flow-based generative model for modality transfer. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 10611–10620 (2019)
13. Van Essen, D.C., Smith, S.M., Barch, D.M., Behrens, T.E., Yacoub, E., Ugurbil, K., Consortium, W.M.H., et al.: The wu-minn human connectome project: an overview. Neuroimage **80**, 62–79 (2013)
14. Welander, P., Karlsson, S., Eklund, A.: Generative adversarial networks for image-to-image translation on multi-contrast mr images-a comparison of cyclegan and unit. arXiv preprint arXiv:1806.07777 (2018)
15. Yuan, Y., Liu, S., Zhang, J., Zhang, Y., Dong, C., Lin, L.: Unsupervised image super-resolution using cycle-in-cycle generative adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. pp. 701–710 (2018)
16. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: IEEE CVPR. pp. 2223–2232 (2017)