# Hierarchical Perception Adversarial Learning Framework for Compressed Sensing MRI

Zhifan Gao, *Member, IEEE*, Yifeng Guo, Jiajing Zhang, Tieyong Zeng, and Guang Yang, *Senior Member, IEEE*

***Abstract*— The long acquisition time has limited the accessibility of magnetic resonance imaging (MRI) because it leads to patient discomfort and motion artifacts. Although several MRI techniques have been proposed to reduce the acquisition time, compressed sensing in magnetic resonance imaging (CS-MRI) enables fast acquisition without compromising SNR and resolution. However, existing CS-MRI methods suffer from the challenge of aliasing artifacts. This challenge results in the noise-like textures and missing the fine details, thus leading to unsatisfactory reconstruction performance. To tackle this challenge, we propose a hierarchical perception adversarial learning framework (HP-ALF). HP-ALF can perceive the image information in the hierarchical mechanism: image-level perception and patch-level perception. The former can reduce the visual perception difference in the entire image, and thus achieve aliasing artifact removal. The latter can reduce this difference in the regions of the image, and thus recover fine details. Specifically, HP-ALF achieves the hierarchical mechanism by utilizing multilevel perspective discrimination. This discrimination can provide the information from two perspectives (overall and regional) for adversarial learning. It also utilizes a global and local coherent discriminator to provide structure information to the generator during training. In addition, HP-ALF contains a context-aware learning block to effectively exploit the slice information between individual images for better reconstruction performance. The experiments validated on three datasets demonstrate the effectiveness of HP-ALF and its superiority to the comparative methods.**

***Index Terms*— MRI reconstruction, compressed sensing, magnetic resonance imaging, generative adversarial networks.**

## I. Introduction

THE long acquisition time in magnetic resonance imaging (MRI) limits the accessibility of this modality [1]. It can lead to patient discomfort and motion artifacts [2], [3]. Several existing MRI techniques have been proposed to reduce the acquisition time, such as parallel imaging (PI), simultaneous multislice (SMS) and compressed sensing in magnetic resonance imaging (CS-MRI). CS-MRI enables a significant reduction in the MRI acquisition time without compromising SNR and resolution with respect to other MRI techniques [4]. It can avoid the deterioration of image quality because it performs nonlinear optimization on highly-undersampled raw data. Thus, CS-MRI allows clinicians to complete multidimensional scans in a clinically feasible scan time. For example, CS-MRI can reduce the scan time by 29.3% in a daily clinical routine study for the brain [5] and reduce the time for volumetric cardiac-resolved flow imaging sequences (4D flow) from an hour to 5−10 minutes [6].

However, CS-MRI still has unsatisfactory reconstruction performance because of aliasing artifacts [7]. This challenge leads to difficulty in applying CS-MRI in the examination of some clinical indications, such as epilepsy and pediatric anesthesia [8], [9]. The aliasing artifacts are derived from the high undersampling in CS-MRI [4]. They cause noise-like textures and the missing fine details to corrupt the reconstructed image [10], as shown in Figure 1(a). First, noise-like texture refers to an irregular pattern that blurs the image globally [4]. It obscures and weakens the appearance of the structure and the edge, and thus interferes with the extraction of this feature information. This interference leads to distortion of the reconstructed image. Second, the fine detail missing refers to the blurry textures in the different parts of the image [11], [12]. It conceals and weakens the structure boundary and

Manuscript received 31 August 2022; revised 21 October 2022 and 16 January 2023; accepted 26 January 2023. Date of publication 30 January 2023; date of current version 1 June 2023. This work was supported in part by the National Key Research and Development Program of China under Grant 2022YFE0209800; in part by the National Natural Science Foundation of China under Grant U1908211 and Grant 62101606; in part by the Shenzhen Science and Technology Program under Grant GXWD20201231165807008, 20200825113400001; in part by ERC IMI under Grant 101005122; in part by H2020 under Grant 952172; in part by MRC under Grant MC/PC/21013; in part by the Royal Society under Grant IEC/NSFC/211235; in part by NVIDIA Academic Hardware Grant Program; in part by NIHR Imperial Biomedical Research Centre under Grant RDA01; in part by the Imperial–Nanyang Technological University Collaboration Fund; in part by UKRI MRC with MSIT and NRF Fund; in part by the UKRI Future Leaders Fellowship under Grant MR/V023799/1, Grant NSFC/RGC N_CUHK 415/19, Grant ITF MHP/038/20, Grant CRF 8730063, Grant RGC 14300219, Grant 14302920, and Grant 14301121; in part by the National Natural Science Foundation of China under Grant 62276282; and in part by the Guangdong Basic and Applied Basic Research Foundation under Grant 2022A1515011384. *(Corresponding author: Guang Yang.)*

Zhifan Gao, Yifeng Guo, and Jiajing Zhang are with the School of Biomedical Engineering, Sun Yat-sen University, Shenzhen 510275, China (e-mail: gaozhifan@mail.sysu.edu.cn; guoyf25@mail2.sysu.edu.cn; zhangjj83@mail2.sysu.edu.cn).

Tieyong Zeng is with the Department of Mathematics, The Chinese University of Hong Kong, Sha Tin, Hong Kong, China (e-mail: zeng@math.cuhk.edu.hk).

Guang Yang is with the Cardiovascular Research Centre, Royal Brompton Hospital, SW3 6NP London, U.K., and also with the National Heart and Lung Institute, Imperial College London, SW3 6LY London, U.K. (e-mail: g.yang@imperial.ac.uk).

Digital Object Identifier 10.1109/TMI.2023.3240862

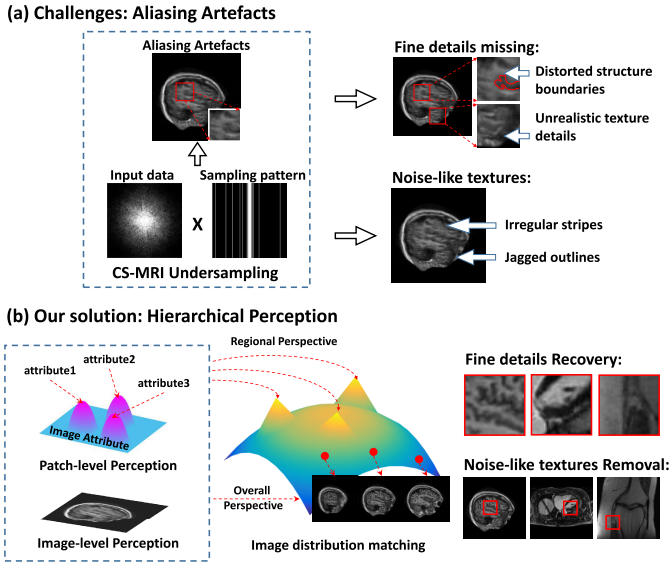


Fig. 1. The contribution of our hierarchical perception adversarial learning framework (HP-ALF). HP-ALF utilizes hierarchical perception to address the challenge of aliasing artifacts. This challenge is derived from the high undersampling in CS-MRI, which applies the sampling patterns to the input data. It results in noise-like texture and missing fine details in the image domain. Hierarchical perception includes image-level perception and patch-level perception. They can reduce the visual perception difference from overall perspective and regional perspective, and thus remove noise-like texture and restore fine details.

the texture details and results in the unrealistic appearance of small structures. The unrealistic image appearance reduces the perceptual quality of the reconstructed image.

Existing CS-MRI methods have difficulty addressing the challenge of aliasing artifacts [11], [13], [14]. Although they perform well under the interference of noise-like texture [15], [16], aliasing artefacts still come from the missing fine details [12], [17]. This is because these methods perceive the visual difference between the reconstructed image and original image in the overall perspective (for removing noise-like texture), rather than in the regional perspective (for restoring the fine details). First, conventional CS-MRI methods tend to focus on low-frequency image information owing to hand-crafted feature extraction rather than high-frequency information corresponding to fine details [14], [15]. This corrupts the reconstruction of the fine details. Second, the existing deep-learning-based CS-MRI methods also face this difficulty, although they enable both high- and low-frequency information extraction [16], [18], [19]. They usually rely on the computation of the pixel-wise distance in the spatial or frequency domain, and thus have to smooth the fine details in the reconstructed image [11], [12], [20]. This may lead to difficulty in reducing the visual perception difference between fine details. Third, adversarial learning methods can reconstruct the image with fine details by bringing in the distribution distance [11], [12], [21], [22], [23]. However, these methods recover fine details from the overall perspective. This is because the information of these details is related to the attributes in different image regions. These attributes represent the parts of the image, including the different levels of aliasing artifacts. The level of aliasing artifacts changes in different image regions and influences the attributes in the reconstruction of fine details. Therefore, the visual perception difference needs to be measured not only from the overall perspective but also from the regional perspective.

In this paper, we propose the hierarchical perception adversarial learning framework (HP-ALF) to tackle the challenge of aliasing artifacts, as shown in Figure 1(b). It builds the perception of image information by the hierarchical mechanism: image-level and patch-level. First, image-level perception refers to the extraction of the image information from the overall perspective. It perceives and calculates the global aliasing artifacts, and thus reduces the visual perception difference in the entire image. Second, patch-level perception refers to the extraction of detailed information from the regional perspective. It perceives and calculates the local aliasing artifacts and thus reduces the visual perception difference in the different regions of the image. Thus, HP-ALF not only builds the image-level perception to reduce the visual perception difference from the overall perspective, but also enables the patch-level perception from the regional perspective. It can remove noise-like textures and restore the fine details simultaneously, thus reducing image distortion and improving perceptual quality. Specifically, HP-ALF is implemented by the multilevel perspective discrimination. It matches the distributions of the reconstructed image and the original image by comparing their quality difference from both the overall and regional perspectives. Therefore, it facilitates the reconstruction of high-quality images.

Our contributions can be summarized as follows:

1. We develop a CS-MRI framework to reconstruct high-quality MRI images. It enables the hierarchical perception by multilevel perspective discrimination to reduce the visual perception difference from the overall perspective and the regional perspectives. This framework can achieve both the noise-like texture removal and fine detail restoration to address the aliasing artifacts owing to the high undersampling.

2. We design a minimization problem for CS-MRI to evaluate the perceptual quality from the overall and regional perspectives. To solve this problem, we extract the global and local structure information as well as the slice information in the image sequence. First, we build a global and local coherent discriminator to provide the detailed per-pixel decision to the generator while maintaining the global coherence of the reconstructed images. Then, the context-aware learning block in the generator exploits the slice information from the MRI sequence.

3. We validate our framework on three datasets for different anatomical structures (brain, heart, knee). The experimental results demonstrate the effectiveness of our framework, as well as its superiority to comparative CS-MRI methods.

This work advances our preliminary work in MICCAI 2020 [24]. First, it extends the minimization problem that is applied to the noise-like texture removal and restores the fine details simultaneously. Second, it designs a novel objective function that transforms the single-level perspective discrimination into the multilevel perspective discrimination for perceptual quality improvement. Third, it extends the discriminator to a U-net-based architecture that can reconstruct globally and locally coherent images for fine detail preservation. Finally, the

experiments are extended to three datasets imaging different organs and three more validations (evaluating the effectiveness of the proposed objective function, the U-net-based discriminator, and the context-aware learning block).

## II. Related Works

The current CS-MRI methods can be broadly classified into conventional CS-MRI methods, traditional deep learning-based CS-MRI methods, and adversarial learning based CS-MRI methods. However, these methods ignore the restoration of fine details.

First, conventional CS-MRI methods have the challenge of extracting high-frequency information for feature representation of fine details. These methods include sparsity-based and dictionary learning-based CS-MRI. First, sparsity-based CS-MRI reconstruction methods have been developed to leverage the sparsity of signals by using predefined and fixed sparse transformations [4], [25] and exploiting spatiotemporal correlations [15]. These methods usually rely on the experience-based determination of what kind of low-frequency image information is helpful. Second, compared to sparsity-based methods, the dictionary learning (DL) can generate data-specific dictionaries and improve image quality [26], [27]. However, the dictionary learning method has difficulty reconstructing high-frequency features because these dictionaries are still designed based on low-frequency image features.

