frontiers in Neuroscience

ORIGINAL RESEARCH
published: 09 September 2021
doi: 10.3389/fnins.2021.653213

Longitudinal Prediction of Infant MR Images With Multi-Contrast Perceptual Adversarial Learning

*Liying Peng[1,2], Lanfen Lin[1], Yusen Lin[3], Yen-wei Chen[4], Zhanhao Mo[5], Roza M. Vlasova[2], Sun Hyung Kim[2], Alan C. Evans[6], Stephen R. Dager[7], Annette M. Estes[8], Robert C. McKinstry[9], Kelly N. Botteron[9,10], Guido Gerig[11], Robert T. Schultz[12], Heather C. Hazlett[2,13], Joseph Piven[2,13], Catherine A. Burrows[14], Rebecca L. Grzadzinski[2,13], Jessica B. Girault[2,13], Mark D. Shen[2,13,15] and Martin A. Styner[2,16]**

[1] Department of Computer Science, Zhejiang University, Hangzhou, China, [2] Department of Psychiatry, UNC School of Medicine, University of North Carolina, Chapel Hill, NC, United States, [3] Department of Electrical and Computer Engineering Department, University of Maryland, College Park, MD, United States, [4] Department of Information Science and Engineering, Ritsumeikan University, Shiga, Japan, [5] Department of Radiology, China-Japan Union Hospital of Jilin University, Changchun, Jilin, China, [6] Montreal Neurological Institute, McGill University, Montreal, QC, Canada, [7] Department of Radiology, University of Washington, Seattle, WA, United States, [8] Department of Speech and Hearing Sciences, University of Washington, Seattle, WA, United States, [9] Mallinckrodt Institute of Radiology, Washington University School of Medicine, St Louis, MO, United States, [10] Department of Psychiatry, Washington University School of Medicine, St. Louis, MO, United States, [11] Department of Computer Science and Engineering, New York University, New York, NY, United States, [12] Center for Autism Research, Department of Pediatrics, Children's Hospital of Philadelphia, and University of Pennsylvania, Philadelphia, PA, United States, [13] Carolina Institute for Developmental Disabilities, University of North Carolina School of Medicine, University of North Carolina-Chapel Hill, Chapel Hill, NC, United States, [14] Department of Pediatrics, University of Minnesota, Minneapolis, MN, United States, [15] UNC Neuroscience Center, University of North Carolina-Chapel Hill, Chapel Hill, NC, United States, [16] Department of Computer Science, University of North Carolina, Chapel Hill, NC, United States

OPEN ACCESS

Edited by:
Federico Giove,
Centro Fermi - Museo storico della fisica e Centro studi e ricerche Enrico Fermi, Italy

Reviewed by:
Young Don Son,
Gachon University Gil Hospital, South Korea
Viktor Vegh,
The University of Queensland, Australia

***Correspondence:**
Martin A. Styner
styner@unc.edu

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: 14 January 2021
Accepted: 09 August 2021
Published: 09 September 2021

Citation:
Peng L, Lin L, Lin Y, Chen Y-w, Mo Z, Vlasova RM, Kim SH, Evans AC, Dager SR, Estes AM, McKinstry RC, Botteron KN, Gerig G, Schultz RT, Hazlett HC, Piven J, Burrows CA, Grzadzinski RL, Girault JB, Shen MD and Styner MA (2021) Longitudinal Prediction of Infant MR Images With Multi-Contrast Perceptual Adversarial Learning. Front. Neurosci. 15:653213. doi: 10.3389/fnins.2021.653213

The infant brain undergoes a remarkable period of neural development that is crucial for the development of cognitive and behavioral capacities (Hasegawa et al., 2018). Longitudinal magnetic resonance imaging (MRI) is able to characterize the developmental trajectories and is critical in neuroimaging studies of early brain development. However, missing data at different time points is an unavoidable occurrence in longitudinal studies owing to participant attrition and scan failure. Compared to dropping incomplete data, data imputation is considered a better solution to address such missing data in order to preserve all available samples. In this paper, we adapt generative adversarial networks (GAN) to a new application: longitudinal image prediction of structural MRI in the first year of life. In contrast to existing medical image-to-image translation applications of GANs, where inputs and outputs share a very close anatomical structure, our task is more challenging as brain size, shape and tissue contrast vary significantly between the input data and the predicted data. Several improvements over existing GAN approaches are proposed to address these challenges in our task. To enhance the realism, crispness, and accuracy of the predicted images, we incorporate both a traditional voxel-wise reconstruction loss as well as a perceptual loss term into the adversarial learning scheme. As the differing contrast changes in T1w and T2w MR images in the first year of life, we

incorporate multi-contrast images leading to our proposed 3D multi-contrast perceptual adversarial network (MPGAN). Extensive evaluations are performed to assess the qualityand fidelity of the predicted images, including qualitative and quantitative assessments of the image appearance, as well as quantitative assessment on two segmentation tasks. Our experimental results show that our MPGAN is an effective solution for longitudinal MR image data imputation in the infant brain. We further apply our predicted/imputed images to two practical tasks, a regression task and a classification task, in order to highlight the enhanced task-related performance following image imputation. The results show that the model performance in both tasks is improved by including the additional imputed data, demonstrating the usability of the predicted images generated from our approach.

Keywords: generative adversarial networks, MRI, longitudinal prediction, machine learning, infant, postnatal brain development, autism, imputation

1. INTRODUCTION

The early postnatal period (neonate to one year of age) is a period of dynamic and rapid brain development with dramatic appearance changes in magnetic resonance images (MRI). This period has been associated with early atypical developmental trajectories in neurodevelopmental disorders, such as autism spectrum disorder (ASD) and schizophrenia (Hazlett et al., 2017; Gilmore et al., 2018). Longitudinal MRI allows the quantification of developmental trajectories over time and plays a critical role in neuroimaging studies of early brain development (Gilmore et al., 2012). However, missing data points is a common issue in longitudinal studies due to MRI scan failure, scheduling issues, or general participant attrition (Laird, 1988). Discarding those study participants with incomplete data significantly reduces the sample size and may even lead to unacceptable levels of bias (Matta et al., 2018). One solution to deal with the issue of missing data is to interpolate/extrapolate the missing data, called data imputation, from the data that is available. Such data imputation can be performed either at the image or at the measurement level.

Low rank matrix completion is a commonly proposed approach for measurement level imputation, for example, Thung et al. (2016) employed it to impute the missing volumetric features. A series of machine learning based approaches also have been proposed in this field, such as Meng et al. (2017) proposed the Dynamically-Assembled Regression Forests (DARF) in order to predict cortical thickness maps at missing time points. Rekik et al. (2016) developed a 4D varifold-based learning framework to predict the cortical shape at the time point in the first year of life using the cortical surface shape at birth. Additionally, a lot of variants of geodesic models (Fishbaugh et al., 2013, 2014; Fletcher, 2013; Singh et al., 2013a) were proposed for longitudinal shape imputation and regression. Compared with measurement-level methods, image-level methods directly predict the image appearance at a missing time point. In (Niethammer et al., 2011; Singh et al., 2013b), the geodesic models were used for longitudinal regression of related image appearance. And Rekik et al. (2015) proposed a sparse patch-based metamorphosis learning framework for regression of MRI appearance and anatomical structures with promising yet limited results.

In this paper, we focus on image-level approaches for infant longitudinal MRI prediction and we treat it as an image synthesis problem, i.e., synthesizing/predicting a missing MR image from an existing image of the same subject at a later or earlier time point. Recently, generative adversarial networks (GANs) have shown great potential in generating visually-realistic images for both natural image synthesis, e.g., in image-to-image translation (Isola et al., 2017; Liu et al., 2017; Yi et al., 2017; Zhu et al., 2017; Huang et al., 2018; Xiong et al., 2019; Emami et al., 2021), generating new plausible samples (Goodfellow et al., 2014; Zhang et al., 2019), generating photographs of human faces (Karras et al., 2017), and medical image synthesis, e.g., cross-modality synthesis (MR-to-CT Nie et al., 2017; Wolterink et al., 2017; Jin et al., 2019, MR-to-PET Pan et al., 2018, 2019, PET-to-MR Choi and Lee, 2018, CT-to-PET Ben-Cohen et al., 2017; Bi et al., 2017; Armanious et al., 2020, 3T-to-7T Qu et al., 2019), cross-site synthesis (Zhao et al., 2019), and multi-contrast MRI synthesis (Dar et al., 2019; Yang et al., 2020). Recently, GANs have also been applied to longitudinal MR image prediction. For example, Xia et al. (2019) proposed a conditional GAN that conditioned on age and health state (status of Alzheimer's Disease) to predict brain aging trajectories. In (Bowles et al., 2018; Ravi et al., 2019), a GAN is used to predict the Alzheimer's related brain degeneration from existing MR images, where biological constraints associated with disease progression are integrated into the framework. These longitudinal prediction approaches are limited to 2D T1w MRI that are hard to generalize to 3D, as well as having been designed for adult brain images related to Alzheimer's disease. Besides, we also notice that GAN-architectures with perceptual loss have been used in a few medical image applications. For example, Armanious et al. applied GAN with perceptual loss to PET-CT translation, MR motion correction and the PET denoising (Armanious et al., 2020). Dar et al. used a GAN with VGGNet-based perceptual loss for a multi-contrast MRI synthesis task (mapping among T1w MRI and T2w MRI) (Dar et al., 2019). Due to the nature

of their tasks, they only focus on single modality 2D data. Also, they utilized VGGNet for the perceptual loss computation. Since VGGNet is a pretrained model based on 2D natural images, it may be not that appropriate for medical image tasks.

In this work, we propose a novel GAN adaptation for a new application: the longitudinal prediction of infant MR images in the first year of life. Since human brain size and shape changes rapidly in the first year of life, 2D methods are not suitable in our task. Thus, we present a fully 3D-based approach for the prediction of infant MR images. In addition, because of the myelination process, the infant brain shows a dramatic change of tissue contrast and anatomical structural shape, which further poses difficulties for prediction. While generative adversarial networks can produce images with realistic textures by enforcing the outputs from the generator to be close to the real data distribution, it cannot ensure the consistency between the outputs and the desired ground-truth images, so that the appearance of predicted image may look different from the ground-truth image. To handle the large variation in appearance, we add a voxel-wise reconstruction constraint, i.e., an L1 loss, to explicitly guide the generator to produce images that match ground-truth images at the voxel level. Although global structures can be well-preserved by harnessing L1 loss, it often results in an over-smoothed output (Pathak et al., 2016). Hence, to alleviate this issue, we also enhance our GAN with a perceptual loss term to maintain appearance consistency at the feature level. We propose to utilize Model Genesis (Zhou et al., 2019), which is a pre-trained model for 3D medical images, for this feature extraction. Finally, in order to tackle the reduced tissue contrast during the first year of life, particularly at about 6 months of age, we propose a multi-contrast framework, so that the complementary information of different contrasts (T1w and T2w images) can be exploited. The source code of our method will be released to the public upon acceptance of this manuscript at https://github.com/liying-peng/MPGAN.

Our main contributions are summarized as follows:

- To the best of our knowledge, this is the first application of deep generative methods for longitudinal prediction of structural MRI in the first year of life.
- Unlike previous 2D-based methods, our method is based on a 3D MRI prediction, where the volumetric spatial information is fully considered.
- To predict sharp, realistic and accurate images, we adopt a GAN with adversarial, pixel-wise reconstruction and perceptual loss. The perceptual loss is computed via features extracted from an application-specific model that has been pre-trained on 3D medical images.
- To leverage complementary information from multi-contrast data, we propose a novel multi-contrast framework to jointly predict T1w and T2w images.
- Extensive experiments demonstrate the effectiveness of our approach for use in longitudinal MRI prediction and imputation in the developing infant brain. We show that when using these imputed MR images to expand the training data in two practical machine learning tasks, we improve the model performance.

The remainder of this paper is organized as follows. In section 2, we introduce the experimental datasets and describe the methodological details of our proposed approach. Experimental results are presented in section 3 and then discussed in section 4. The conclusions are shown in section 5.

2. METHODS

In this section, we first introduce a brief background of generative adversarial networks. Subsequently, we formulate our problem and then define our objective functions. Finally, our network architectures are discussed.

2.1. Generative Adversarial Network

The Generative Adversarial Network (GAN) is a generative deep learning model that was proposed by Goodfellow et al. (2014). The aim of it on image-to-image translation tasks is to learn a mapping from the input image x to the target image y, i.e., $x \rightarrow y$. It consists of two separate components, each a neural network, specifically a generator G and a discriminator D network. In the training stage, these two networks compete with each other, where a) G attempts to fool D by generating a fake image $G(x)$ that looks similar to a real target image y, and b) D aims to distinguish between the real image y and the fake image $G(x)$. As the two networks face off, $G(x)$ generates more realistic images that get closer to the real data distribution and $D(x)$ becomes more skilled at differentiating images. At the end, the algorithm will converge to a Nash equilibrium (Nash et al., 1950). This two-player *minmax* game is formulated as: $\min_G \max_D \mathcal{L}_{adv}(G, D)$, where the adversarial loss $\mathcal{L}_{adv}$ can be defined as.

$$\mathcal{L}_{adv}(G, D) = \mathbb{E}_x[(1 - D(G(x)))^2] + \mathbb{E}_y[D(y)^2] \quad (1)$$

2.2. Objective Design

We consider two settings in this work: (1) a single-input-single-output setting when using single contrast images and (2) a multi-input-multi-output setting when using multiple contrasts jointly. In the former setting, suppose $\{x_i, y_i\}_{i=1}^N$ is a series of paired instances, where x_i is a T1w or a T2w image at age a_1, y_i is the corresponding T1w or T2w image at age a_2 and N is the number of paired subjects in the training set. Our goal is to learn the mapping $G: x \rightarrow y$. In the latter setting, assume $\{x_i^{T1}, x_i^{T2}, y_i^{T1}, y_i^{T2}\}_{i=1}^N$ is a set of paired subjects, where x_i^{T1}, x_i^{T2} indicate the T1w and T2w images at age a_1 and y_i^{T1}, y_i^{T2} stand for the corresponding T1w and T2w images at age a_2. The aim is then to learn two mapping functions: $G_{T1}: \{x^{T1}, x^{T2}\} \rightarrow y^{T1}$ and $G_{T2}: \{x^{T1}, x^{T2}\} \rightarrow y^{T2}$.

2.2.1. Adversarial Loss

In the single-input-single-output setting, in order to learn the mapping $G: x \rightarrow y$, we can employ the adversarial loss function of the original GAN (see Equation 1). In the multi-input-multi-output setting, the basic idea is same as the original GAN, but here we define two generators, i.e., $G_{T1}: \{x^{T1}, x^{T2}\} \rightarrow y^{T1}$ and $G_{T2}: \{x^{T1}, x^{T2}\} \rightarrow y^{T2}$. G_{T1} and G_{T2} aim at generating fake T1w and T2w images that look similar as real images, respectively. We also define two discriminators D_{T1} and D_{T2}, where the

intention of D_{T1} is to differentiate the real T1w image y^{T1} from the generated T1w image $G_{T1}(x^{T1}, x^{T2})$. Similarly, D_{T2} attempts to distinguish between y^{T2} and $G_{T2}(x^{T1}, x^{T2})$. With respect to generator G_{T1} and its discriminator D_{T1}, the adversarial loss can be formulated as

$$\begin{aligned}\mathcal{L}_{adv}(G_{T1}, D_{T1}) = \mathbb{E}_{x^{T1},x^{T2}}[(1 - D_{T1}(G_{T1}(x^{T1}, x^{T2})))^2] \\ +\mathbb{E}_{y^{T1}}[D_{T1}(y^{T1})^2]\end{aligned} \tag{2}$$

The adversarial loss $\mathcal{L}_{adv}(G_{T2}, D_{T2})$ can be expressed similarly.

2.2.2. Voxel-Wise Reconstruction Loss

While the adversarial loss can optimize the generated output images closer to the real data distribution, it cannot ensure consistency between the outputs and the desired ground-truth images, so that a predicted image may not share the details of its corresponding ground-truth image. To deal with this problem, we further restrict the generator with a voxel-wise reconstruction loss. Here we choose a traditional L1 loss, as recommended in Zhao et al. (2015), which directly penalizes the voxel-wise differences between the two images. For the single-input-single-output setting, the voxel-wise reconstruction loss is given by

$$\mathcal{L}_{vr}(G) = \mathbb{E}_{x,y}[\|y - G(x)\|_1] \tag{3}$$

For the multi-input-multi-output setting, the voxel-wise reconstruction loss is expressed as.

$$\begin{aligned}\mathcal{L}_{vr}(G_{T1}, G_{T2}) = \mathbb{E}_{x^{T1},x^{T2},y^{T1}}[\|y^{T1} - G_{T1}(x^{T1}, x^{T2})\|_1] \\ +\mathbb{E}_{x^{T1},x^{T2},y^{T2}}[\|y^{T2} - G_{T2}(x^{T1}, x^{T2})\|_1]\end{aligned} \tag{4}$$

2.2.3. Perceptual Loss

Although the voxel-wise reconstruction loss enforces voxelwise consistency between the real and generated images, it prefers an over-smoothed solution (Pathak et al., 2016). In other words, this loss commonly leads to outputs with well-preserved low-frequency information, e.g., global structures, at the expense of the high-frequency crispness. To alleviate this problem, we add a perceptual loss (Johnson et al., 2016) to the generator, which results in sharper images. The perceptual loss calculates the difference between two images in feature space in place of voxel space. Thus, it forces the generated images to be perceptually similar to the real images, instead of matching intensities exactly at the voxel level. Suppose $\phi_m(x)$ is the output from the m-th layer of a feature extractor ϕ when processing the image x. For the single-input-single-output setting, the perceptual loss can be written as

$$\mathcal{L}_p(G) = \mathbb{E}_{x,y}[\|\phi_m(y) - \phi_m(G(x))\|_1] \tag{5}$$

For the multi-input-multi-output setting, the perceptual loss is formulated as

$$\begin{aligned}\mathcal{L}_p(G_{T1}, G_{T2}) = \mathbb{E}_{x^{T1},x^{T2},y^{T1}}[\|\phi_m(y^{T1}) - \phi_m(G_{T1}(x^{T1}, x^{T2}))\|_1] \\ + \mathbb{E}_{x^{T1},x^{T2},y^{T2}}[\|\phi_m(y^{T2}) - \phi_m(G_{T2}(x^{T1}, x^{T2}))\|_1]\end{aligned} \tag{6}$$

Overall objective: By combining the above loss functions, we can define the final objective in the single-input-single-output setting as

$$\min_G \max_D \mathcal{L}(G, D) = \mathcal{L}_{adv}(G, D) + \alpha\mathcal{L}_{vr}(G) + \beta\mathcal{L}_p(G) \tag{7}$$

Similar, we can define the total objective in the multi-input-multi-output setting as

$$\begin{aligned}\min_{G_{T1},G_{T2}} \max_{D_{T1},D_{T2}} \mathcal{L}(G_{T1}, G_{T2}, D_{T1}, D_{T2}) = \mathcal{L}_{adv}(G_{T1}, D_{T1}) \\ +\mathcal{L}_{adv}(G_{T2}, D_{T2}) + \alpha\mathcal{L}_{vr}(G_{T1}, G_{T2}) + \beta\mathcal{L}_p(G_{T1}, G_{T2})\end{aligned} \tag{8}$$

where α and β are the coefficients to weight the loss contributions.

2.3. Network Architectures

2.3.1. Perceptual Adversarial Network

Figure 1A illustrates the architecture of the perceptual adversarial network (PGAN) that is designed for the single-input-single-output setting. It consists of a generator G, a discriminator D and a feature extractor ϕ. We utilize a traditional 3D-Unet (Çiçek et al., 2016) as generator. The 3D-Unet is an end-to-end convolutional neural network that was originally developed for medical image segmentation. It includes an analysis path (encoder) and a synthesis path (decoder). The encoder part contains four convolutional layers each of which includes two repeated $3 \times 3 \times 3$ convolution operations, followed by a $2 \times 2 \times 2$ max pooling for downsampling (except for the last layer). The decoder part is basically the same as the encoder part, but it replaces all downsampling with upsampling. Skip-connections are build between the layers of the encoder and their counterparts of the decoder. The architecture of our discriminator D is the same as the one in (Isola et al., 2017). It contains four stride-2 convolutional layers, with 64, 128, 256, and 512 channels, respectively. The output layer of it is a stride-1 convolutional layer with one channel, followed by a *sigmoid* activation function. Instance normalization (Ulyanov et al., 2016) is applied to the convolutional layers in both generator and discriminator.

For natural images, pretrained VGG networks are often adopted as the feature extractors. However, for medical image tasks, feature extractors based on pretrained VGG-Nets have the following limitations: (1) 3D medical images would have to be reformulated into a 2D format to fit VGG-Nets (2D networks), leading to the loss of rich 3D anatomical information. (2) Perceptual differences between natural projection images (such as photos) and 3D tomographic, medical images, are not captured. To overcome these limitations, we employ an existing application-specific model as our feature extractor ϕ, specifically Model Genesis (Zhou et al., 2019), which is built directly from 3D medical images. Note that Model Genesis is a U-Net style network and here we only use its encoder part for feature extraction. The details of PGAN are shown in **Table 1**.

2.3.2. Multi-Contrast Perceptual Adversarial Network

As shown in **Figure 1C**, the multi-contrast perceptual adversarial network (MPGAN) contains two generators G_{T1} and G_{T2}, two

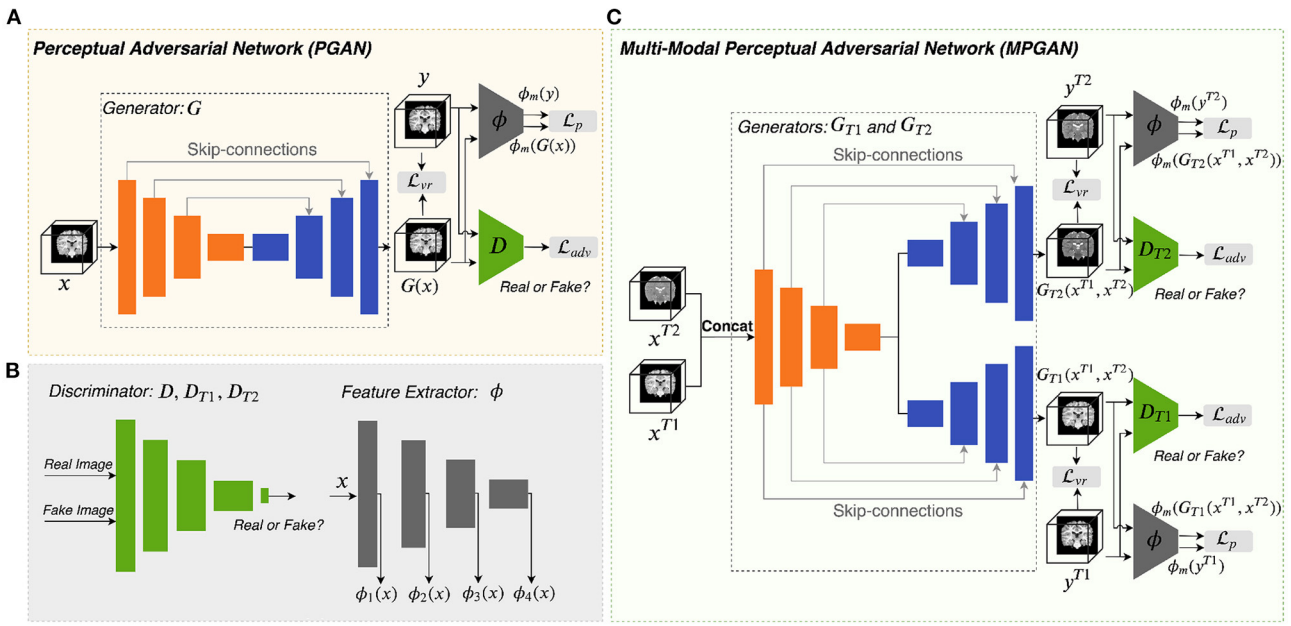

FIGURE 1 | Overview of the proposed methods. Panels **(A,C)** are the architectures of PGAN and MPGAN, respectively. Panel **(B)** shows the discriminator and the feature extractor setup for the perceptual loss computation.

TABLE 1 | The network architecture of perceptual adversarial network.