In addition, existing CS-MRI methods still difficulty reconstructing realistic fine details, although deep learning shows high potential in many medical image applications [28], [29], [30], [31], [32]. Compared with conventional CS-MRI methods, existing deep learning-based methods can extract the high- and low-frequency features. These methods can be divided into two classes [13]. First, end-to-end optimization models the inverse acquisition to achieve the fast MRI reconstruction. For example, Feng et al. [16] introduced an end-to-end task transformer network (T2Net) utilizing an $\ell_1$ loss function. Second, unrolled optimization incorporates prior domain knowledge about the expected properties of MR images. For example, Qin et al. [18] proposed a convolutional recurrent neural network (CRNN) method based on unrolled optimization with the pixel-wise distance in the spatial domain. Guo et al. [33] proposed an unrolled model based on the novel convolutional recurrent neural network (OUCR) with $\ell_1$ loss. Hu et al. [34] proposed a self-supervised unrolled model (SSL-MRI) based on the parallel network training framework with a pixel-wise loss. However, these methods fail to reconstruct the realistic fine details. This is because it is difficult to improve the perceptual quality of fine details by reducing the visual perception difference. The visual perception difference is associated with the distance between the reconstructed image distribution and the original image distribution [20]. These deep-learning-based methods have difficulty calculating this distribution distance [20].

Furthermore, adversarial learning can introduce the distribution distance to further reduce the visual perception difference [11], [12], [21], [22], [35]. However, it is still unsatisfactory for recovering fine details. Adversarial learning methods enable the computation of the distribution distance [20]. Thus, these methods can capture and reduce visual perception differences to improve the perceptual quality of the reconstructed image. The existing methods can be divided into two classes. First, the loss-variant model utilizes auxiliary penalties to improve the perceptual quality of the reconstructed image. For instance, Yang et al. [11] proposed a combination of pixel-wise, perceptual, and GAN losses to achieve fast CS-MRI utilizing the conditional generative adversarial network-based model (DAGAN). Mardani et al. [12] proposed a mixture of pixel-wise and least-squares GAN (GANCS) losses in which the least-squares GAN learns the texture details and the pixel-wise loss suppresses high-frequency noise. Second, the architecture-variant model relies on the MRI data characteristics to improve the reconstruction performance. For example, Shaul et al. [17] proposed a two-stage GAN framework (Sub-GAN) including a cascade of a $k$-space and an image-space U-Net with a mixture loss. Korkmaz et al. [36] proposed a novel unsupervised MRI reconstruction based on an unconditional deep adversarial network (SLATER) utilizing a mixture loss. Wei et al. [37] introduced a two-stage generative adversarial network utilising cross-domain learning with $\ell_1$ and $\ell_2$ pixel-wise loss. However, these methods have difficulty improving the perceptual quality of fine details. This is because they reduce the visual perception difference only from the overall perspective and thus lead to the unrealistic reconstruction of the fine details.

## III. Method

### A. Problem Statement

Let $\mathbf{x} \in \mathbb{C}^N$ be the slice of 2D images to be reconstructed, where each slice consists of $\sqrt{N} \times \sqrt{N}$ pixels for one image. The problem is to reconstruct $\mathbf{x}$ from $\mathbf{y} \in \mathbb{C}^M$ ($M \ll N$), undersampled measurements in $k$-space, such that $\mathbf{y} = \mathbf{F}_u\mathbf{x}+\epsilon$. $\mathbf{F}_u$ is the undersampling Fourier encoding operator and $\epsilon$ is complex Gaussian noise [7]. However, such measurements are underdetermined even in the absence of noise because of the violation of the Nyquist-Shannon sampling theorem [4]. Therefore, the corresponding linear inversion for CS-MRI $\mathbf{x}_u = \mathbf{F}_u^H\mathbf{y}$ is usually ill-posed, where $H$ denotes the Hermitian transpose operation. $\mathbf{x}_u$ suffers from the challenge of aliasing artifacts. The artifacts result in noise-like textures and missing fine details in the image domain. The existing CS-MRI methods address the challenge by formulating a minimization problem [4], [11]:

$$\min_{\mathbf{x}} \lambda \|\mathbf{y} - \mathbf{F}_u\mathbf{x}\|_2^2 + \mathcal{R}(\mathbf{x}), \tag{1}$$

where $\|\mathbf{y} - \mathbf{F}_u\mathbf{x}\|_2^2$ is the data fidelity term [7] and $\mathcal{R}(\mathbf{x})$ is the regularization term. $\lambda$ is the regularization parameter. However, these methods have difficulty removing noise-like textures and restoring fine details simultaneously. This is because Equation (1) tends to realize noise-like texture removal but neglects fine detail recovery. The detailed image information is related to the attributes in different image regions. These methods can only perceive the image from the overall perspective. Therefore, CS-MRI reconstruction requires both

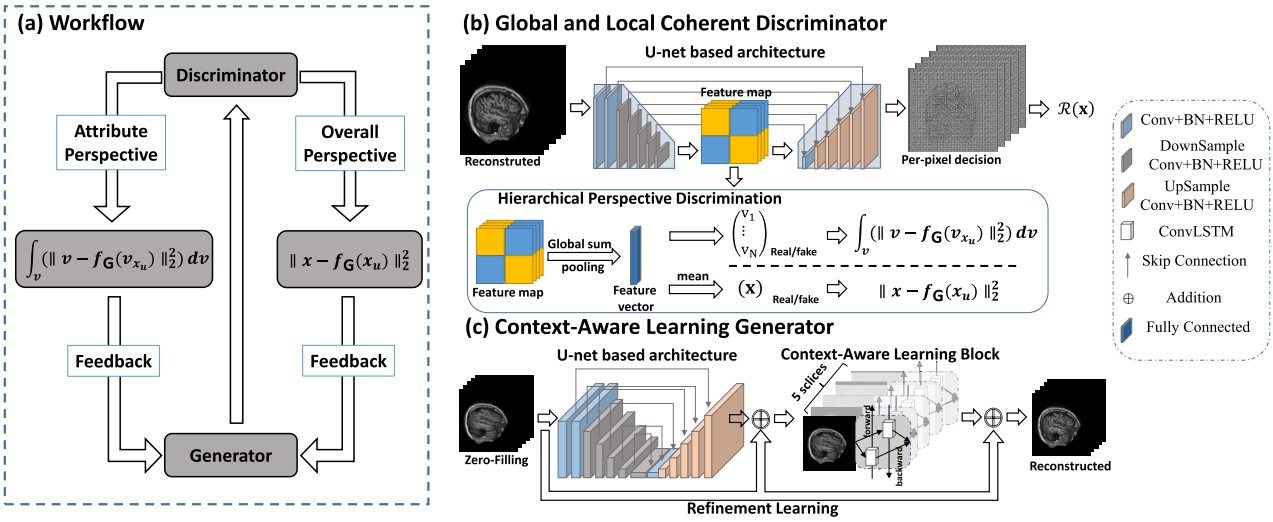


Fig. 2. Overview of our HP-ALF. (a) HP-ALF utilizes multilevel perspective discrimination to achieve the hierarchical perception by providing information from the overall perspective and the regional perspective. (b) Our global and local coherent discriminator utilizes a U-net-based architecture to provide fine details information during training. It is achieved by using the decoder of U-net to provide a detailed per-pixel decision to the generator. (c) Context-aware learning generator includes a U-net-based architecture and a context-aware learning block. U-net can utilize 2D spatial information while the context-aware learning block can exploit 3D spatial feature from the sequential MRI data.

noise-like texture removal and fine detail recovery from different perspectives. HP-ALF builds an additional regularization term (i.e., a hierarchical perception term) in Equation (1). This term can reduce the visual perception difference from the overall perspective and the regional perspective. The CS-MRI reconstruction problem can be reformulated as a different minimization problem:

$$\begin{aligned}\min_{\mathbf{x},\mathbf{v}} &\int_{\mathbf{v}} \left(\left\|\mathbf{v}_{\mathbf{x}} - f_{\mathrm{G}}\left(\mathbf{v}_{\mathbf{x}_u}\right)\right\|_2^2\right) d\mathbf{v} + \left\|\mathbf{x} - f_{\mathrm{G}}\left(\mathbf{x}_u\right)\right\|_2^2 \\ &+ \lambda \left\|\mathbf{y} - \mathbf{F}_u\mathbf{x}\right\|_2^2 + \mathcal{R}(\mathbf{x}), \\ &\mathcal{R}(\mathbf{x}) = \sum_{i,j} \left\|\left[f_{\mathrm{D}}\left(\mathbf{x}\right)\right]_{i,j} - f_{\mathrm{D}}\left(f_{\mathrm{G}}\left(\mathbf{x}_u\right)\right)\right]_{i,j}\right\| \end{aligned} \tag{2}$$

where $f_{\mathrm{G}} : \mathbb{C}^N \mapsto \mathbb{C}^N$ is the model that reconstructs images from $\mathbf{x}_u$ to address the challenge of aliasing artifacts. $f_{\mathrm{D}}$ represents the model that outputs local (per-pixel) changes between the reconstructed image and the original image. $\mathbf{x}_u$ represents the $n$ slices of sequential MRI images. Further explanations of the variable $\mathbf{v}$, hierarchical perception term and regularization term are as follows.

*1) The Variable* $\mathbf{v}$*:* The variable $\mathbf{v}$ is the mathematical symbol of the attribute. The attribute represents the parts of the image with different levels of aliasing artifacts. Thus, $\mathbf{v}_{\mathbf{x}}$ represents the parts of the reconstructed image $\mathbf{x}$ with different levels of aliasing artifacts. $\mathbf{v}_{\mathbf{x}_u}$ represents those of the zero-filled reconstruction $\mathbf{x}_u$. Then, the term $\int_{\mathbf{v}} \left(\left\|\mathbf{v}_{\mathbf{x}} - f_{\mathrm{G}}\left(\mathbf{v}_{\mathbf{x}_u}\right)\right\|_2^2\right) d\mathbf{v}$ in Equation (2) aims to compute the difference in the attributes between the reconstructed image and the original image. This facilitates Equation (2) in focusing on the removal of local aliasing artifacts in different regions. Specifically, Equation (2) utilizes the $\ell_2$ distance to calculate the difference of the variable $\mathbf{v}$. Then, it sums up all the difference results to obtain the integral result of $\int_{\mathbf{v}} \left(\left\|\mathbf{v}_{\mathbf{x}} - f_{\mathrm{G}}\left(\mathbf{v}_{\mathbf{x}_u}\right)\right\|_2^2\right) d\mathbf{v}$.

*2) Hierarchical Perception Term:* This term is the additional regularization term in Equation (1), including two parts. First, $\int_{\mathbf{v}} \left(\left\|\mathbf{v}_{\mathbf{x}} - f_{\mathrm{G}}\left(\mathbf{v}_{\mathbf{x}_u}\right)\right\|_2^2\right) d\mathbf{v}$ represents the difference in the local aliasing artifact between the reconstructed image and the original image. Therefore, it can calculate and reduce the visual perception difference in the different regions of the image and thus recover fine details. Second, $\left\|\mathbf{x} - f_{\mathrm{G}}\left(\mathbf{x}_u\right)\right\|_2^2$ represents the difference in the global aliasing artifact between the reconstructed image and the original image. Therefore, it can calculate and reduce the visual perception difference in the entire image, and thus achieve the aliasing artifact removal.

*3) Regularization Term:* The term $\mathcal{R}(\mathbf{x})$ in Equation (2) represents the local difference of the images at pixel $(i, j)$ calculated by the model $f_{\mathrm{D}}$. It provides the detailed per-pixel decision during the optimization process while maintaining the global coherence of the reconstructed images.

Our HP-ALF achieves Equation (2) by image-level perception and the patch-level perception. Image-level perception and patch-level perception are represented by the first term and the second term in Equation (2), respectively. They can reduce the visual perception difference from the overall perspective and the regional perspective. Specifically, HP-ALF constructs the encoder of the global and local coherent discriminator to achieve image-level perception and patch-level perception. The decoder of this discriminator can preserve the fine details. Then, a context-aware learning block in the generator exploits the slice information. In addition, the loss function in HP-ALF can be optimized for Equation (2). Figure 2 shows the details of the HP-ALF.

## B. Multilevel Perspective Discrimination

We propose multilevel perspective discrimination in HP-ALF to achieve the patch-level perception and image-level perception in Equation (2), as shown in Figure 2(a). It is built by the objective function of the generative adversarial network (GAN) including image-level perspective discrimination and patch-level discrimination. First, the patch-level

perspective discrimination refers to $\int_{\mathbf{v}} \left( \left\| \mathbf{v}_{\mathbf{x}} - f_{\mathrm{G}} \left( \mathbf{v}_{\mathbf{x}_u} \right) \right\|_2^2 \right) d\mathbf{v}$ in Equation (2). It measures the difference in $\mathbf{v}$ between the reconstructed image and the original image. Therefore, it can provide information from the regional perspective for distribution matching in adversarial learning. Second, image-level perspective discrimination refers to $\|\mathbf{x} - f_{\mathrm{G}}(\mathbf{x}_u)\|_2^2$ in Equation (2). It measures the difference between the reconstructed image and the original image. Therefore, it can provide information from the overall perspective for distribution matching in adversarial learning. The encoder of the discriminator in HP-ALF is constructed based on $\|\mathbf{x} - f_{\mathrm{G}}(\mathbf{x}_u)\|_2^2$ and $\int_{\mathbf{v}} \left( \left\| \mathbf{v}_{\mathbf{x}} - f_{\mathrm{G}} \left( \mathbf{v}_{\mathbf{x}_u} \right) \right\|_2^2 \right) d\mathbf{v}$. The optimization process of the encoder is presented as a two-player min-max value function. This value function includes the discriminator loss and the generator loss for the model training.

Specifically, the output of the encoder in the discriminator is applied as the input of this discrimination. The output of the encoder is divided into two parts. The first part is a scalar as that in the traditional adversarial learning method [21]. It is the mean value of the feature map obtained from the last layer of the encoder. This scalar represents the difference between the entire reconstructed image and the entire original image. Therefore, it can be used to achieve image-level perspective discrimination. The second part is the feature map obtained from the last layer in the encoder. It can be flattened as a discrete distribution $p_{\text{perspective}}$. This distribution is constructed based on the variable $\mathbf{v}$. Each element in this distribution corresponds to an image attribute. Therefore, this distribution can be applied for patch-level perspective discrimination. For the input sample $\boldsymbol{x}$, the corresponding discriminator output can be represented as $D(\boldsymbol{x}) = \left\{ p_{\text{perspective}}(\boldsymbol{x}, \mathbf{v}); \mathbf{v} \in \Omega \right\}$, where $\Omega$ is the set of outcomes of $p_{\text{perspective}}$. Each outcome $\mathbf{v}$ corresponds to the attribute in different regions. To measure the distance between distributions, HP-ALF constructs two normal distributions with a positive skew and a negative skew as $\mathcal{R}_1$ (real) and $\mathcal{R}_0$ (fake), respectively, similar to the virtual ground-truth scalars 0 and 1 in the standard GAN. $\mathcal{R}_1$ and $\mathcal{R}_0$ are also defined on $\Omega$. Accordingly, the JS divergence in the standard GAN is replaced with the Kullback-Leibler (KL) divergence. The min-max game between $G$ and $D$ thus becomes:

$$\begin{aligned} &\max_G \min_D V(G, D) \\ &\quad = \mathbb{E}_{\boldsymbol{x} \sim p_{\text{data}}} \left[ \mathcal{D}_{\mathrm{KL}} \left( \mathcal{R}_1(\mathbf{v}) \| D(\boldsymbol{x}) \right) + \log(D(\boldsymbol{x})) \right] \\ &\qquad + \mathbb{E}_{\boldsymbol{x} \sim p_g} \left[ \mathcal{D}_{\mathrm{KL}} \left( \mathcal{R}_0(\mathbf{v}) \| D(\boldsymbol{x}) \right) + \log(1 - D(\boldsymbol{x})) \right]. \end{aligned} \tag{3}$$

where $D$ in Equation (3) corresponds to the encoder of the discriminator. It aims to maximize the value function in Equation (3). Thus, the loss for $D$ can be formulated as:

$$\begin{aligned} \mathcal{L}_{D_{enc}^U} = &-\mathbb{E}_{\boldsymbol{x} \sim p_{\text{data}}} \left[ \mathcal{D}_{\mathrm{KL}} (\mathcal{R}_1(\mathbf{v}) \| D_{enc}^U(\boldsymbol{x})) + \log(D_{enc}^U(\boldsymbol{x})) \right] \\ &- \mathbb{E}_{\boldsymbol{x} \sim p_g} \left[ \mathcal{D}_{\mathrm{KL}} (\mathcal{R}_0(\mathbf{v}) \| D_{enc}^U(f_{\mathrm{G}}(\mathbf{x}))) \right. \\ &\left. + \log(1 - D_{enc}^U(f_{\mathrm{G}}(\mathbf{x}))) \right], \end{aligned} \tag{4}$$

where $D_{enc}^U$ is the encoder of the discriminator (i.e., $D$ in Equation (3)). $f_{\mathrm{G}}$ represents the generator $G$. Correspondingly, $G$ aims to minimize the value function in Equation (3). Thus, the adversarial loss for $G$ becomes:

$$\begin{aligned} \mathcal{L}_{\text{adv1}} = &-\mathbb{E}_{\boldsymbol{x} \sim p_g} [\mathcal{D}_{\mathrm{KL}} (\mathcal{R}_0(\mathbf{v}) \| D_{enc}^U(f_{\mathrm{G}}(\mathbf{x}))) \\ &+ \log(1 - D_{enc}^U(f_{\mathrm{G}}(\mathbf{x})))], \end{aligned} \tag{5}$$

The optimization of Equation (3) requires that HP-ALF reaches the Nash equilibrium. This further leads to the optimality of the generator and the discriminator. Specifically, Theorem 1 states that $D$ can reach its optimality for any given generator $G$. Then, in Theorem 2, $G$ can also reach its optimality when $D$ satisfies this optimality condition. The proofs for Theorems 1 and 2 are presented in Appendix I.

*Theorem 1:* When $G$ is fixed, for any outcome $\mathbf{v}$ and input sample $x$, the optimal discriminator $D$ satisfies

$$\begin{aligned} D_G^{\star}(\boldsymbol{x}, v) = &\frac{p_{\text{data}}(\boldsymbol{x})}{p_{\text{data}}(\boldsymbol{x}) + p_g(\boldsymbol{x})} \\ &+ \frac{\mathcal{R}_1(v) p_{\text{data}}(\boldsymbol{x}) + \mathcal{R}_0(v) p_g(\boldsymbol{x})}{p_{\text{data}}(\boldsymbol{x}) + p_g(\boldsymbol{x})} \end{aligned}$$

*Theorem 2:* When $D = D_G^{\star}$, and there exists an outcome $v \in \Omega$ such that $\mathcal{R}_1(v) \neq \mathcal{R}_0(v)$, the maximum of $V\left(G, D_G^{\star}\right)$ is achieved if and only if $p_g = p_{\text{data}}$

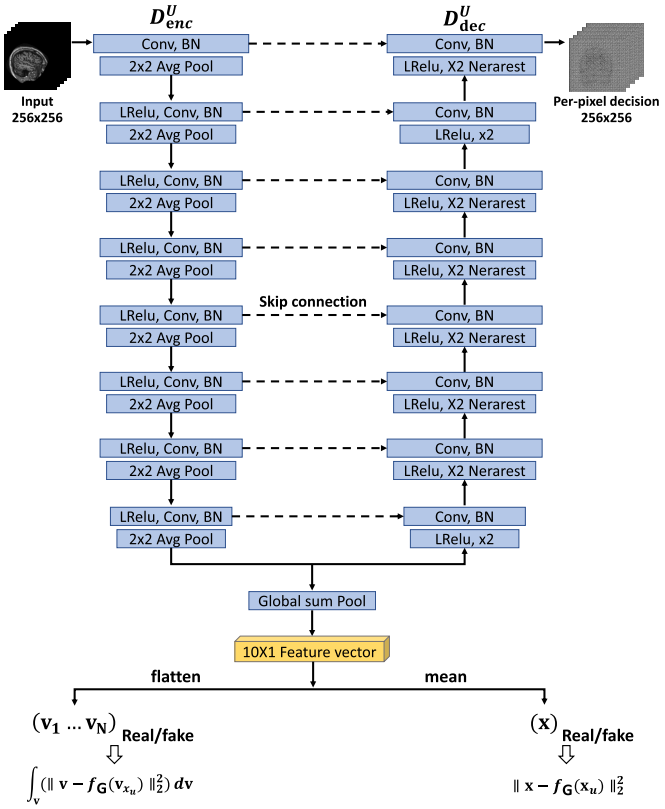


Fig. 3. The network details of the global and local coherent discriminator. The encoder in our discriminator represents the features and downsamples it in each level. The decoder in our discriminator represents the features and then upsamples it in each level. This discriminator also performs the multi-level discrimination for the reconstructed images.

## C. Global and Local Coherent Discriminator

HP-ALF utilizes the decoder of the global and local coherent discriminator (shown in Figure 3) to achieve the regularization term $\sum_{i,j} \left\| [f_{\mathrm{D}}(\mathbf{x})]_{i,j} - [f_{\mathrm{D}}(f_{\mathrm{G}}(\mathbf{x}_u))]_{i,j} \right\|$ in

Equation (2). This decoder can improve the ability to reconstruct globally and locally coherent images in existing adversarial learning. This is because, in the existing adversarial learning, these discriminators often focus either on the global structure or local details and thus provide insufficient information for the generator [38]. In contrast, HP-ALF builds the global and local coherent discriminator to provide the detailed per-pixel decision to the generator while maintaining the global coherence of reconstructed images. The decoder of this discriminator in HP-ALF is constructed based on the term $\sum_{i,j} \left\| [f_{\mathrm{D}}(\mathbf{x})]_{i,j} - [f_{\mathrm{D}}(f_{\mathrm{G}}(\mathbf{x}_u))]_{i,j} \right\|$. It outputs the classification on every pixel $(i, j)$ and then calculates the classification difference over all pixels between $x_t$ and $f_{\mathrm{G}}(x_u)$. The optimization process of the decoder can be formulated in a loss function. Then, this loss function can construct an adversarial loss for the generator to receive the global and local feedback from the decoder.