Generator	
Encoder network	**Decoder (generator) network**
(3 × 3 × 3) 20 Conv, IN, RL	(2 × 2 × 2) ↑, (3 × 3 × 3) 160 Conv, IN, RL
(3 × 3 × 3) 40 Conv, IN, RL, (2 × 2 × 2) ↓	(3 × 3 × 3) 160 Conv, IN, RL
(3 × 3 × 3) 40 Conv, IN, RL	(2 × 2 × 2) ↑, (3 × 3 × 3) 80 Conv, IN, RL
(3 × 3 × 3) 80 Conv, IN, RL, (2 × 2 × 2) ↓	(3 × 3 × 3) 80 Conv, IN, RL
(3 × 3 × 3) 80 Conv, IN, RL	(2 × 2 × 2) ↑, (3 × 3 × 3) 40 Conv, IN, RL
(3 × 3 × 3) 160 Conv, IN, RL, (2 × 2 × 2) ↓	(3 × 3 × 3) 40 Conv, IN, RL
(3 × 3 × 3) 160 Conv, IN, RL	(1 × 1 × 1) 1 Conv, tanh
(3 × 3 × 3) 320 Conv, IN, RL	
Discriminator	
(4 × 4 × 4) 64 stride 2 Conv, LR, (4 × 4 × 4) 128 stride 2 Conv, IN, LR	
(4 × 4 × 4) 256 stride 2 Conv, IN, LR, (4 × 4 × 4) 512 stride 2 Conv, IN, LR	
(4 × 4 × 4) 1 stride 1 Conv, sigmoid	
Feature extractor	
(3 × 3 × 3) 32 Conv, IN, RL, (3 × 3 × 3) 64 Conv, IN, RL, (2 × 2 × 2) ↓	
(3 × 3 × 3) 64 Conv, IN, RL, (3 × 3 × 3) 128 Conv, IN, RL, (2 × 2 × 2) ↓	
(3 × 3 × 3) 128 Conv, IN, RL, (3 × 3 × 3) 256 Conv, IN, RL, (2 × 2 × 2) ↓	
(3 × 3 × 3) 256 Conv, IN, RL, (3 × 3 × 3) 512 Conv, IN, RL	

Conv, convolution; IN, Instance Normlization; RL, ReLu; LR, Leaky ReLU; ↓ and ↑, represent down- and upsampling, respectively; sigmoid, sigmoid activation function; tanh, tanh activation function.

discriminators D_{T1} and D_{T2} and one feature extractor ϕ. The feature extractor ϕ and the architectures of D_{T1} and D_{T2} are the same as for PGAN. G_{T1} and G_{T2} are both based on 3D-Unets that utilize a shared encoder and two independent decoders with skip-connections. The shared encoder learns complementary information from both T1w and T2w images

and skip connections are used to transfer this information from the shared encoder to different decoders. We combine T1w and T2w images before feeding them into generators by applying a channel-wise concatenation.

3. EXPERIMENTS AND RESULTS

3.1. Materials

The data used in this work is collected from the "Infant Brain Imaging Study" (IBIS) database (https://www.ibis-network.org) and the raw MR images are available on NDA (https://nda.nih.gov). All MR images were clinically evaluated by an expert neuroradiologist (RCM) and subjects with visible clinical pathology were excluded from the study. Data collection sites had approved study protocols by their Institutional Review Boards (IRB), and all enrolled subjects had informed consent provided by their parent/guardian. MR imaging parameters are as follows: (1) 3T Siemens Tim Trio at 4 sites; (2) T1w MRI: TR/TE = 2,400/3.16 ms, 256 × 256 × 160, 1 mm^3 resolution; (3) T2w MRI: TR/TE = 3,200/499 ms, same matrix and resolution as T1w. A series of preprocessing steps were adopted, i.e., ICBM alignment, bias correction, geometry correction, skull stripping (see Hazlett et al., 2017 for details), and intensity normalization to range (−1,1). For our main dataset which is used for longitudinal prediction, a total of 289 subjects with two complete scans at 6 and 12 months were selected. The dataset was split into three sets: training set (231 subjects), validation set (29 subjects), and test set (29 subjects).

We also build two additional datasets to evaluate the applicability of our predicted/imputed images. In the first application, we aim at classifying subject image data into different Autism Diagnostic Observation Schedule social affect (ADOS-SA-CSS) based groups. Thus, only those subjects with valid ADOS-SA-CSS measures were employed. In addition, we reduced the size of the typical developing group for group size balancing. This resulted in 77 subjects with complete scans at 6 and 12 months and 103 subjects with scans either at 6 or 12 months. In the second application, we estimated a subject's gestational age (GA) at birth from its MRI data. Only subjects with known GA were selected and this resulted in 134 subjects with complete scan pairs at 6 and 12 months, as well as 76 subjects with scans either at 6 or 12 months. In both applications, we employ the imputed datasets as additional training data. No imputed images are used in the testing datasets.

3.2. Implementation Details

Our experiments were performed on a lambdalab GPU server with four NVIDIA TITAN RTX GPU with 24GB oncard memory. All the networks were implemented in Tensorflow and trained via Adam optimization (Kingma and Ba, 2014). The batch size was set to 1. The learning rate was initially set to 2e-4 for the first 44 epochs and decayed every 22 epochs with a base of 0.5 for an additional 176 epochs. The trade-off parameters α and β in Equation (7) and (8) were set to 25 and $\phi_1(x)$ was used for computation of the perceptual loss based on a grid search. The details of the grid search are shown in the **Supplementary Materials**. Two longitudinal prediction tasks were performed in this work, i.e., prediction of 6-month images from 12-month images and prediction of 12-month images from 6-month images.

3.3. Alternative Networks for Comparison

In this paper, we also trained five additional networks for the purpose of comparison: (1) CycleGAN: 3D extension of original CycleGAN (Zhu et al., 2017). (2) Unet($\mathcal{L}_{vr}$): 3D-Unet (Çiçek et al., 2016) trained with $\mathcal{L}_{vr}$. (3) Unet ($\mathcal{L}_{vr} + \mathcal{L}_p$): 3D-Unet trained with both $\mathcal{L}_{vr}$ and $\mathcal{L}_p$. (4) GAN: original GAN (Goodfellow et al., 2014). (5) GAN+$\mathcal{L}_{vr}$: original GAN with additional $\mathcal{L}_{vr}$ term. To enable fair comparisons, we implemented these networks with parameters optimized the same way as our proposed methods. Further, the 3D-Unet was used as the backbone of Unet variant methods, i.e., (2) and (3), and it was also used as the generator of cycleGAN, GANs and our methods. The discriminators for (1), (4), and (5) are the same as for our models.

3.4. Evaluation via Appearance Based Metrics

In this section, images predicted by different methods are evaluated both in qualitative and quantitative fashion, focusing on the image appearance. In addition, we conducted a human perceptual study, where the participants were required to rate the predicted images based on visual realism and closeness to the ground truth images.

3.4.1. Qualitative Results

The qualitative results for different methods are given in **Figure 2** (6-to-12 months prediction task) and **Figure 3** (12-to-6 months prediction task). The following findings were obtained from both tasks. (1) The images predicted by Unet($\mathcal{L}_{vr}$) and GAN+$\mathcal{L}_{vr}$ are globally consistent with the ground-truth images, but they appear overly smoothed, resulting in a poor visual quality. (2) Unet($\mathcal{L}_{vr} + \mathcal{L}_p$) outperforms Unet($\mathcal{L}_{vr}$) with the resultant images showing more high-frequency details. This indicates that adding the perceptual loss $\mathcal{L}_p$ into training process helps the model to produce sharper details. However, the visual quality of the images generated by Unet($\mathcal{L}_{vr} + \mathcal{L}_p$) is still unsatisfactory due to unrealistic textured appearance. (3) GAN produces the least anatomically accurate images, albeit with sharp details. This may be due to the reason that GAN is trained without any additional constraints to enforce appearance consistency between ground-truth and generated images. (4) Our PGAN and MPGAN show a superior performance compared with the other methods. They produce more realistic images with sharp and refined details from a visual perspective. (5) Compared to PGAN, MPGAN can predict finer details, especially for T2w images. This implies that multi-contrast learning can further improve the image quality by combining complementary information from T1w and T2w images.

3.4.2. Quantitative Results

The development of optimal evaluation metrics for generated images is an challenging problem. Recently, a new learning-based metric, i.e., Learned Perceptual Image Patch Similarity (LPIPS), was proposed (Zhang et al., 2018) to assess the similarity between

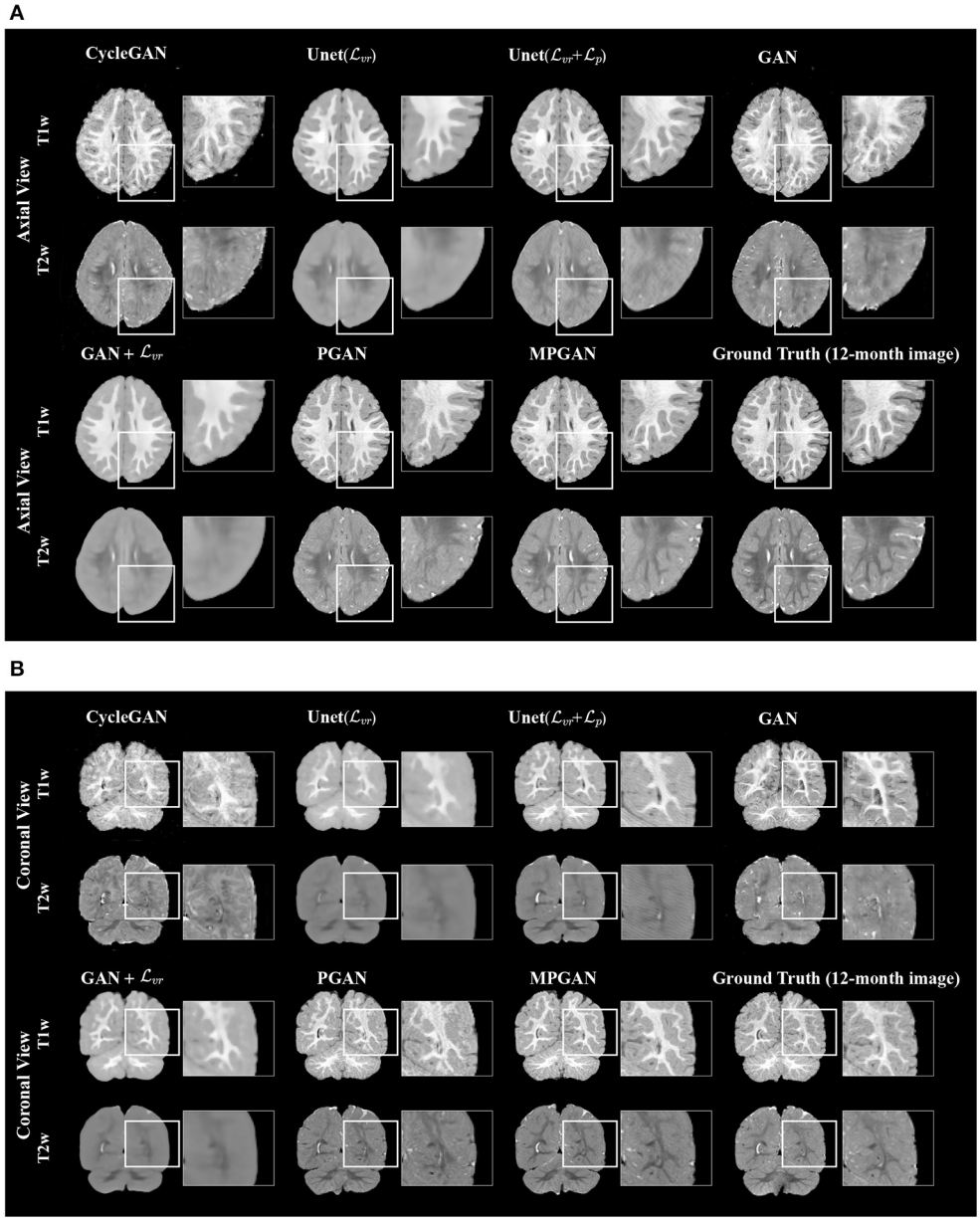

FIGURE 2 | Examples of predicted MR images at 12 months (from 6 months MRI) compared across seven methods and the corresponding ground truth. **(A)** Axial view. **(B)** Coronal view.

two images, which shows a superior performance compared to the traditional metrics. In this section, all of the methods are quantitatively compared based on LPIPS, which is shown in **Figure 4**. Note that LPIPS is a "similarity distance" calculated between the ground-truth image and the predicted image and lower value reflects a higher similarity. One can see that our PGAN and MPGAN give a notable improvement of LPIPS compared to other approaches, for both 6-to-12 months and 12-to-6 months prediction tasks. Specifically, MPGAN achieves the best performance. Paired t-tests showed statistically significant improvements ($p < 0.05$) of MPGAN over all other methods.