Specifically, the discriminator $D^U(x)$ has a U-net-based architecture including the encoder $D^U_{enc}$ and the decoder $D^U_{dec}$. The output of encoder $D^U_{enc}$ includes two parts. The first part is the feature map at the last layer in the encoder. It is applied for the multilevel perspective discrimination. This enables the encoder $D^U_{enc}$ to act as a discriminator to classify the image as real or fake, as mentioned in Section III-B. The second part includes the feature maps from the multiple levels in the encoder. The input of the decoder $D^U_{dec}$ includes the two parts of the output of the encoder. The first part is fed into the decoder for subsequent upsampling. The second part of the encoder output is fed to the corresponding levels in the decoder by skip connection. It can perform the classification on a per-pixel basis, segmenting image $x$ into real and fake regions. Therefore, the discriminator loss can be formulated by taking the decisions from both $D^U_{enc}$ and $D^U_{dec}$: $\mathcal{L}_{D^U} = \mathcal{L}_{D^U_{enc}} + \mathcal{L}_{D^U_{dec}}$, where $D^U_{enc}$ outputs the discrimination distribution score from the multilevel perspective discrimination and $D^U_{dec}$ outputs the per-pixel decision. The loss for the decoder $\mathcal{L}_{D^U_{dec}}$ is formulated as the mean decision over all pixels:

$$\begin{aligned}\mathcal{L}_{D^U_{dec}} = &-\mathbb{E}_{\boldsymbol{x}\sim p_{\mathrm{data}}}\Big[\sum_{i,j}\log\Big[D^U_{dec}(x)\Big]_{i,j}\Big] \\ &-\mathbb{E}_{\boldsymbol{x}\sim p_{\mathrm{g}}}\Big[\sum_{i,j}\log\Big(1-\Big[D^U_{dec}(G(x))\Big]_{i,j}\Big)\Big], \quad (6)\end{aligned}$$

where $\left[D^U_{dec}(x)\right]_{i,j}$ and $\left[D^U_{dec}(G(z))\right]_{i,j}$ are the local (per-pixel) decisions of the images at pixel $(i, j)$. HP-ALF encourages the discriminator to provide the detailed per-pixel decision to the generator. Therefore, the discriminator can help the generator maintain the global coherence of the reconstructed images by providing global image feedback. Correspondingly, the adversarial loss in Equation (6) becomes:

$$\mathcal{L}_{\mathrm{adv2}} = -\mathbb{E}_{\boldsymbol{x}\sim p_g}\Big[\sum_{i,j}\log\Big[D^U_{dec}\left(f_{\mathrm{G}}(\mathbf{x})\right)\Big]_{i,j}\Big]. \quad (7)$$

### D. Context-Aware Learning Generator

The context-aware learning generator $f_{\mathrm{G}}$ in Equation (2) aims to reconstruct the artifacts-free MR images from $\mathbf{x}_u$. As shown in Figure 2(c), the context-aware learning generator includes a U-net-based architecture $f_{\mathrm{Unet}}$ and a context-aware learning block $f_{\mathrm{CAL}}$. The generator is defined as follows:

$$\hat{\mathbf{x}}_u = f_{\mathrm{Unet}}(\mathbf{x}_u), \mathbf{x}_{rec} = f_{\mathrm{CAL}}\left(\hat{\mathbf{x}}_u\right), \quad (8)$$

where $\mathbf{x}_u$ is the input for the generator. $\hat{\mathbf{x}}_u$ and $\mathbf{x}_{rec}$ are the outputs of $f_{\mathrm{Unet}}$ and $f_{\mathrm{CAL}}$, respectively. The U-net based architecture fully utilizes the 2D spatial information in each slice, but neglects the correlation between adjacent 2D slices. The insufficient prior information leads to the inaccurate reconstruction of the fine details [24]. Therefore, the context-aware learning block exploits the 3D spatial feature from the input sequence of the MRI data. Specifically, this block utilizes the enhancement ConvLSTM (i.e., Bi-ConvLSTM) to achieve the exploration of 3D semantic knowledge. The LSTM unit contains a memory cell $\mathcal{C}_t$, an input gate $i_t$, a forget gate $f_t$, an output gate $o_t$, and an output state $\mathcal{H}_t$. However, ConvLSTM replaces LSTM by fully connected transformations with spatial local convolutions. ConvLSTM can be formulated as follows:

$$\begin{aligned} i_t &= \sigma\left(\mathbf{W}_{xi} * \mathcal{X}_t + \mathbf{W}_{hi} * \mathcal{H}_{t-1} + \mathbf{W}_{ci} * \mathcal{C}_{t-1} + b_i\right) \\ f_t &= \sigma\left(\mathbf{W}_{xf} * \mathcal{X}_t + \mathbf{W}_{hf} * \mathcal{H}_{t-1} + \mathbf{W}_{cf} * \mathcal{C}_{t-1} + b_f\right) \\ \mathcal{C}_t &= f_t \circ \mathcal{C}_{t-1} + i_t \tanh\left(\mathbf{W}_{xc} * \mathcal{X}_t + \mathbf{W}_{hc} * \mathcal{H}_{t-1} + b_c\right) \\ o_t &= \sigma\left(\mathbf{W}_{xo} * \mathcal{X}_t + \mathbf{W}_{ho} * \mathcal{H}_{t-1} + \mathbf{W}_{co} \circ \mathcal{C}_t + b_c\right) \\ \mathcal{H}_t &= o_t \circ \tanh\left(\mathcal{C}_t\right), \end{aligned} \quad (9)$$

where $\sigma$ and $\circ$ denote the convolution and Hadamard functions, respectively. $\mathcal{X}_t$ is the input tensor. Bi-ConvLSTM uses two ConvLSTMs to process the input data into two directions of the forward and backward paths and then makes a decision for the current input by dealing with the data dependencies in both directions. The output of Bi-ConvLSTM can be calculated as:

$$\mathbf{Y}_t = \tanh(\mathbf{W}_y^{\overrightarrow{\mathcal{H}}} * \overrightarrow{\mathcal{H}}_t + \mathbf{W}_y^{\overleftarrow{\mathcal{H}}} \overleftarrow{\mathcal{H}}_t + b), \quad (10)$$

where $\overrightarrow{\mathcal{H}}_t$ and $\overleftarrow{\mathcal{H}}_t$ denote the hidden state tensors for the forward state and the backward state, respectively, $\mathbf{W}_y^{\overrightarrow{\mathcal{H}}}$ and $\mathbf{W}_y^{\overleftarrow{\mathcal{H}}}$ denote the weight parameters for $\overrightarrow{\mathcal{H}}_t$ and $\overleftarrow{\mathcal{H}}_t$, respectively. $b$ is the bias term, and $\mathbf{Y}_t$ indicates the final output considering bidirectional information. Through the Bi-ConvLSTM module, HP-ALF can learn the fine details of MRI data slices.

### E. Loss Function and Implementation Details

The loss function for Equation (2) consists of a content loss and an adversarial loss. The content loss function is basically made up of a frequency-domain MSE loss and a perceptual VGG loss. The whole loss function can be formulated as

$$\mathcal{L}_{\mathrm{TOTAL}} = \alpha\mathcal{L}_{\mathrm{fMSE}} + \beta\mathcal{L}_{\mathrm{VGG}} + \mathcal{L}_{\mathrm{adv}}, \quad (11)$$

where $\alpha$ and $\beta$ are hyperparameters. $\mathcal{L}_{\mathrm{adv}}$ represents the adversarial loss. $\mathcal{L}_{\mathrm{fMSE}}$ is the frequency-domain MSE loss and $\mathcal{L}_{\mathrm{VGG}}$ is the perceptual VGG loss.

The adversarial loss in Equation (11) includes two parts. First, $\mathcal{L}_{\mathrm{adv1}}$ is the adversarial loss based on the optimization process in Equation (3). It can optimise

$\int_{\mathbf{v}} \left( \left\| \mathbf{v}_{\mathbf{x}} - f_{\mathrm{G}} \left( \mathbf{v}_{\mathbf{x}_u} \right) \right\|_2^2 \right) d\mathbf{v}$ and $\| \mathbf{x} - f_{\mathrm{G}} (\mathbf{x}_u) \|_2^2$ in Equation (2). Second, $\mathcal{L}_{\mathrm{adv2}}$ is the adversarial loss based on the loss function in Equation (6). It can optimize the regularization term $\mathcal{R}(\mathbf{x})$ in Equation (2). The whole adversarial loss function can be formulated as:

$$\mathcal{L}_{\mathrm{adv}} = -\mathbb{E}_{\boldsymbol{x} \sim p_g} \left[ \mathcal{D}_{\mathrm{KL}} \left( \mathcal{R}_0 \| D_{enc}^U (f_{\mathrm{G}} (\mathbf{x})) \right) + \log(1 - D_{enc}^U (f_{\mathrm{G}} (\mathbf{x}))) - \sum_{i,j} \log \left[ D_{dec}^U (f_{\mathrm{G}} (\mathbf{x})) \right]_{i,j} \right]. \tag{12}$$

The content loss can improve the perceptual quality of reconstruction. It includes a VGG loss and a frequency-domain MSE loss as constraints formulated as:

$$\mathcal{L}_{\mathrm{fMSE}} = \frac{1}{2} \left\| \mathrm{y}_t - \hat{\mathrm{y}}_u \right\|_2^2, \mathcal{L}_{\mathrm{VGG}} = \frac{1}{2} \| f_{\mathrm{VGG}} (\mathrm{x}_t) - f_{\mathrm{VGG}} \left( \hat{x}_u \right) \|_2^2. \tag{13}$$

$\mathcal{L}_{\mathrm{fMSE}}$ represents the difference between the reconstructed image and the original image in the frequency domain. Therefore, it can be applied to achieve the data fidelity term $\lambda \| \mathbf{y} - \mathbf{F}_u \mathbf{x} \|_2^2$ in Equation (2). $\mathcal{L}_{\mathrm{VGG}}$ is an additional regularisation term $\mathcal{R}(\mathbf{x})$ to constrain the solution space, where $f_{\mathrm{VGG}}$ denotes VGG feature maps of VGG16.

HP-ALF consists of the context-aware learning generator and the global and local coherent discriminator. The generator $G$ includes a U-net-based architecture and a context-aware learning block. The U-Net-based architecture consists of eight convolutional layers (encoder layers) and eight corresponding deconvolutional layers (decoder layers). The numbers of the filters are 64, 128, 256, 512, 512, 512, and 512 in the encoder layers and 1024, 1024, 1024, 1024, 512, 256, and 128 in the decoder layers. Each layer in $G$ uses a kernel with size $k = 3$ with the batch normalization and leaky ReLU layers behind it. The subsequent module of the U-Net-based architecture is the context-aware learning block. It has a Bi-ConvLSTM block, where the kernel size is $k = 3$ and the feature map channel inside is 32. $G$ also applies refinement learning to connect layers between the input and the output of the context-aware learning block. The architecture of $D$ is similar to the U-net-based architecture of $G$. In addition, $D$ cascades three dense convolutional layers after the encoder layer, and the sigmoid activation function outputs the classification results.

HP-ALF uses 5 consecutive 2D slices from 3D data as the input sequence. It also normalizes the intensities of all 2D slices into the range between $-1$ and 1 [11]. These adjacent slices are fed into the network in chip orders. Then, the output of the U-net is reshaped into the input of the context-aware learning block. The discriminator uses the original or generated image as input with a size of $256 \times 256$. In the discriminator, the encoder represents and downsamples the input image to a feature map (channels=10, columns=4, rows=4). This feature map is then transformed to a $10 \times 1$ feature vector by global sum pooling. The decoder applies the feature maps as input and upsamples it in each level until it reaches the original image size of $256 \times 256$.

## IV. Experiment and Results

### A. Data Collection

The experiments are carried out on three MRI datasets. (1) Brain MRI dataset: This is the MICCAI 2013 grand challenge dataset.[1] It contains 150 3D patient data with 42750 slices. Each patient data includes about 285 slices with $256 \times 256$ pixels. (2) Cardiac MRI dataset: This is the 2018 Atrial Segmentation Challenge dataset.[2] It contains 100 3D LGE MRI patient data with 5920 slices. Each patient data includes about 60 slices with $256 \times 256$ pixels. A whole-body MRI scanner is used for this dataset. The image acquisition resolution is $0.625\mathrm{mm}^2$. (3) Knee MRI dataset: This is the FastMRI dataset.[3] It contains 96 3D patient data with 3270 slices. Each patient data includes about 35 slices with $256 \times 256$ pixels. Each dataset is divided into training data (70%), validation data (20%) and test data (10%). In all 3D data, we exclude the slices at the edge, where the number of void pixels is greater than 90%. For the three datasets, the DICOM data are collected, rather than the raw $k$-space data. Not using the raw $k$-space data results in the inequality of image quality between the ground truth image and the fully sampled raw data. This is because the raw $k$-space data directly correspond to the originally measured raw data [39], [40]. However, the use of DICOM data considers the reproducibility and generality of the reconstruction method. This is because clinical centers usually save the image data [40].