3.4.3. Human Perceptual Study

We performed a perceptual study based on 116 sets of images, including 29 sets of 6-month T1w images, 29 sets of 6-month T2w images, 29 sets of 12-month T1w images, and 29 sets of 12-month T2w images. For each image set, the ground-truth image and the predicted images of seven different methods were shown to human raters for visual assessment. We asked 22 human raters (6 radiologists, 5 neuroscientists, 3 biomedical researchers, and 8 computer scientists with medical imaging background) to rate the image quality of the predicted images using a 7-point score, with 7 being the most realistic and

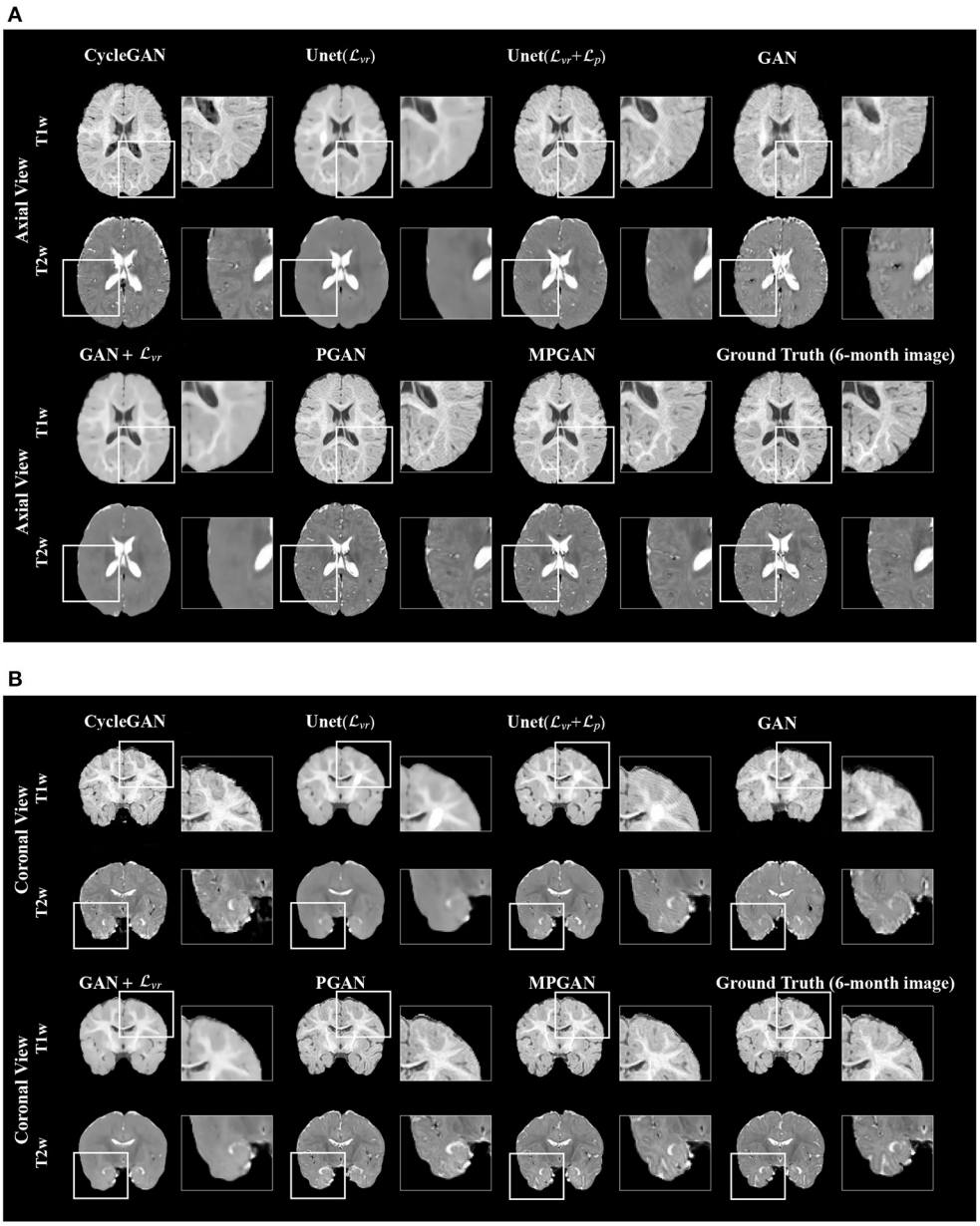

FIGURE 3 | Examples of predicted MR images at 6 months (from 12 months MRI) compared across seven methods and the corresponding ground truth. **(A)** Axial view. **(B)** Coronal view.

closest to the ground-truth image (ties are allowed). All the images were shown initially in a random order and presented in axial, coronal, and sagittal views. The visualization order was continuously updated by sorting according to the current scores. The results of the perceptual study are shown in **Table 2**. Of all the studied methods, our MPGAN achieves the highest quality score across different images, with a statistical significance in Wilcoxon signed-rank test ($p < 0.05$ vs. other methods). The second-best performance is yielded by PGAN ($p < 0.05$ vs. other methods). While MPGAN are PGAN are close for T1w image prediction, MPGAN outperforms PGAN by a large margin for predicting T2w images (both 6 and 12 months), demonstrating the benefits of the multi-contrast architecture.

3.5. Evaluation on Segmentation Task

In this section, we assess the quality of the predicted images in two segmentation tasks. We conducted subcortical and tissue segmentation on both predicted and ground-truth images at 12 months using an existing multi-atlas segmentation method (Wang et al., 2014). For the tissue segmentation task, the brain

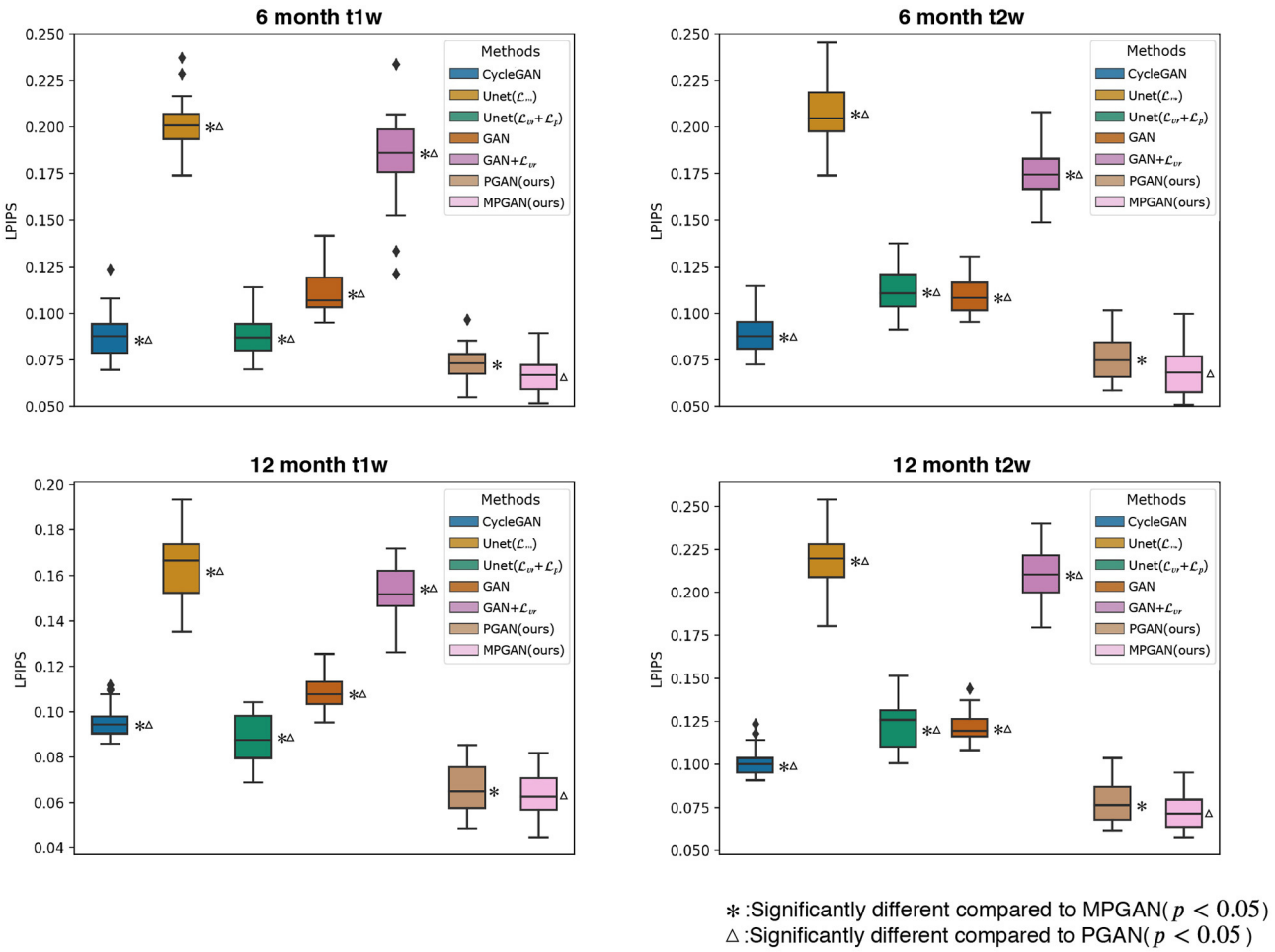

FIGURE 4 | Quantitative comparison results of different methods based on Learned Perceptual Image Patch Similarity (LPIPS). A lower LPIPS value reflects a higher similarity between ground-truth and predicted images.

TABLE 2 | The average score results of human assessments based on the appearance of images predicted by different methods.

Method	Unet($\mathcal{L}_{vr}$)	GAN+$\mathcal{L}_{vr}$	GAN	Unet($\mathcal{L}_{vr}+\mathcal{L}_p$)	CycleGAN	PGAN	MPGAN
6-month T1w	1.48 ± 0.64	2.18 ± 0.68	2.58 ± 1.09	3.44 ± 1.11	5.24 ± 0.91	6.13 ± 0.84	**6.43 ± 0.72**
6-month T2w	1.87 ± 0.80	2.14 ± 0.78	3.29 ± 1.54	3.32 ± 0.84	5.34 ± 0.82	5.96 ± 0.76	**6.59 ± 0.67**
12-month T1w	2.83 ± 1.27	2.15 ± 1.46	3.68 ± 1.75	4.20 ± 1.36	3.92 ± 1.64	6.22 ± 0.71	**6.64 ± 0.64**
12-month T2w	2.13 ± 1.07	1.73 ± 0.88	2.63 ± 1.29	3.29 ± 1.06	4.01 ± 1.40	5.89 ± 0.64	**6.92 ± 0.27**

The scores range from 1 to 7, with higher scores indicating a higher degree of realism and closer appearance to the ground truth. Methods with best performance are bolded for each setting (significantly better than other measurements, $p < 0.5$).

was segmented into four types of tissue, i.e., white matter, cortical gray matter, deep gray matter, and cerebrospinal fluid (CSF). For the subcortical segmentation task, 12 subcortical structure labels were computed: left and right hemispheric caudate, putamen, pallidum, thalamus, amygdala, and hippocampus. Examples of tissue and subcortical segmentation results are shown in **Figures 5**, **6**, respectively. Our quantitative evaluation is based on following three metrics that measure the segmentation similarity between two segmentation results (S_1 and S_2): relative absolute volume difference (AVD, in %), average symmetric surface distance (ASD, in *mm*), and Dice coefficient. The relative absolute volume difference is as

$$AVD = \frac{|V_{S_1} - V_{S_2}|}{V_{S_1}} \times 100\% \tag{9}$$

where V_{S_1} is the volume of S_1 and V_{S_2} can be defined similarly. Suppose B_{S_1} and B_{S_2} are the borders of S_1 and S_2, respectively. The average symmetric surface distance (ASD) (Van Ginneken