To reduce the extra computational burden, the strategy in [11] is applied to handle the complex-valued data. Specifically, the real-valued information can be embedded into the complex space using an operator $\mathrm{Re}^* : \mathbb{R}^N \mapsto \mathbb{C}^N$ such that $\mathrm{Re}^*(\mathrm{x}) = \mathrm{x} + 0i$, and therefore the MRI forward operator can be expressed as $\mathrm{F} : \mathbb{R}^N \overset{\mathrm{Re}^*}{\mapsto} \mathbb{C}^N \mapsto \mathcal{F}\mathbb{C}^N \mapsto \mathcal{U}\mathbb{C}^M$, where $\mathrm{F}_u$ combines the Fourier transform $\mathcal{F}$ and random undersampling operators $\mathcal{U}$.

The experiments utilize the single-coil MRI data for training, although most existing methods utilize the multi-coil MRI data. This is because HP-ALF utilizes the zero-filling images that come from the preprocessing of raw multi-coil or single-coil data, while existing methods utilize the multi-coil MRI data to extract the coil sensitivity [39]. Therefore, it does not affect HP-ALF whether the data input is single-coil or multi-coil in the reconstruction process. Moreover, the single-coil MRI image obtained by data preprocessing is of lower quality at the same acceleration factor [39]. Therefore, it is more challenging for HP-ALF to use the single-coil MRI data as input compared with the multi-coil MRI data. In addition, HP-ALF, similar to other GAN-based methods, can combine the parallel imaging strategy and transfer learning for multichannel imaging [43], [44]. The extension to multi-coil MRI data will be considered in the future studies.

[1] http://masiweb.vuse.vanderbilt.edu/workshop2013/index.php/
[2] https://atriaseg2018.cardiacatlas.org/
[3] https://fastmri.org/dataset/

### B. Evaluation Metrics and Training Details

The evaluation metrics include the peak signal-to-noise ratio (PSNR), the structural similarity index (SSIM), the Fréchet inception distance (FID) and the perceptual similarity measure (PSIM). PSNR evaluates the perceptual quality of reconstruction [20]. SSIM measures the perceptual similarity of images [45]. FID is a similarity measure between two datasets that correlates well with human judgments of visual quality. It evaluates the similarity between the set of generated images and the corresponding fully sampled images [46]. PSIM is a perceptual image quality assessment (IQA) metric based on the human visual system. It evaluates the similarity of local details between the input original and distorted images [47].

HP-ALF uses the Adam optimizer with a batch of four subjects per step and an initial learning rate of 0.0003 during the training process. To balance the weights of different losses in Equation (11) into similar scales, $\alpha$ is set to 15 and $\beta$ is set to 0.1 according to [11]. The learning rate is halved every 5 epochs. Early stopping is used when the validation loss stops decreasing for 50 epochs.

### C. Results

**Comparison with CS-MRI Methods.** We compare HP-ALF with seven CS-MRI methods, including conventional methods and deep-learning-based methods. The conventional methods include the total variation (TV) [14] and ADMM [41]. The deep-learning-based methods include Deep ADMM [42], DAGAN [11], CRNN [18], Sub-GAN [17] and SSL-MRI [34]. All comparison methods and HP-ALF use the baseline zero-filling reconstruction for initialization to achieve the fair comparison. In addition, all comparison methods use the default parameters recommended in their papers.

Figure 4 shows the results of the method comparison under different undersampling conditions. The image data are first transformed to the $k$-space data. Then the $k$-space data are undersampled using three masks: 1D Gaussian (G1D), 2D Gaussian (G2D), and 2D Poisson disc (P2D). Each mask retains 10%, 30%, and 50% of the data to achieve the 10×, 3.3×, and 2× acceleration, respectively. The results show that HP-ALF achieves the best scores in PSNR, SSIM and FID, especially for the 1D Gaussian disc mask at 10× speed-up. The values of PSNR, SSIM, FID and PSIM obtained by HP-ALF are 32.42, 0.94, 90.79, and 0.90, respectively, in brain data; 33.41, 0.95, 80.72, and 0.91, respectively, in cardiac data; and 35.75, 0.99, 77.54, and 0.93, respectively, in knee data. These values are better than the other methods. However, DAGAN, deep ADMM and SSL-MRI can achieve higher PSIM values in knee data as 0.94, 0.93, and 0.94, respectively, at 3.3× and 0.95, 0.94, 0.95, respectively, at 2× speed-up.

Figure 5 shows the visualization of the method comparison for the sample reconstructed images. The results show that HP-ALF returns the sharpest images with fine details in brain data and cardiac data, as apparent from the magnified regions. Although CRNN and DAGAN can also return the sharp images, the reconstruction of the fine tissue structure is less detailed than that reconstructed by HP-ALF. Then, Sub-GAN and SSL-MRI reconstruct the textures with less

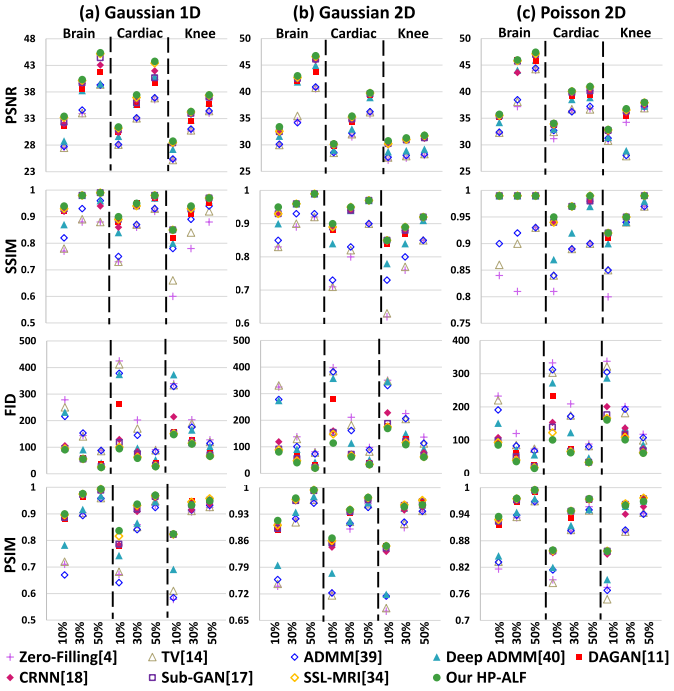


Fig. 4. Quantitative results (PSNR, SSIM, FID and PSIM) of the comparison study using different random undersampling masks (Gaussian 1D, Gaussian 2D and Poisson 2D). 10%, 30%, 50% represent the percentage of the data sampling in the $k$-space data obtained from original image data.

noise, but with unsatisfactory performance in reconstructing the fine details. For example, the areas within the block in Figure 5 show the unrealistic texture details in Sub-GAN and SSL-MRI. Moreover, ADMM and zero-filling reconstruction have difficulty inhibiting the remaining noise-like textures. However, Deep ADMM, DAGAN, Sub-GAN and SSL-MRI can achieve smaller error maps and retain fewer aliasing artifacts. Furthermore, these methods have better PSIM performance in knee data as the values are 0.9252, 0.9477, 0.9356 and 0.9450, respectively. This is because the evaluation of the local and detailed information by PSNR, SSIM and FID are affected by global similarity from the perspective of the single-scale structure. Then this evaluation may be also affected by the nonsalient area (e.g., background) with fewer aliasing artifacts, when the remaining aliasing artifacts mainly appear in the bone areas of the local image In addition, the Knee dataset contains more random noise. The random noise and texture details are indistinguishable with respect to other datasets. Thus, HP-ALF may perceive random noise as the texture detail of the image during learning.

Table I presents the number of parameters and the reconstruction time of all methods. The results show that HP-ALF has the third largest number of parameters and the third fastest speed among all methods. Although Deep ADMM and CRNN have fewer numbers of the parameters than HP-ALF, they do not complete the reconstruction at the same time. Similarly, although DAGAN is faster than HP-ALF, it has more parameters. SSL-MRI can achieve the lowest number of parameters and the fastest speed in all methods, but its performance does not reach that of HP-ALF.

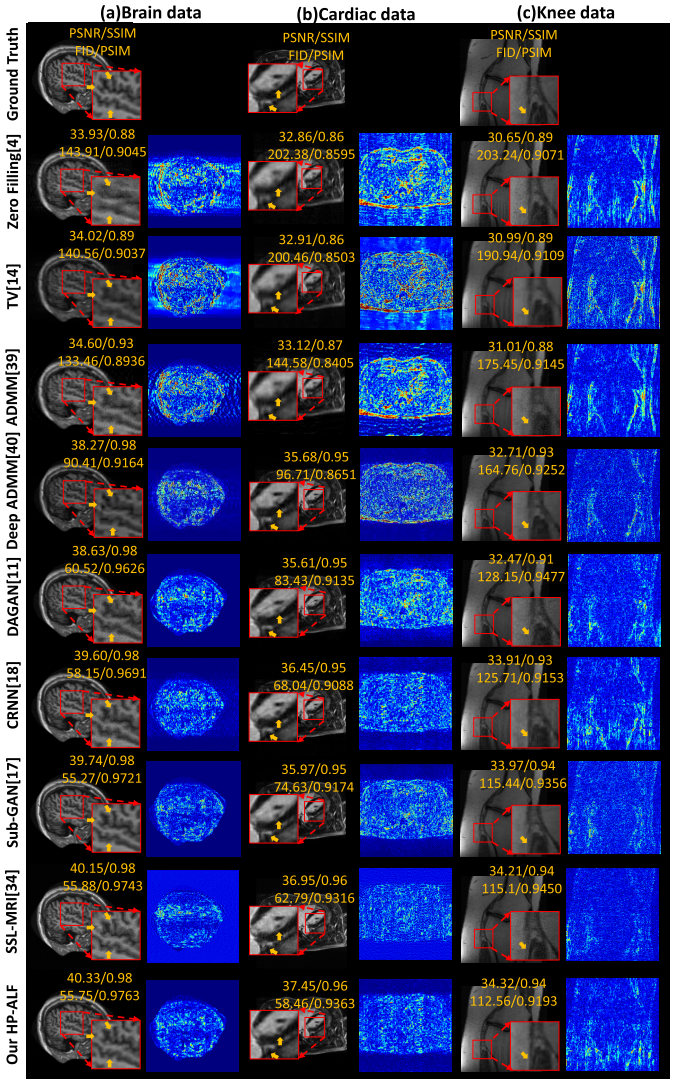


Fig. 5. Qualitative comparison with some representative methods using 30% of the $k$-space data and 1D Gaussian mask. Red boxes illustrate the enlarged view. The right panel images in different datasets illustrate the difference view. The numbers in the images are the PSNR, SSIM, FID and PSIM values.