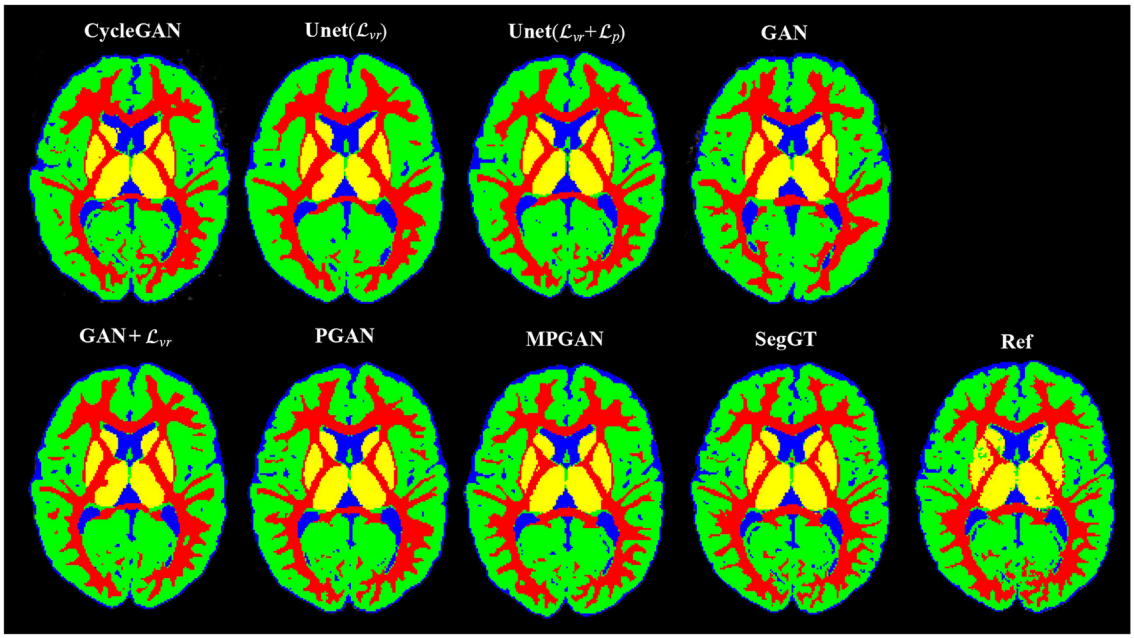

FIGURE 5 | Examples of the tissue segmentation results. The first four columns are the automatic segmentations of the predicted and the ground-truth images (SegGT). The last column is the reference manual segmentation (Ref).

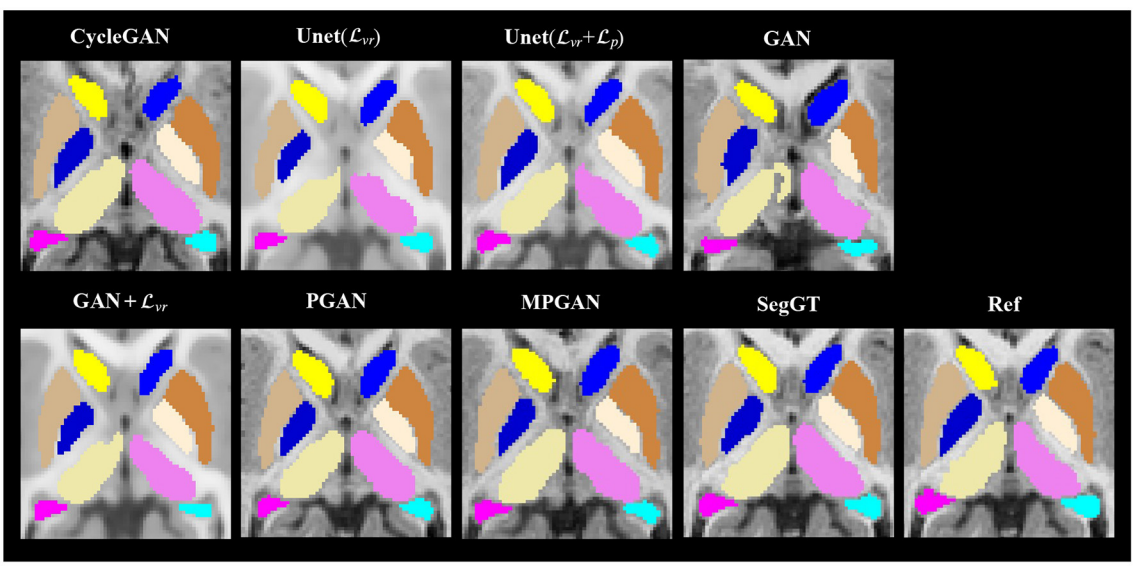

FIGURE 6 | Examples of the subcortical segmentation results. The first four columns are the automatic segmentations of the predicted and the ground-truth images (SegGT). The last column is the reference manual segmentation (Ref).

et al., 2007) is the mean of the closest distances from voxels on B_{S_1} to B_{S_2} and from voxels on B_{S_2} to B_{S_1}, respectively. It can be defined as

$$ASD = \frac{1}{|B_{S_1}| + |B_{S_2}|} \times \left(\sum_{x \in B_{S_1}} d(x, B_{S_2}) + \sum_{x \in B_{S_2}} d(x, B_{S_1}) \right) \quad (10)$$

The Dice coefficient evaluates the spatial overlap between two segmentation results, which is defined as

$$Dice = \frac{2|S_1 \cap S_2|}{|S_1| + |S_2|} \quad (11)$$

In order to obtain a overall evaluation criterion, we also combine the multiple metrics into a single fused score

TABLE 3 | Segmentation consistency across different approaches on subcortical segmentation task.

	***AVD* ↓**	***ASD* ↓**	***Dice* ↑**	***FusedScore* ↓**
$SegPredict_{GAN}$ vs. SegGT	$^{*\Delta}15.436 \pm 7.683$	$^{*\Delta}0.800 \pm 0.069$	$^{*\Delta}0.738 \pm 0.077$	$^{*\Delta}2.760 \pm 0.727$
$SegPredict_{Unet(\mathcal{L}_{vr})}$ vs. SegGT	$^{*\Delta}19.224 \pm 7.934$	$^{*\Delta}0.672 \pm 0.071$	$^{*\Delta}0.781 \pm 0.058$	$^{*\Delta}2.712 \pm 0.690$
$SegPredict_{GAN+\mathcal{L}_{vr}}$ vs. SegGT	$^{*\Delta}15.656 \pm 5.545$	$^{*\Delta}0.632 \pm 0.039$	$^{*\Delta}0.794 \pm 0.052$	$^{*\Delta}2.415 \pm 0.494$
$SegPredict_{CycleGAN}$ vs. SegGT	$^{*\Delta}12.909 \pm 4.947$	$^{*\Delta}0.678 \pm 0.094$	$^{*\Delta}0.779 \pm 0.057$	$^{*\Delta}2.345 \pm 0.534$
$SegPredict_{Unet(\mathcal{L}_{vr}+\mathcal{L}_p)}$ vs. SegGT	$^{*\Delta}7.566 \pm 2.107$	0.555 ± 0.021	0.820 ± 0.044	$^{*\Delta}1.761 \pm 0.261$
$SegPredict_{PGAN}$ vs. SegGT	$^{*}5.638 \pm 2.049$	0.555 ± 0.030	0.820 ± 0.047	$^{*}1.644 \pm 0.274$
$SegPredict_{MPGAN}$ vs. SegGT	$^{\Delta}\mathbf{5.153 \pm 1.767}$	0.556 ± 0.025	0.820 ± 0.043	$^{\Delta}\mathbf{1.618 \pm 0.235}$

*$SegPredict_{CycleGAN}$, $SegPredict_{Unet(\mathcal{L}_{vr})}$, $SegPredict_{Unet(\mathcal{L}_{vr}+\mathcal{L}_p)}$, $SegPredict_{GAN}$, $SegPredict_{GAN+\mathcal{L}_{vr}}$, $SegPredict_{PGAN}$ and $SegPredict_{MPGAN}$ denote automatic segmentations on the predicted images from CycleGAN, Unet($\mathcal{L}_{vr}$), Unet($\mathcal{L}_{vr}+\mathcal{L}_p$), GAN, GAN+$\mathcal{L}_{vr}$, PGAN, and MPGAN, respectively. SegGT is automatic segmentation on the ground-truth image. *Significantly different compared to MPGAN ($p < 0.05$). $^{\Delta}$Significantly different compared to PGAN ($p < 0.05$). ↓, lower is better. ↑, higher is better. The methods are sorted by the fused score. Methods with best performance are bolded for each metric (significantly better than other measurements, $p < 0.5$).*

TABLE 4 | Segmentation consistency across different approaches on tissue segmentation task.

	***AVD* ↓**	***ASD* ↓**	***Dice* ↑**	***FusedScore* ↓**
$SegPredict_{GAN}$ vs. SegGT	$^{*\Delta}4.556 \pm 2.043$	$^{*\Delta}0.850 \pm 0.148$	$^{*\Delta}0.709 \pm 0.129$	$^{*\Delta}2.247 \pm 0.451$
$SegPredict_{CycleGAN}$ vs. SegGT	$^{*\Delta}6.486 \pm 3.571$	$^{*\Delta}0.744 \pm 0.101$	$^{*\Delta}0.739 \pm 0.120$	$^{*\Delta}2.151 \pm 0.399$
$SegPredict_{Unet(\mathcal{L}_{vr})}$ vs. SegGT	$^{*\Delta}6.538 \pm 4.722$	$^{*\Delta}0.651 \pm 0.052$	$^{*\Delta}0.777 \pm 0.103$	$^{*\Delta}1.939 \pm 0.408$
$SegPredict_{GAN+\mathcal{L}_{vr}}$ vs. SegGT	$^{*\Delta}4.797 \pm 3.040$	$^{*\Delta}0.613 \pm 0.041$	$^{*\Delta}0.779 \pm 0.097$	$^{*\Delta}1.783 \pm 0.285$
$SegPredict_{Unet(\mathcal{L}_{vr}+\mathcal{L}_p)}$ vs. SegGT	$^{*\Delta}3.496 \pm 1.263$	$^{*\Delta}0.591 \pm 0.038$	$^{*\Delta}0.783 \pm 0.104$	$^{*\Delta}1.668 \pm 0.328$
$SegPredict_{PGAN}$ vs. SegGT	2.958 ± 1.566	$^{*}0.583 \pm 0.038$	$^{*}0.786 \pm 0.108$	$^{*}1.614 \pm 0.384$
$SegPredict_{MPGAN}$ vs. SegGT	2.933 ± 1.404	$^{\Delta}\mathbf{0.564 \pm 0.024}$	$^{\Delta}\mathbf{0.791 \pm 0.103}$	$^{\Delta}\mathbf{1.576 \pm 0.351}$

*$SegPredict_{CycleGAN}$, $SegPredict_{Unet(\mathcal{L}_{vr})}$, $SegPredict_{Unet(\mathcal{L}_{vr}+\mathcal{L}_p)}$, $SegPredict_{GAN}$, $SegPredict_{GAN+\mathcal{L}_{vr}}$, $SegPredict_{PGAN}$ and $SegPredict_{MPGAN}$ denote automatic segmentations on the predicted images from CycleGAN, Unet($\mathcal{L}_{vr}$), Unet($\mathcal{L}_{vr}+\mathcal{L}_p$), GAN, GAN+$\mathcal{L}_{vr}$, PGAN, and MPGAN, respectively. SegGT is automatic segmentation on the ground-truth image. *Significantly different compared to MPGAN ($p < 0.05$). $^{\Delta}$Significantly different compared to PGAN ($p < 0.05$). ↓, lower is better. ↑, higher is better. The methods are sorted by the fused score. Methods with best performance are bolded for each metric (significantly better than other measurements, $p < 0.5$).*

(FS). We follow (Van Ginneken et al., 2007) for the fused score, and thus use the *TanimotoError* as a measure of overlap instead of Dice coefficient when calculating FS. FS is formulated as

$$FS = \frac{1}{3}\left(\frac{AVD}{refAVD} + \frac{ASD}{refASD} + \frac{TanimotoError}{refTanimotoError}\right) \tag{12}$$

where *TanimotoError* is

$$TanimotoError = \frac{|S_1 \cup S_2| - |S_1 \cap S_2|}{|S_1 \cup S_2|} \tag{13}$$

As in Van Ginneken et al. (2007), *refAVD*, *refASD*, and *refTanimotoError* are set to 5.6%, 0.27 mm, and 15.8%, respectively, based on the manual segmentation variance among human experts.