TABLE I

RECONSTRUCTION TIME AND PARAMETERS OF THE COMPARISON STUDY. "NUM OF PARAM" MEANS THE NUMBERS OF THE PARAMETERS IN THE COMPARED MODELS. "±" MEANS STANDARD DEVIATION

| Methods | Testing Time CPU(sec)/GPU(ms) | Num of Param |
|---|---|---|
| Zero-Filling [4] | 0.002±0.003/ - | - |
| TV [14] | 10.5±1.2/ - | - |
| ADMM [41] | 10.2±2.3/ - | - |
| Deep ADMM [42] | 3.2±0.2/ - | 397.19K |
| DAGAN [11] | 0.2±0.1/5.4±0.1 | 564.76M |
| CRNN [18] | 0.2±0.1/6.3±0.3 | 1.14M |
| Sub-GAN [17] | 0.2±0.1/5.9±0.4 | 566.7M |
| SSL-MRI [34] | 0.2±0.1/4.2±0.5 | 2.68M |
| Our HP-ALF | 0.2±0.1/5.7±0.1 | 217.9M |

**Noise Suppression Comparison.** It compares the noise suppression performance of HP-ALF with that of the comparative methods. Gaussian noise is used because it is suitable to simulate the natural noise of MRI. The natural noise of MRI mainly comes from the thermal noise of the scanned object. Gaussian noise is commonly used for simulating thermal

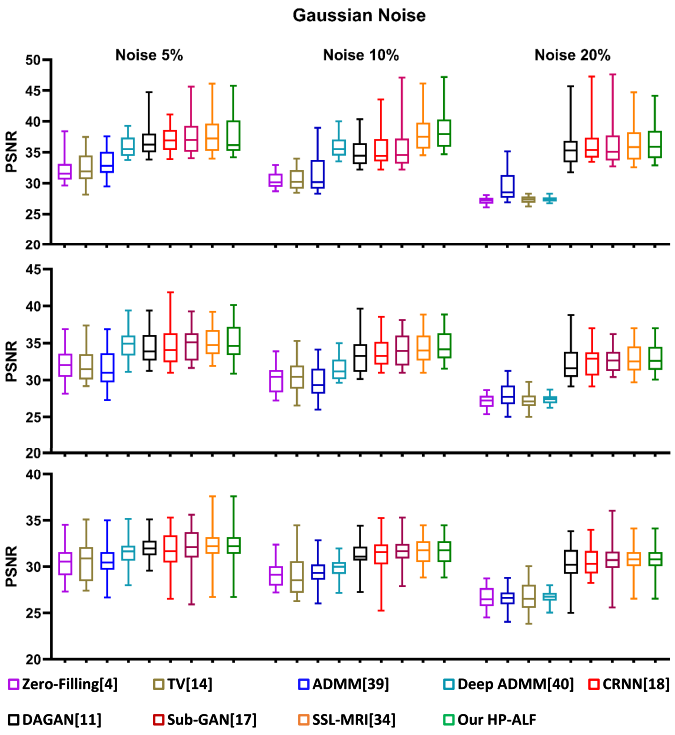


Fig. 6. The comparison of reconstruction performance using Gaussian noise in different noise levels. The comparison uses 1D Gaussian mask and 30% percentage for data sampling. The upper, middle, and below panels show the results of the brain MRI dataset, the cardiac MRI dataset, and the knee MRI dataset respectively.

noise in $k$-space [48], [49], [50]. Moreover, Gaussian noise is also widely used in existing CS-MRI methods [51], [52], [53]. These methods motivate us to insert Gaussian noise in $k$-space for the noise suppression experiment. Specifically, this experiment adds additive white Gaussian noise to the $k$-space before applying the undersampling during training and testing. For the comparative analysis, the Gaussian noise is varied from 5% to 20%. In addition, the comparison study uses a 1D Gaussian mask and is performed on 30% data sampling. Additionally, the noise level estimation method in [54] is applied to evaluate the residual noise after noise suppression.

Figure 6 shows the noise suppression performance of HP-ALF at different noise levels by PSNR. The PSNR of HP-ALF is a higher mean value than those of the other methods. In particular, HP-ALF achieves the best PSNR values in all the datasets at the highest Gaussian noise level (i.e., noise 20%) with 36.28 for the brain data, 32.91 for the cardiac data and 30.78 for the knee data. Figure 7 shows the noise suppression performance of HP-ALF at different noise levels by the residual noise level. Especially for the highest Gaussian noise level (i.e., noise 20%), HP-ALF achieves the lowest value of the residual noise levels in all the datasets with 0.02 for the brain data, 0.04 for the cardiac data, and 0.09 for the knee data. These results indicate that HP-ALF can effectively reduce the residual noise in the reconstructed images.

**Performance of Different Components in HP-ALF.** The ablation study evaluates how changes in the main components of HP-ALF affects its performance. (1) without MPD:

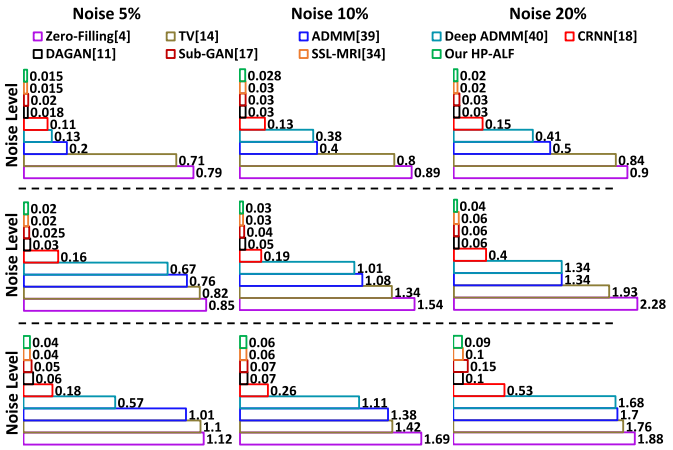


Fig. 7. The comparison of noise reduction performance using Gaussian noise in different noise levels by the residual noise level. The comparison uses 1D Gaussian mask and 30% percentage for data sampling. The upper, middle, and lower panels show the results of the brain MRI dataset, the cardiac dataset, and the knee dataset, respectively. The estimated value of the noise level is proportional to the residual noise.

the method without the multilevel perspective discrimination; (2) without CAL: the method without thecontext-aware learning block; (3) without GLC: the method without the global and local coherent discriminator; (4) HP-ALF: the method with the multilevel perspective discrimination, the context-aware learning block and the global and local coherent discriminator (5) with TAL: the method without the multilevel perspective discrimination and the global and local coherent discriminator, utilizing the traditional adversarial loss instead of the proposed adversarial loss. In addition, we compare the different loss components in Equation (11) to validate the effectiveness of the loss in HP-ALF. (1) HP-ALF: the whole loss function in Equation (11); (2) without MPD: the loss function without $\mathcal{L}_{\text{adv1}}$ (i.e., without the multilevel perspective discrimination); (3) without LfMSE: the loss function without $\mathcal{L}_{\text{fMSE}}$; (4) without Lvgg: the loss function without $\mathcal{L}_{\text{VGG}}$; (5) with TAL: the loss function using the traditional adversarial loss instead of the proposed adversarial loss in Equation (11); (6) without GLC: the loss function without $\mathcal{L}_{\text{adv2}}$ (i.e., without the decoder of the discriminator).

Table II shows the effectiveness of the current configuration in HP-ALF for all evaluation metrics. In particular, the values of PSNR in HP-ALF for the three datasets are 40.33, 37.45, 34.32. The FID values for the Brain dataset, the Cardiac dataset and the Knee dataset are 55.75, 58.46, and 112.56, respectively. These results indicate that the current configuration of HP-ALF achieves superior reconstruction performance.

Figure 9(a) shows that all loss functions increase quickly and converge after several epochs. Moreover, HP-ALF with the whole loss function enables faster convergence with respect to other loss functions. For the loss function without the content loss, HP-ALF without $\mathcal{L}_{\text{VGG}}$ and $\mathcal{L}_{\text{fMSE}}$ obtain the third and fourth fastest convergence, respectively. For the loss function without the proposed adversarial loss, HP-ALF without GLC, HP-ALF without GLC and HP-ALF without MPD obtains the second, fifth and sixth fastest convergence, respectively. These results indicate that the proposed adversarial training loss (i.e., the combination of all above loss functions except traditional adversarial loss) achieves the fastest convergence.

TABLE II
QUANTITATIVE RESULTS (PSNR AND SSIM) OF THE ABLATION STUDY USING 1D GAUSSIAN MASK. 10%, 30%, 50% REPRESENTS THE PERCENTAGES OF THE DATA SAMPLING IN THE *k*-SPACE DATA OBTAINED FROM ORIGINAL IMAGE DATA. "NUM OF PARAM" MEANS THE NUMBERS OF THE PARAMETERS. THE ABBREVIATIONS OF THE COMPARATIVE METHODS ARE EXPLAINED IN SECTION IV-C. "±" MEANS STANDARD DEVIATION

| Task | Mask | Methods | PSNR | SSIM | Num of Param |
|---|---|---|---|---|---|
| Brain data | 10% | with TAL | 28.01±2.31 | 0.87±0.02 | 217.9m |
| | | without MPD | 31.86±3.09 | 0.93±0.02 | 217.9m |
| | | without GLC | 32.04±3.35 | 0.94±0.02 | **217.6m** |
| | | without CAL | 32.19±2.98 | 0.94±0.02 | 217.9m |
| | | HP-ALF | **32.42±3.28** | **0.94±0.02** | 217.9m |
| | 30% | with TAL | 35.01±4.48 | 0.95±0.02 | 217.9m |
| | | without MPD | 39.18±3.40 | 0.98±0.01 | 217.9m |
| | | without GLC | 39.83±3.13 | 0.98±0.01 | **217.6m** |
| | | without CAL | 39.36±3.32 | 0.98±0.01 | 217.9m |
| | | HP-ALF | **40.33±3.37** | **0.98±0.01** | 217.9m |
| | 50% | with TAL | 39.67±2.85 | 0.96±0.01 | 217.9m |
| | | without MPD | 43.36±2.51 | 0.99±0.001 | 217.9m |
| | | without GLC | 44.68±2.54 | 0.99±0.001 | **217.6m** |
| | | without CAL | 44.05±2.10 | 0.99±0.001 | 217.9m |
| | | HP-ALF | **45.38±2.56** | **0.99±0.001** | 217.9m |
| Cardiac data | 10% | with TAL | 29.47±3.16 | 0.87±0.03 | 217.9m |
| | | without MPD | 30.21±2.78 | 0.86±0.02 | 217.9m |
| | | without GLC | 31.19±3.27 | 0.90±0.02 | **217.6m** |
| | | without CAL | 30.86±2.99 | 0.89±0.02 | 217.9m |
| | | HP-ALF | **31.41±3.18** | **0.90±0.02** | 217.9m |
| | 30% | with TAL | 35.43±4.12 | 0.95±0.02 | 217.9m |
| | | without MPD | 36.18±2.56 | 0.95±0.01 | 217.9m |
| | | without GLC | 37.30±2.43 | 0.96±0.01 | **217.6m** |
| | | without CAL | 37.01±2.76 | 0.96±0.01 | 217.9m |
| | | HP-ALF | **37.45±1.99** | **0.96±0.01** | 217.9m |
| | 50% | with TAL | 39.37±2.11 | 0.92±0.01 | 217.9m |
| | | without MPD | 41.98±2.84 | 0.97±0.01 | 217.9m |
| | | without GLC | 42.60±2.50 | 0.97±0.01 | **217.6m** |
| | | without CAL | 42.46±2.74 | 0.97±0.01 | 217.9m |
| | | HP-ALF | **42.77±2.34** | **0.97±0.01** | 217.9m |
| Knee data | 10% | with TAL | 27.45±2.31 | 0.80±0.06 | 217.9m |
| | | without MPD | 27.05±3.39 | 0.83±0.06 | 217.9m |
| | | without GLC | 28.07±3.41 | 0.85±0.06 | **217.6m** |
| | | without CAL | 27.49±2.76 | 0.85±0.06 | 217.9m |
| | | HP-ALF | **28.75±2.22** | **0.85±0.06** | 217.9m |
| | 30% | with TAL | 31.99±1.97 | 0.91±0.03 | 217.9m |
| | | without MPD | 32.98±2.04 | 0.93±0.03 | 217.9m |
| | | without GLC | 34.02±2.13 | 0.94±0.03 | **217.6m** |
| | | without CAL | 33.79±1.79 | 0.93±0.03 | 217.9m |
| | | HP-ALF | **34.32±1.98** | **0.94±0.03** | 217.9m |
| | 50% | with TAL | 34.86±2.71 | 0.95±0.01 | 217.9m |
| | | without MPD | 37.01±2.51 | 0.97±0.01 | 217.9m |
| | | without GLC | 37.28±2.54 | 0.97±0.01 | **217.6m** |
| | | without CAL | 37.19±2.10 | 0.97±0.01 | 217.9m |
| | | HP-ALF | **37.44±2.18** | **0.97±0.01** | 217.9m |

**Performance of Multilevel Perspective Discrimination.** This experiment includes four parts. First, the training process is presented to validate the convergence properties for multilevel perspective discrimination in HP-ALF. Second, the different objectives of GAN are compared with the multilevel perspective discrimination in HP-ALF, including the standard GAN (GAN) [21], WGAN [55], HingeGAN [56] and LSGAN [57]. Third, according to section III-B, increasing the number of the outcomes $\mathbf{v}$ in $p_{\text{perspective}}$ can provide more information to the generator to improve the perceptual quality of the reconstruction results. To this end, the number of $\mathbf{v}$ is changed while keeping the model unchanged to validate that the setting of this number (i.e., the default number is 10) in HP-ALF is

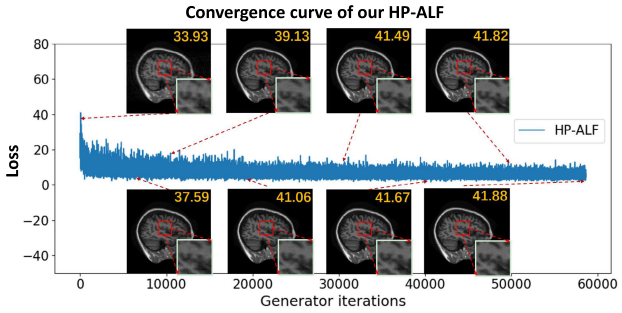


Fig. 8. Convergence validation for multi-level perspective discrimination using training curves and samples. It can be seen a clear correlation between lower error and better sample perception quality. The numbers in the images are the PSNR values.

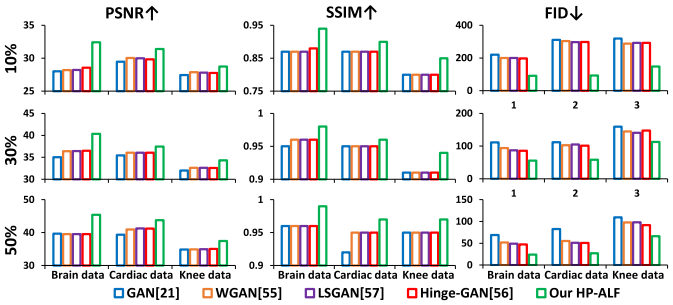


Fig. 10. The ablation study for different GAN objectives functions using 1D Gaussian mask. 10%, 30%, 50% represent the percentages of the data sampling.

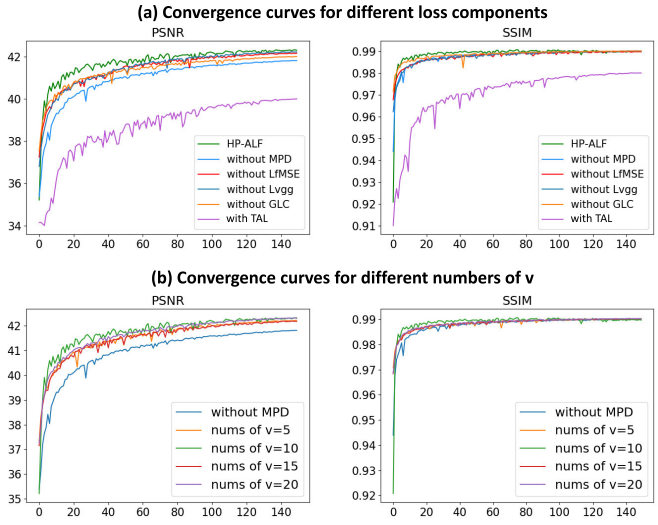


Fig. 9. (a) Convergence curves for different loss components in Equation (11). It can be seen the image quality for all loss functions increase quickly and converge after several epochs. (b) The comparison of the different number of $\mathbf{v}$ in $p_{\text{perspective}}$. When the number of $\mathbf{v}$ exceeds 10 (i.e., the default setting), the image quality does not improve significantly.

appropriate. Finally, the performance of the patch-level reduction of aliasing artifacts is visualized. It shows the relationship by displaying the output of the discriminator (corresponding to $\mathbf{v}$) and the reconstruction error (corresponding to the image region). This is because each element in the output vector of the discriminator corresponds to an image attribute. The experiments set the different elements of this output as zero to make HP-ALF not focus on the reduction of aliasing artifacts in certain image regions.

Figure 8 shows the convergence process of HP-ALF. The figure shows that these curves correlate well with the perceptual quality of the generated samples. At the first 10000 iterations, the loss of the generator significantly decreases, and the fine details in the generated samples (enlarged area in the figure) are sharpened. The loss of the generator converges in a small range in the subsequent process. These results indicate that multilevel perspective discrimination in HP-ALF can reach optimality.

Figure 10 presents the comparison results of the GAN baseline methods. The results indicate that our HP-ALF with multilevel perspective discrimination obtains the better scores in all the evaluation metrics compared to the GAN baseline methods. Especially for the FID metric, the values for the Brain dataset, the Cardiac dataset, and the Knee dataset at the 10-fold undersampling rate are 90.79, 92.79 and 147.54, respectively. At 3.3× speed-up, the FID values for the Brain dataset, the Cardiac dataset, and the Knee dataset are 55.75, 58.46, and 112.56, respectively. In addition, the values for the Brain dataset, the Cardiac dataset, and the Knee dataset at 2× speed-up are 24.09, 27.04 and 66.16, respectively. These results indicate that HP-ALF can achieve better reconstruction performance than other GAN baseline methods.

Figure 9(b) shows that image quality improves as the number of $\mathbf{v}$ increases. The PSNR and SSIM values for HP-ALF with multilevel perspective discrimination are better than those for HP-ALF with TAL during the training process. The above results indicate that $\mathbf{v}$ can provide stronger guidance for $G$ to match the original image distribution. When the number of $\mathbf{v}$ exceeds 10, the image quality is not improved significantly. Moreover, the convergence speed of the model is faster than others when the number of $\mathbf{v}$ is 10.

Figure 11 presents the visual performance of the patch-level reduction of aliasing artifacts. It shows that during the training procedure with the increase of epochs, the overall quality is gradually improved in the entire image. However, the region within the image also changes (where the aliasing artifacts are not significantly reduced), when the selected element (i.e., set to zero) of the output vector changes. Specifically, the image reconstruction at the edge of the brain does not perform well (i.e., has large error), when the selected element is located at the left or right side of the output vector (i.e., the first row and last row in the figure). Then, the image reconstruction at the center of the brain has large errors when the selected element is located at the peak of the output vector (i.e., the second row in the figure). Therefore, these results indicate that different $\mathbf{v}$ corresponds to the attributes in different image regions.

**Performance of Global and Local Coherent Discriminator.** The experiments show the effectiveness of the discriminator by the influence of the per-pixel decision of the discriminator on the reconstruction error of the generator. The per-pixel decision investigates how the discriminator guides the generator to improve perceptual quality.

Figure 12 shows that the background of the per-pixel decision (top row) first became bright, and then the region containing fine details became bright as the epoch increases.

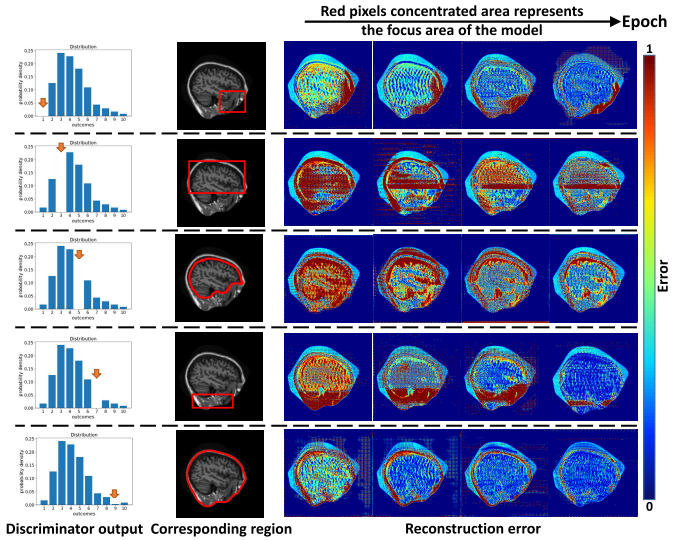


Fig. 11. The visual performance of the patch-level reduction of aliasing artifacts. It shows the relationship between the discriminator output and the reconstruction error. The experiment uses 1D Gaussian mask and is performed on 50% data sampling. The color bar for the difference images is shown on the right. The changes of the selected elements (i.e., the discriminator output) lead to the changes of the image region (i.e., corresponding region). The aliasing artifacts in these image regions are not significantly reduced (i.e., reconstruction error).

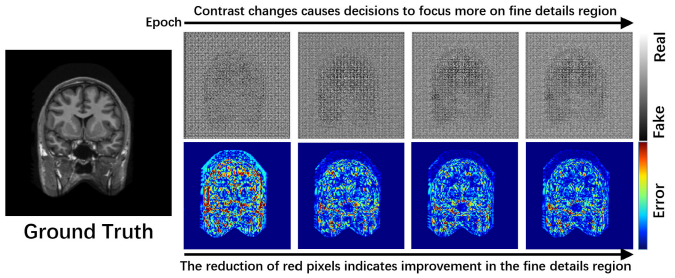


Fig. 12. Correlations between the per-pixel decision (top row) and the difference images (bottom row) investigate how the discriminator guides the generator. Color bars for the difference images and the feedback image are separately shown on the left and right. Darker colors in the per-pixel decision correspond to the discriminator confidence of pixel being fake, and then discriminator provides confidence to the generator to correct the fake regions (as shown in the difference images).

Additionally, the output image becomes bright overall as the epoch increases. The brighter (resp. darker) colors in the per-pixel decision correspond to the discriminator confidence of pixels being real (resp. being fake). These results indicate that the background of the reconstructed image is recognized as fake by the discriminator. The focused area of the discriminator is gradually transferred from the image background to the region with fine details. The per-pixel decision is correlated well with the difference (the bottom row in Figure 12) between the reconstructed image and ground truth. This indicates that the decision of the discriminator can guide the generator during the training process.

**Performance of Context-Aware Learning Generator.** The experiments present the effectiveness of the proposed context-aware generator by considering the kind of generator, the numbers of input slices, and the feature representation map.

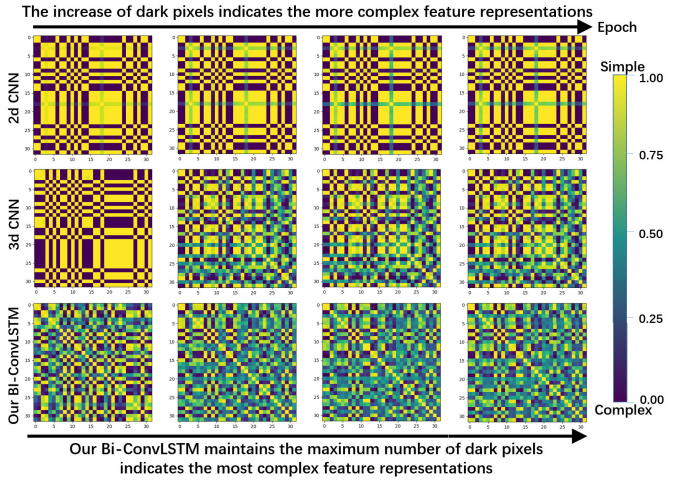


Fig. 13. The comparison of feature representations extracted by the different context-aware learning generators (U-net+2D CNN, U-net+3D CNN and U-net+Bi-ConvLSTM). The 2.5D U-net cannot the feature representations because it does not contain the block. The color of the pixel represents the similarity of the different features evaluated by cosine distance (darker colors indicate complex feature representations and brighter indicate simple representations). The increase of dark pixels shows that the feature representations of the module are gradually enriched.