3.5.1. Segmentation Consistency Analysis

In this section, we aim at assessing the quality of predicted images by evaluating how well the automatic segmentations of the predicted images match the ones of the ground-truth images. The intuition is that if two images are segmented by the same algorithm, the more similar the two images are, the more similar their segmentation results should be. The comparison results on subcortical segmentation task are presented in **Table 3**. We observe that, with respect to AVD, our PGAN and MPGAN significantly outperform all the other methods ($p < 0.05$) and MPGAN achieves the best performance. We can also see that, for ASD and Dice, there are no significant differences among Unet($\mathcal{L}_{vr} + \mathcal{L}_p$), PGAN, and MPGAN, but our PGAN and MPGAN show a superior performance ($p < 0.05$) compared to the remaining four methods. While the images generated by Unet($\mathcal{L}_{vr} + \mathcal{L}_p$) are visually of lower quality owing to blurred and unrealistic details (see **Figures 2**, **3**), in this segmentation analysis, Unet($\mathcal{L}_{vr} + \mathcal{L}_p$) performs at acceptable level for ASD and Dice coefficient. A possible explanation for this is that the segmentation algorithm we applied is robust to the image quality, to some extent, for this subcortical segmentation task. As for the fused score, our PGAN and MPGAN methods significantly outperform than other methods and MPGAN achieves the best score.

Table 4 lists the comparison results on the brain tissue segmentation task. The results confirmed the statistically significant better performance of our proposed methods (both PGAN and MPGAN) vs. the other methods, with respect to all metrics. The results clearly demonstrate the effectiveness of

TABLE 5 | Segmentation accuracy analysis on predicted images from different methods on subcortical segmentation task.

	***AVD* ↓**	***ASD* ↓**	***Dice* ↑**	***FusedScore* ↓**
$SegPredict_{Unet(\mathcal{L}_{vr})}$ vs. Ref	$^{*\Delta}24.304 \pm 9.173$	$^{*\Delta}0.832 \pm 0.144$	$^{*\Delta}0.731 \pm 0.085$	$^{*\Delta}3.349 \pm 0.861$
$SegPredict_{GAN}$ vs. Ref	$^{*\Delta}18.573 \pm 10.597$	$^{*\Delta}0.940 \pm 0.177$	$^{*\Delta}0.699 \pm 0.096$	$^{*\Delta}3.218 \pm 0.937$
$SegPredict_{GAN+\mathcal{L}_{vr}}$ vs. Ref	$^{*\Delta}20.715 \pm 7.776$	$^{*\Delta}0.822 \pm 0.122$	$^{*\Delta}0.737 \pm 0.071$	$^{*\Delta}3.110 \pm 0.670$
$SegPredict_{CycleGAN}$ vs. Ref	$^{*\Delta}18.023 \pm 6.998$	$^{*\Delta}0.749 \pm 0.128$	$^{*\Delta}0.757 \pm 0.060$	$^{*\Delta}2.807 \pm 0.617$
$SegPredict_{Unet(\mathcal{L}_{vr}+\mathcal{L}_{p})}$ vs. Ref	$^{*\Delta}13.122 \pm 5.901$	$^{*\Delta}0.748 \pm 0.115$	$^{*\Delta}0.762 \pm 0.060$	$^{*\Delta}2.502 \pm 0.508$
$SegPredict_{PGAN}$ vs. Ref	$^{*}10.390 \pm 6.217$	0.720 ± 0.110	0.771 ± 0.056	$^{*}2.279 \pm 0.505$
$SegPredict_{MPGAN}$ vs. Ref	$^{\Delta}\mathbf{9.254 \pm 5.691}$	0.721 ± 0.109	0.771 ± 0.051	$^{\Delta}\mathbf{2.215 \pm 0.460}$
SegGT vs. Ref	6.000 ± 4.396	0.666 ± 0.119	0.788 ± 0.055	1.902 ± 0.390

*$SegPredict_{CycleGAN}$, $SegPredict_{Unet(\mathcal{L}_{vr})}$, $SegPredict_{Unet(\mathcal{L}_{vr}+\mathcal{L}_{p})}$, $SegPredict_{GAN}$, $SegPredict_{GAN+\mathcal{L}_{vr}}$, $SegPredict_{PGAN}$ and $SegPredict_{MPGAN}$ denote automatic segmentations on the predicted images from CycleGAN, Unet($\mathcal{L}_{vr}$), Unet($\mathcal{L}_{vr}+\mathcal{L}_{p}$), GAN, GAN+$\mathcal{L}_{vr}$, PGAN, and MPGAN, respectively. SegGT is automatic segmentation on the ground-truth image and Ref denotes manual segmentation. *Significantly different compared to MPGAN ($p < 0.05$). $^{\Delta}$Significantly different compared to PGAN ($p < 0.05$). The methods are sorted by the fused core. Methods with best performance are bolded for each metric (significantly better than other measurements, $p < 0.5$).*

TABLE 6 | Segmentation accuracy analysis on predicted images from different methods on tissue segmentation task.

	***AVD* ↓**	***ASD* ↓**	***Dice* ↑**	***FusedScore* ↓**
$SegPredict_{GAN}$ vs. Ref	4.090 ± 2.507	$^{*\Delta}0.904 \pm 0.143$	$^{*\Delta}0.689 \pm 0.151$	$^{*\Delta}2.346 \pm 0.575$
$SegPredict_{CycleGAN}$ vs. Ref	5.493 ± 2.908	$^{*\Delta}0.747 \pm 0.093$	$^{*\Delta}0.724 \pm 0.143$	$^{*\Delta}2.130 \pm 0.458$
$SegPredict_{Unet(\mathcal{L}_{vr})}$ vs. Ref	$^{*\Delta}8.431 \pm 5.839$	$^{*\Delta}0.687 \pm 0.026$	$^{*\Delta}0.768 \pm 0.123$	$^{*\Delta}2.120 \pm 0.634$
$SegPredict_{GAN+\mathcal{L}_{vr}}$ vs. Ref	$^{*\Delta}5.302 \pm 2.105$	$^{*\Delta}0.636 \pm 0.024$	$^{*\Delta}0.776 \pm 0.118$	$^{*\Delta}1.849 \pm 0.384$
$SegPredict_{Unet(\mathcal{L}_{vr}+\mathcal{L}_{p})}$ vs. Ref	3.610 ± 2.109	$^{*}0.589 \pm 0.031$	$^{*}0.784 \pm 0.119$	$^{*}1.667 \pm 0.373$
$SegPredict_{PGAN}$ vs. Ref	3.817 ± 1.506	$^{*}0.583 \pm 0.031$	$^{*}0.788 \pm 0.115$	$^{*}1.662 \pm 0.352$
$SegPredict_{MPGAN}$ vs. Ref	3.948 ± 1.285	$^{\Delta}\mathbf{0.531 \pm 0.049}$	$^{\Delta}\mathbf{0.800 \pm 0.112}$	$^{\Delta}\mathbf{1.574 \pm 0.343}$
SegGT vs. Ref	4.158 ± 2.604	0.376 ± 0.075	0.871 ± 0.058	1.185 ± 0.270

*$SegPredict_{CycleGAN}$, $SegPredict_{Unet(\mathcal{L}_{vr})}$, $SegPredict_{Unet(\mathcal{L}_{vr}+\mathcal{L}_{p})}$, $SegPredict_{GAN}$, $SegPredict_{GAN+\mathcal{L}_{vr}}$, $SegPredict_{PGAN}$ and $SegPredict_{MPGAN}$ denote automatic segmentations on the predicted images from CycleGAN, Unet($\mathcal{L}_{vr}$), Unet($\mathcal{L}_{vr}+\mathcal{L}_{p}$), GAN, GAN+$\mathcal{L}_{vr}$, PGAN, and MPGAN, respectively. SegGT is automatic segmentation on the ground-truth image and Ref denotes manual segmentation. *Significantly different compared to MPGAN ($p < 0.05$). $^{\Delta}$Significantly different compared to PGAN ($p < 0.05$). The methods are sorted by the fused score. Methods with best performance are bolded for each metric (significantly better than other measurements, $p < 0.5$).*

our proposed methods. Furthermore, MPGAN offers superior performance to PGAN in this task, which indicates that MPGAN benefits from using complementary information from multiple contrast data when performing a full brain tissue segmentation.

3.5.2. Segmentation Accuracy Analysis

In this section, we computed the AVD, ASD, and Dice coefficient between the reference manual segmentation ("Ref") and the automatic segmentation of the ground-truth images or the predicted images. The values of these metrics here reflect the segmentation accuracy of the given image. Our goal is to evaluate the predicted images by comparing their segmentation accuracy with the one of the ground-truth image. Intuitively, the image which is more similar to the ground-truth image should have a closer segmentation accuracy to the ground-truth image. The segmentation accuracy comparison results on the subcortical segmentation task are shown in **Table 5**. We can observe that the images predicted by MPGAN achieve the closest AVD, Dice coefficient and fused score to the ground-truth images. **Table 6** illustrates the segmentation accuracy comparison results on the tissue segmentation task. Similar to the subcortical segmentation task, MPGAN outperforms the other methods across most of the metrics (except AVD).

3.6. Efficacy of Data Imputation

The issue of missing scans is a common, practical problem in longitudinal studies. Subjects with incomplete scans cannot be used as training samples for machine learning applications as well as also for statistical methods that need complete data. Thus, the training size is significantly reduced due to these missing scans. Intuitively, using a larger training set is expected to improve performance, because adding training samples can bring more information and increase the diversity of the dataset. In this section, we use our method to predict missing subject scans for the purpose of machine learning tasks. After completing the data, these subjects can then be added to the training set for methods necessitating complete longitudinal subject data. While increasing the size of training samples via such imputation can improve the model performance, the imputed data has to be of high quality and representing the longitudinal data distribution appropriately. Poorly imputed data is expected to reduce model performance. Here, we investigated whether adding our imputed

data to increase the size of the training set is beneficial to two practical tasks:

1. Classification of the severity group according to the social affect (SA) calibrated severity score (CSS) of the Autism Diagnostic Observation Schedule (ADOS, second edition) (Lord et al., 2012) at 24 months of age from prior longitudinal image data at 6 and 12 months
2. Regression of the gestational age at birth from later longitudinal image data at 6 and 12 months.

We compared the results of two settings: "non-imputed" and "imputed." For the "non-imputed" setting, the classifier/regressor was trained with only real image pairs, i.e., both 6-month and 12-month images are real. For the "imputed" setting, in addition to the real training pairs used in the first setting, the classifier/regressor was also trained on "mixed" image pairs, i.e., real 6-month and predicted 12-month images, or predicted 6-month and real 12-month images. No imputed data was employed in the testing set. Thus, any differences in performance would stem from the additional inclusion of the imputed/generated datasets in the training set.

In our experiments, the real image pairs were divided into 4 folds and a 4-fold cross-validation was employed to evaluate the performance. Each time one fold was used for testing, and the other three were used for training. For the "imputed" setting, an additional set of "mixed" image pairs were included in all training folds of the cross-validation scheme. Since MPGAN performed the best in previous experiments, here we only employed MPGAN for image imputation. The Extreme Gradient Boosting (Xgboost) algorithm (Chen and Guestrin, 2016) was applied in both classification and regression tasks, which was implemented using the scikit-learn Python libraries (Raschka, 2015). Instead of directly feeding the raw images into the Xgboost model, which would result in an extremely high dimensional feature space, we employed the features extracted by the model genesis encoding (the 3D deep learning model previously used for the perceptual loss in section 2.4.1) as our inputs.

3.6.1. ADOS-SA-CSS Group Classification at 24 Months With Imputed Data

Our goal in this experiment was to classify subject image data into one of three social affect severity groups (typical: score 1–2, low: score 3–4, moderate-to-high: score 5–10 Hus et al., 2014) at 24 months of age using 6 and 12 months MR image pairs. The ADOS-SA-CSS is a calibrated score that was developed to capture the severity of symptoms in social affect in children with ASD. We selected ADOS-SA-CSS as prediction measure instead of other calibrated ADOS scores (Restricted and Repetitive Behavior, RRB, and total severity score) as the ADOS-SA-CSS was observed to have a smoother distribution in our sample, as well as prior work indicate wide-scale associations with atypical social behavior in ASD (Sato and Uono, 2019).

We assessed whether data imputation improves the classification performance via F1 score, AUC score, and balanced accuracy metrics. Before data imputation, 77 subjects with complete scans at 6 and 12 month can be used for training.

TABLE 7 | Effects of imputed longitudinal data on the ADOS-SA-CSS group classification task.

Setting	F_1*Score* ↑	*AUC* ↑	*BalancedAccuracy* ↑
Non-imputed (77 subjects)	0.590	0.655	0.594
Imputed (77+103 subjects)	0.694	0.671	0.699

↑, *higher is better.*

After data imputation, an additional 103 subjects can be added to the training set. **Table 7** summarizes the cross-validated results in the "non-imputed" and "imputed" setting. We found that adding imputed data into training process can improve the model performance, such that the F1 score is increased from 0.590 to 0.694, the AUC score is increased from 0.655 to 0.671, and the balanced accuracy is increased from 0.594 to 0.699. We show the confusion matrices of the "non-imputed" and "imputed" setting in **Figure 7**. It is observed that the number of correctly classified samples is clearly improved across all groups, but especially in the moderate-to-high ADOS-SA-CSS group.

3.6.2. Regression of Gestational Age at Birth With Imputed Data

In this part, we employed the 6 and 12 months MR images to regress the gestational age at birth (39.03 ± 1.50 weeks). The mean absolute error (MAE) and relative error (RE) of the regressed GA were used as metrics. The 4-fold cross-validated results of "non-imputed" and "imputed" setting are shown in **Table 8**. Before data imputation, 134 subjects with complete scans at 6 and 12 month are used for training. After data imputation, an additional 76 subjects are added to the training set. Incorporating imputed data into the training process offers a slight improvement on the regression performance, though this improvement is of smaller magnitude than in the ADOS-SA-CSS classification.

4. DISCUSSION

In this paper, we present a novel adaptation of GAN to a new application: the longitudinal prediction of infant MR images. We validated and compared our technique with five alternative networks from multiple perspectives: qualitative and quantitative assessments of the image appearance, as well as quantitative assessment on two segmentation tasks. A consistent superior performance of our method has been shown in these evaluations, indicating its effectiveness. The images predicted by our method were then used for expanding the train set for ADOS-SA-CSS group classification and gestational age regression experiments. These experiments show that the imputed data brings a performance boost, highlighting the potential of our image prediction method when applied to a practical task.

Looking at **Figures 2**, **3**, one can see that the 3D-Unet trained with only the voxel-wise reconstruction loss ($\mathcal{L}_{vr}$) does a good job in keeping low-frequency information, and thus global structures are well-preserved. However, the images produced by

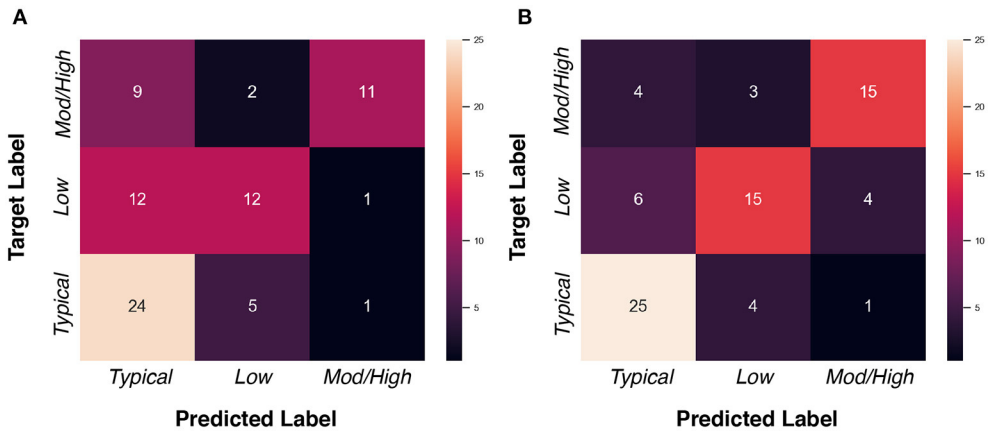

FIGURE 7 | Confusion matrices of ADOS-SA-CSS group classification task. **(A)** "Non-imputed" setting. **(B)** "Imputed" setting.

TABLE 8 | Effects of imputed longitudinal data on the gestational age (in weeks) regression task.

Setting	*RE* ↓	*MAE* ↓
Non-imputed (134 subjects)	2.559	0.971
Imputed (134 + 76 subjects)	2.327	0.886

↓, *lower is better.*

it appear blurry and significantly lack high-frequency details, which is also reflected quantitatively with it achieving the worst LPIPS scores in **Figure 4**. We found that integrating the perceptual loss $\mathcal{L}_p$ into the training can effectively alleviate this problem, as Unet($\mathcal{L}_{vr} + \mathcal{L}_p$) predicts sharper images, compared to Unet($\mathcal{L}_{vr}$). Nevertheless, the images generated by Unet($\mathcal{L}_{vr} + \mathcal{L}_p$) still have unrealistic appearance from visual perspective. Using an adversarial training scheme, our proposed PGAN employs a voxel-wise reconstruction loss, a perceptual loss, and an adversarial loss jointly to produce sharper and more realistic images. This results in a statistically significant improvement in the quantitative assessment (see **Figure 4**). This may be due to the following reason: with respect to the adversarial learning strategy, the discriminator is optimized to differentiate the real and fake images. In order to fool the discriminator, the generator has to push the output distribution closer to the distribution of real data. As a result, the outputs of the generator are visually realistic.

As T1w and T2w images encompass rich information that is different and complementary to each other, we further propose a multi-contrast version, called MPGAN, which produces even finer details as well as achieves a better quantitative score, compared to PGAN. In particular, we observe a loss in the cortical contrast information in T2w images in most of evaluated methods, while it appears well-preserved for MPGAN. In summary, the combination of the voxel-wise reconstruction loss, the perceptual loss, the adversarial loss and the use of multi-contrast information allows our MPGAN to produce realistic images with accurate details where both low and high frequency information is well-preserved.

Also, we report analyses of the segmentation consistency and accuracy for the different methods in both subcortical and tissue segmentation tasks. As shown in **Tables 3**, **4**, our MPGAN achieves the best segmentation consistency across most of the used metrics, where the differences are significant at the $p < 0.05$ level. The second-best performance is yielded by our single contrast PGAN method. As shown in **Tables 5**, **6**, our MPGAN results in the closest segmentation accuracy to the ground truth images across most of the considered metrics. Besides, we also found that, with respect to ASD and Dice coefficient, there is no significant difference between PGAN and MPGAN for both segmentation consistency and accuracy analysis on the subcortical segmentation task. However, MPGAN clearly performs better than PGAN for tissue segmentation task. This may be attributed to the reason that subcortical segmentation is relatively simple because of consistent shape of subcortical structures, compared to the folded, complex cortex assessed in the tissue segmentation task.

To investigate the applicability of our predicted/imputed images, we employed our predicted image data for data imputation in two practical tasks, i.e., one on ADOS-SA-CSS group classification and the other on regression of gestational age. The results show that the model performance can be boosted with the help of imputed data for the both tasks. We did not employ the imputed data for testing purposes, so all gains in classification/regression are due to the inclusion of the imputed data. This finding indicates that the predicted data was sufficiently close to the true data that it provided valuable information to the training process.

It is further noteworthy that any image data prediction is biased by the training data. Thus, we expect our method

to potentially perform poorly for brain images with atypical morphometry or neuropathology that are unknown to the trained model as such data was not included in the training. This indicates the necessity to develop additional safeguards to ensure that the input data and the trained model are appropriately matched. In the results presented here, we apply our methods in a fairly narrow subject population (typically developing children and children at familial risk for ASD) and all MRI data was inspected by a neuroradiologist for the presence of visible neuropathology.

While our method shows very promising results, it is not without limitations. For one, the current approach needs paired longitudinal data of the same subject for training. A further computation limitation is that we are directly feeding 3D data into our networks, which requires large amounts of memory and thus a high performance GPU server.

5. CONCLUSION

This paper introduces a novel multi-contrast perceptual generative network (MPGAN) for longitudinal prediction of infant MRI data. To the best of our knowledge, this is the first time that deep generative methods are applied for longitudinal prediction of structural MRI in the first year of life. Our approach improves the realism, sharpness, and accuracy of predicted images by merging the adversarial learning scheme with the voxel-wise reconstruction loss and the perceptual loss, as well as taking the multi-contrast information into account. In our qualitative and quantitative assessments, our method yielded a better performance than the alternative approaches studied in this work.

Longitudinal data is crucial to capture appropriate developmental trajectories in studies of the first year of life. Missing data is a major issue and our proposed method achieves highly promising results to impute such missing data for training data augmentation in classification or regression tasks. The improvement in performance when classifying subjects into categories of severity of social affect symptoms from image data only is quite impressive.

Our future work will focus on extending the paired approach at consistent time points (here 6 and 12 months of age) to a time regression based approach to overcome our current limitation of discrete time points and model imputation along the full first year of life. Furthermore, additional experiments to quantify the value of adding real data (i.e., acquiring additional subjects) vs. adding imputed data (i.e., imputing incomplete data as performed here) will need to be performed.

DATA AVAILABILITY STATEMENT

Code: The source code of our MPGAN method is available as open source at https://github.com/liying-peng/MPGAN. Data: All raw MRI datasets and associated demographic information employed in this study is available at NIH/NDA: https://nda.nih.gov/edit_collection.html?id=19.

ETHICS STATEMENT

This study was reviewed and approved by the Institutional Review Boards (IRB) of all data collection sites (University of North 1654 Carolina, University of Washington, Children's Hospital of 1655 Pennsylvania, Washington University). Written informed consent was provided by the parent/guardian of all enrolled subjects. The data used in this work is collected from the Infant Brain Imaging Study (IBIS) database (https://www.ibis-network.org).

AUTHOR CONTRIBUTIONS

LP, LL, YL, Y-wC, GG, and MAS contributed to methodological development. LP, MAS, HH, CB, RG, MDS, and JG contributed to the study design. ACE, SD, AME, RM, KB, RS, HH, and JP contributed to the image data collection. LP, ZM, RV, SK, and MAS contributed to the experiment data collection. LP and MAS contributed to the statistical analysis. LP and MAS wrote the first draft of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

This study was supported by grants from the Major Scientific Project of Zhejiang Lab (No. 2018DG0ZX01), the National Institutes of Health (R01-HD055741, T32-HD040127, U54-HD079124, U54-HD086984, R01-EB021391, and P50-HD103573), Autism Speaks, and the Simons Foundation (140209). MDS was supported by NIH career development award K12-HD001441, as is JG K01-MH122779. The sponsors had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

ACKNOWLEDGMENTS

We are sincerely grateful to all the families and children who have participated in the Infant Brain Imaging Study (IBIS). The Infant Brain Imaging Study (IBIS) Network is an NIH funded Autism Centers of Excellence project and consists of a consortium of 9 universities in the U.S. and Canada. Members and components of the IBIS Network include: JP (IBIS Network PI), Clinical Sites: University of North Carolina: HH, C. Chappell, MDS, M. Swanson; University of Washington: SD, AME, D. Shaw, T. St. John; Washington University: KB, J. Constantino; Children's Hospital of Philadelphia: RS, J. Pandey. Behavior Core: University of Washington: AME; University of Alberta: L. Zwaigenbaum; University of Minnesota: J. Elison, J. Wolff. Imaging Core: University of North Carolina: MAS; New York University: GG; Washington University in St. Louis: RM, J. Pruett. Data Coordinating Center: Montreal Neurological Institute: ACE, D. L. Collins, V. Fonov, L. MacIntyre; S. Das. Statistical Analysis Core: K. Truong. Environmental risk core: John Hopkins University: H. Volk. Genetics Core: John Hopkins University: D. Fallin; University of North Carolina: MDS. We

would also like to thank Y. Gong, M. W. Ren, H. Sui, R. H. Ma, L. Liu, M. Bagonis, Y. Panikratova, R. Rozovskaya, M. Egorova, M. Foster, K. A. Ali, A. Rumple, G. R. Wu, J. Z. Chen, A. Q. Chen, H. Shah, Y. Zhang, D. Liang, and H. Zheng for their participation in the human perceptual assessment study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2021.653213/full#supplementary-material

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

Ben-Cohen, A., Klang, E., Raskin, S. P., Amitai, M. M., and Greenspan, H. (2017). "Virtual pet images from ct data using deep convolutional networks: initial results," in *International Workshop on Simulation and Synthesis in Medical Imaging* (Quebec, QC: Springer), 49–57. doi: 10.1007/978-3-319-68127-6_6

Bi, L., Kim, J., Kumar, A., Feng, D., and Fulham, M. (2017). "Synthesis of positron emission tomography (PET) images via multi-channel generative adversarial networks (GANs)," in *Molecular Imaging, Reconstruction and Analysis of Moving Body Organs, and Stroke Imaging and Treatment* (Quebec, QC: Springer), 43–51. doi: 10.1007/978-3-319-67564-0_5

Bowles, C., Gunn, R., Hammers, A., and Rueckert, D. (2018). "Modelling the progression of Alzheimer's disease in MRI using generative adversarial networks," in *Medical Imaging 2018: Image Processing* (Houston, TX: International Society for Optics and Photonics).

Chen, T., and Guestrin, C. (2016). "Xgboost: a scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (San Francisco, CA), 785–794. doi: 10.1145/2939672.2939785

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

Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., and Ronneberger, O. (2016). "3D U-Net: learning dense volumetric segmentation from sparse annotation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Athens: Springer), 424–432.

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. Imaging* 38, 2375–2388. doi: 10.1109/TMI.2019.2901750

Emami, H., Aliabadi, M. M., Dong, M., and Chinnam, R. B. (2021). SPA-GAN: spatial attention GAN for image-to-image translation. In: *IEEE Transactions on Multimedia*. Vol. 23, 391–401. doi: 10.1109/TMM.2020.2975961

Fishbaugh, J., Prastawa, M., Gerig, G., and Durrleman, S. (2013). "Geodesic shape regression in the framework of currents," in *International Conference on Information Processing in Medical Imaging* (Asilomar, CA: Springer), 718–729. doi: 10.1007/978-3-642-38868-2_60

Fishbaugh, J., Prastawa, M., Gerig, G., and Durrleman, S. (2014). "Geodesic regression of image and shape data for improved modeling of 4D trajectories," in *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)* (Bejing: IEEE), 385–388. doi: 10.1109/ISBI.2014.6867889

Fletcher, P. T. (2013). Geodesic regression and the theory of least squares on riemannian manifolds. *Int. J. Comput. Vision* 105, 171–185. doi: 10.1007/s11263-012-0591-y

Gilmore, J. H., Knickmeyer, R. C., and Gao, W. (2018). Imaging structural and functional brain development in early childhood. *Nat. Rev. Neurosci.* 19:123. doi: 10.1038/nrn.2018.1

Gilmore, J. H., Shi, F., Woolson, S. L., Knickmeyer, R. C., Short, S. J., Lin, W., et al. (2012). Longitudinal development of cortical and subcortical gray matter from birth to 2 years. *Cereb. Cortex* 22, 2478–2485. doi: 10.1093/cercor/bhr327

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 27*, eds Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger (Montreal, QC: Curran Associates, Inc.), 2672–2680.

Hasegawa, C., Takahashi, T., Yoshimura, Y., Nobukawa, S., Ikeda, T., Saito, D. N., et al. (2018). Developmental trajectory of infant brain signal variability: a longitudinal pilot study. *Front. Neurosci.* 12:566. doi: 10.3389/fnins.2018.00566

Hazlett, H. C., Gu, H., Munsell, B. C., Kim, S. H., Styner, M., Wolff, J. J., et al. (2017). Early brain development in infants at high risk for autism spectrum disorder. *Nature* 542, 348–351. doi: 10.1038/nature21369

Huang, X., Liu, M.-Y., Belongie, S., and Kautz, J. (2018). "Multimodal unsupervised image-to-image translation," in *Proceedings of the European Conference on Computer Vision (ECCV)* (Munich), 172–189. doi: 10.1007/978-3-030-01219-9_11

Hus, V., Gotham, K., and Lord, C. (2014). Standardizing ados domain scores: separating severity of social affect and restricted and repetitive behaviors. *J. Autism Dev. Disord.* 44, 2400–2412. doi: 10.1007/s10803-012-1719-1

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* (Honolulu, HI), 1125–1134. doi: 10.1109/CVPR.2017.632

Jin, C.-B., Kim, H., Liu, M., Jung, W., Joo, S., Park, E., et al. (2019). Deep CT to MR synthesis using paired and unpaired data. *Sensors* 19:2361. doi: 10.3390/s19102361

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

Karras, T., Aila, T., Laine, S., and Lehtinen, J. (2017). Progressive growing of gans for improved quality, stability, and variation. *arXiv [preprint]. arXiv:1710.10196*.

Kingma, D. P., and Ba, J. (2014). Adam: a method for stochastic optimization. *arXiv [preprint]. arXiv:1412.6980*.

Laird, N. M. (1988). Missing data in longitudinal studies. *Stat. Med.* 7, 305–315.

Liu, M.-Y., Breuel, T., and Kautz, J. (2017). "Unsupervised image-to-image translation networks," in *Proceedings of the 31st International Conference on Neural Information Processing Systems* (Long Beach, CA: Curran Associates Inc.), 700–708.

Lord, C., Rutter, M., DiLavore, P. C., Risi, S., and Gotham, K. (2012). *Autism Diagnostic Observation Schedule, Second Edition (ADOS-2) Manual (Part I): Modules 1–4*. Torrance, CA: Western Psychological Service.

Matta, T. H., Flournoy, J. C., and Byrne, M. L. (2018). Making an unknown unknown a known unknown: missing data in longitudinal neuroimaging studies. *Dev. Cogn. Neurosci.* 33, 83–98. doi: 10.1016/j.dcn.2017.10.001

Meng, Y., Li, G., Rekik, I., Zhang, H., Gao, Y., Lin, W., et al. (2017). Can we predict subject-specific dynamic cortical thickness maps during infancy from birth? *Hum. Brain Mapp.* 38, 2865–2874. doi: 10.1002/hbm.23555

Nash, J. F., et al. (1950). Equilibrium points in n-person games. *Proc. Natl. Acad. Sci. U.S.A.* 36, 48–49.

Nie, D., Trullo, R., Lian, J., Petitjean, C., Ruan, S., Wang, Q., et al. (2017). "Medical image synthesis with context-aware generative adversarial networks," in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Quebec, QC: Springer), 417–425. doi: 10.1007/978-3-319-66179-7_48

Niethammer, M., Huang, Y., and Vialard, F.-X. (2011). "Geodesic regression for image time-series," in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Toronto, ON: Springer), 655–662. doi: 10.1007/978-3-642-23629-7_80

Pan, Y., Liu, M., Lian, C., Xia, Y., and Shen, D. (2019). "Disease-image specific generative adversarial network for brain disease diagnosis with incomplete multi-modal neuroimages," in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Shenzhen: Springer), 137–145. doi: 10.1007/978-3-030-32248-9_16

Pan, Y., Liu, M., Lian, C., Zhou, T., Xia, Y., and Shen, D. (2018). "Synthesizing missing pet from mri with cycle-consistent generative adversarial networks for Alzheimer's disease diagnosis," in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Granada: Springer), 455–463.

Pathak, D., Krahenbuhl, 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* (Las Vegas, NV), 2536–2544. doi: 10.1109/CVPR.2016.278

Qu, L., Wang, S., Yap, P.-T., and Shen, D. (2019). "Wavelet-based semi-supervised adversarial learning for synthesizing realistic 7T from 3T MRI," in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Shenzhen: Springer), 786–794. doi: 10.1007/978-3-030-32251-9_86

Raschka, S. (2015). *Python Machine Learning*, 1st Edition. Birmingham, UK: Packt Publishing Ltd.

Ravi, D., Alexander, D. C., Oxtoby, N. P., and Alzheimer's Disease Neuroimaging Initiative (2019). "Degenerative adversarial neuroimage nets: generating images that mimic disease progression," in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Shenzhen: Springer), 164–172. doi: 10.1007/978-3-030-32248-9_19

Rekik, I., Li, G., Lin, W., and Shen, D. (2016). Predicting infant cortical surface development using a 4D varifold-based learning framework and local topography-based shape morphing. *Med. Image Anal.* 28, 1–12. doi: 10.1016/j.media.2015.10.007

Rekik, I., Li, G., Wu, G., Lin, W., and Shen, D. (2015). "Prediction of infant mri appearance and anatomical structure evolution using sparse patch-based metamorphosis learning framework," in *International Workshop on Patch-based Techniques in Medical Imaging* (Munich: Springer), 197–204. doi: 10.1007/978-3-319-28194-0_24

Sato, W., and Uono, S. (2019). The atypical social brain network in autism: advances in structural and functional MRI studies. *Curr. Opin. Neurol.* 32, 617–621. doi: 10.1097/WCO.0000000000000713

Singh, N., Hinkle, J., Joshi, S., and Fletcher, P. T. (2013a). "A hierarchical geodesic model for diffeomorphic longitudinal shape analysis," in *International Conference on Information Processing in Medical Imaging* (Asilomar, CA: Springer), 560–571.

Singh, N., Hinkle, J., Joshi, S., and Fletcher, P. T. (2013b). "A vector momenta formulation of diffeomorphisms for improved geodesic regression and atlas construction," in *2013 IEEE 10th International Symposium on Biomedical Imaging* (San Francisco, CA: IEEE), 1219–1222.

Thung, K.-H., Wee, C.-Y., Yap, P.-T., and Shen, D. (2016). Identification of progressive mild cognitive impairment patients using incomplete longitudinal MRI scans. *Brain Struct. Funct.* 221, 3979–3995. doi: 10.1007/s00429-015-1140-6

Ulyanov, D., Vedaldi, A., and Lempitsky, V. (2016). Instance normalization: the missing ingredient for fast stylization. *arXiv [preprint]. arXiv:1607.08022*.

Van Ginneken, B., Heimann, T., and Styner, M. (2007). "3D segmentation in the clinic: a grand challenge," in *MICCAI Workshop on 3D Segmentation in the Clinic: A Grand Challenge* (Brisbane), 7–15.

Wang, J., Vachet, C., Rumple, A., Gouttard, S., Ouziel, C., Perrot, E., et al. (2014). Multi-atlas segmentation of subcortical brain structures via the autoseg software pipeline. *Front. Neuroinform.* 8:7. doi: 10.3389/fninf.2014.00007

Wolterink, J. M., Dinkla, A. M., Savenije, M. H., Seevinck, P. R., van den Berg, C. A., and Išgum, I. (2017). "Deep MR to CT synthesis using unpaired data," in *International Workshop on Simulation and Synthesis in Medical Imaging* (Quebec, QC: Springer), 14–23. doi: 10.1007/978-3-319-68127-6_2

Xia, T., Chartsias, A., Wang, C., and Tsaftaris, S. A. (2019). Learning to synthesise the ageing brain without longitudinal data. *Med. Image Anal.* 73:102169 doi: 10.1016/j.media.2021.102169

Xiong, F., Wang, Q., and Gao, Q. (2019). Consistent embedded GAN for image-to-image translation. *IEEE Access* 7, 126651–126661. doi: 10.1109/ACCESS.2019.2939654

Yang, Q., Li, N., Zhao, Z., Fan, X., Eric, I., Chang, C., et al. (2020). MRI cross-modality image-to-image translation. *Sci. Rep.* 10, 1–18.

Yi, Z., Zhang, H., Tan, P., and Gong, M. (2017). "DualGAN: unsupervised dual learning for image-to-image translation," in *Proceedings of the IEEE International Conference on Computer Vision* (Venice), 2849–2857. doi: 10.1109/ICCV.2017.310

Zhang, H., Goodfellow, I., Metaxas, D., and Odena, A. (2019). "Self-attention generative adversarial networks," in *International Conference on Machine Learning* (Long Beach, CA), 7354–7363.

Zhang, R., Isola, P., Efros, A. A., Shechtman, E., and Wang, O. (2018). "The unreasonable effectiveness of deep features as a perceptual metric," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Salt Lake City, UT), 586–595. doi: 10.1109/CVPR.2018.00068

Zhao, F., Wu, Z., Wang, L., Lin, W., Xia, S., Shen, D., et al. (2019). "Harmonization of infant cortical thickness using surface-to-surface cycle-consistent adversarial networks," in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Shenzhen: Springer), 475–483. doi: 10.1007/978-3-030-32251-9_52

Zhao, H., Gallo, O., Frosio, I., and Kautz, J. (2015). Loss functions for neural networks for image processing. *arXiv [preprint]. arXiv:1511.08861*.

Zhou, Z., Sodha, V., Siddiquee, M. M. R., Feng, R., Tajbakhsh, N., Gotway, M. B., et al. (2019). "Models genesis: generic autodidactic models for 3D medical image analysis," in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Shenzhen: Springer), 384–393. doi: 10.1007/978-3-030-32251-9_42

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* (Venice), 2223–2232. doi: 10.1109/ICCV.2017.244

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.