Table III shows the comparison results of different context-aware learning generators. They include the proposed U-net+Bi-ConvLSTM (HP-ALF with Bi-ConvLSTM), U-net+2D CNN (HP-ALF with 2D CNN), U-net+3D CNN (HP-ALF with 3D CNN) and 2.5D U-net (HP-ALF with 2.5D U-net). Specifically, the 2.5D U-net utilizes the same U-net architecture as those in the comparative generators. The results for HP-ALF with 3D CNN, HP-ALF with 2.5D U-net and HP-ALF with 2D CNN occupy the second, third and fourth places in the most cases, respectively. However, HP-ALF with 2.5D U-net achieves the lowest scores of 28.36, 0.83, 166.12, and 0.82 in PSNR, SSIM, FID and PSIM, respectively, at 10× speed-up for the 5 input slices in the Knee dataset. These results indicate that HP-ALF with Bi-ConvLSTM achieves the best performance with respect to other generators.

Table III also presents the comparison results using different numbers of input slices. The adjacent slices are considered as the input sequence. This table presents the effect of the number of input adjacent slices (set to 3∼7) on different generators. The results show that HP-ALF achieves the best performance when the number of input adjacent slices is 5 for all comparative generators. For example, when the dataset is brain at 10× acceleration, the PSNR, SSIM, FID and PSIM are 32.42, 0.94, 90.79, 0.90 for HP-ALF, 32.18, 0.94, 95.98, 0.90 for HP-ALF with 2.5D U-net, 31.74, 0.93, 94.90, and 0.90, respectively, for HP-ALF with 3D CNN and 31.67, 0.93, 97.81, and 0.88, respectively, for HP-ALF with 2D CNN, respectively. The best performance in the above results is reached when the input has five adjacent slices for all datasets.

Figure 13 shows the feature representation ability of our Bi-ConvLSTM during the training process. It displays the similarity of the feature maps obtained by the context-aware learning block by the cosine distance $d(A, B) = A^T B/\|A\|\|B\| = \cos(\theta)$ [18]. If two feature maps are orthogonal, then

TABLE III

THE ABLATION STUDY WITH DIFFERENT CONTEXT-AWARE LEARNING GENERATORS (U-NET+2D CNN, U-NET+3D CNN, U-NET+BI-CONVLSTM AND 2.5D U-NET) USING 1D GAUSSIAN MASK. HP-ALF WITH BI-CONVLSTM IS OUR METHOD. 10%, 30%, 50% REPRESENT THE PERCENTAGES OF THE DATA SAMPLING. THIS STUDY IS PERFORMED UNDER THE DIFFERENT NUMBERS OF INPUT ADJACENT SLICES (3, 4, 5, 6, 7)

| **3 slice** | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Comparison** | **Brain data** | | | | | | | | | | | |
| | 10% | | | | 30% | | | | 50% | | | |
| | PSNR | SSIM | FID | PSIM | PSNR | SSIM | FID | PSIM | PSNR | SSIM | FID | PSIM |
| HP-ALF with 2D CNN | 31.25 | 0.93 | 98.89 | 0.86 | 38.10 | 0.98 | 67.27 | 0.97 | 42.95 | 0.99 | 35.41 | 0.99 |
| HP-ALF with 3D CNN | 31.59 | 0.93 | 96.49 | 0.89 | 39.82 | 0.98 | 63.49 | 0.97 | 43.89 | 0.99 | 32.71 | 0.99 |
| HP-ALF with 2.5D U-net | 31.53 | 0.93 | 96.78 | 0.89 | 38.72 | 0.98 | 68.20 | 0.97 | 44.05 | 0.99 | 30.03 | 0.99 |
| HP-ALF with Bi-ConvLSTM | **32.20** | **0.94** | **91.56** | **0.90** | **40.05** | **0.98** | **60.65** | **0.98** | **44.88** | **0.99** | **28.74** | **0.99** |
| | **Cardiac data** | | | | | | | | | | | |
| HP-ALF with 2D CNN | 30.33 | 0.88 | 120.79 | 0.85 | 36.11 | 0.95 | 81.19 | 0.95 | 39.20 | 0.97 | 70.98 | 0.96 |
| HP-ALF with 3D CNN | 30.59 | 0.89 | 118.84 | 0.87 | 36.43 | 0.95 | 76.53 | 0.96 | 41.91 | 0.97 | 65.02 | 0.96 |
| HP-ALF with 2.5D U-net | 30.78 | 0.89 | 99.00 | 0.87 | 36.09 | 0.95 | 79.50 | 0.95 | 40.05 | 0.97 | 69.84 | 0.96 |
| HP-ALF with Bi-ConvLSTM | **31.15** | **0.90** | **97.08** | **0.89** | **36.74** | **0.96** | **72.36** | **0.96** | **42.67** | **0.97** | **53.68** | **0.97** |
| | **Knee data** | | | | | | | | | | | |
| HP-ALF with 2D CNN | 28.35 | 0.83 | 155.51 | 0.82 | 32.98 | 0.91 | 134.81 | 0.94 | 35.69 | 0.95 | 85.64 | 0.92 |
| HP-ALF with 3D CNN | 28.46 | 0.83 | 151.64 | 0.82 | 33.17 | 0.92 | 131.83 | 0.95 | 36.21 | 0.95 | 79.85 | 0.92 |
| HP-ALF with 2.5D U-net | 28.23 | 0.82 | 170.11 | 0.82 | 32.08 | 0.90 | 140.16 | 0.95 | 35.40 | 0.95 | 83.03 | 0.92 |
| HP-ALF with Bi-ConvLSTM | **28.55** | **0.85** | **149.68** | **0.82** | **33.37** | **0.92** | **125.43** | **0.96** | **36.37** | **0.96** | **74.03** | **0.93** |
| **4 slice** | | | | | | | | | | | | |
| **Comparison** | **Brain data** | | | | | | | | | | | |
| HP-ALF with 2D CNN | 31.48 | 0.93 | 98.74 | 0.88 | 38.72 | 0.98 | 65.02 | 0.97 | 42.12 | 0.99 | 35.69 | 0.99 |
| HP-ALF with 3D CNN | 31.67 | 0.93 | 95.02 | 0.90 | 39.63 | 0.98 | 63.19 | 0.97 | 43.87 | 0.99 | 30.77 | 0.99 |
| HP-ALF with 2.5D U-net | 31.89 | 0.94 | 96.55 | 0.90 | 39.40 | 0.98 | 66.67 | 0.97 | 43.89 | 0.99 | 31.02 | 0.99 |
| HP-ALF with Bi-ConvLSTM | **32.37** | **0.94** | **90.81** | **0.90** | **40.20** | **0.98** | **58.41** | **0.98** | **45.11** | **0.99** | **25.99** | **0.99** |
| | **Cardiac data** | | | | | | | | | | | |
| HP-ALF with 2D CNN | 30.51 | 0.88 | 112.21 | 0.85 | 36.51 | 0.95 | 80.01 | 0.95 | 40.1 | 0.97 | 60.64 | 0.96 |
| HP-ALF with 3D CNN | 30.51 | 0.89 | 106.5 | 0.87 | 36.88 | 0.95 | 69.90 | 0.96 | 42.49 | 0.97 | 49.76 | 0.96 |
| HP-ALF with 2.5D U-net | 30.96 | 0.89 | 97.01 | 0.88 | 36.10 | 0.95 | 76.21 | 0.95 | 41.55 | 0.97 | 55.4 | 0.96 |
| HP-ALF with Bi-ConvLSTM | **31.31** | **0.90** | **95.68** | **0.89** | **36.98** | **0.96** | **64.21** | **0.96** | **43.39** | **0.98** | **33.64** | **0.97** |
| | **Knee data** | | | | | | | | | | | |
| HP-ALF with 2D CNN | 28.37 | 0.82 | 150.19 | 0.82 | 33.10 | 0.91 | 130.71 | 0.94 | 36.50 | 0.96 | 77.96 | 0.92 |
| HP-ALF with 3D CNN | 28.51 | 0.82 | 152.11 | 0.82 | 33.59 | 0.92 | 128.43 | 0.95 | 37.01 | 0.96 | 75.12 | 0.92 |
| HP-ALF with 2.5D U-net | 28.26 | 0.83 | 162.00 | 0.82 | 32.88 | 0.91 | 131.11 | 0.95 | 36.12 | 0.96 | 81.11 | 0.92 |
| HP-ALF with Bi-ConvLSTM | **28.60** | **0.85** | **144.99** | **0.82** | **34.20** | **0.92** | **111.36** | **0.96** | **37.11** | **0.97** | **73.10** | **0.94** |
| **5 slice** | | | | | | | | | | | | |
| **Comparison** | **Brain data** | | | | | | | | | | | |
| HP-ALF with 2D CNN | 31.67 | 0.93 | 97.81 | 0.88 | 38.92 | 0.98 | 64.04 | 0.97 | 42.95 | 0.99 | 34.55 | 0.99 |
| HP-ALF with 3D CNN | 31.74 | 0.93 | 94.90 | 0.90 | 39.52 | 0.98 | 61.43 | 0.97 | 43.89 | 0.99 | 28.19 | 0.99 |
| HP-ALF with 2.5D U-net | 32.18 | 0.94 | 95.98 | 0.90 | 39.62 | 0.98 | 62.07 | 0.98 | 44.05 | 0.99 | 25.97 | 0.99 |
| HP-ALF with Bi-ConvLSTM | **32.42** | **0.94** | **90.79** | **0.90** | **40.33** | **0.98** | **55.75** | **0.98** | **45.38** | **0.99** | **24.09** | **0.99** |
| | **Cardiac data** | | | | | | | | | | | |
| HP-ALF with 2D CNN | 30.57 | 0.88 | 120.79 | 0.79 | 36.61 | 0.95 | 76.71 | 0.90 | 40.76 | 0.97 | 34.83 | 0.96 |
| HP-ALF with 3D CNN | 30.59 | 0.89 | 118.84 | 0.79 | 37.15 | 0.95 | 67.43 | 0.91 | 42.96 | 0.98 | 31.69 | 0.96 |
| HP-ALF with 2.5D U-net | 31.12 | 0.89 | 96.30 | 0.82 | 36.37 | 0.95 | 73.08 | 0.90 | 42.46 | 0.97 | 33.28 | 0.96 |
| HP-ALF with Bi-ConvLSTM | **31.41** | **0.90** | **94.63** | **0.84** | **37.45** | **0.95** | **58.46** | **0.94** | **43.77** | **0.98** | **27.04** | **0.97** |
| | **Knee data** | | | | | | | | | | | |
| HP-ALF with 2D CNN | 28.43 | 0.82 | 155.43 | 0.82 | 33.47 | 0.92 | 127.15 | 0.97 | 36.71 | 0.97 | 75.30 | 0.92 |
| HP-ALF with 3D CNN | 28.59 | 0.82 | 153.20 | 0.82 | 34.11 | 0.92 | 123.82 | 0.97 | 37.14 | 0.97 | 72.82 | 0.93 |
| HP-ALF with 2.5D U-net | 28.36 | 0.83 | 166.12 | 0.82 | 33.27 | 0.92 | 130.76 | 0.97 | 36.68 | 0.97 | 77.53 | 0.92 |
| HP-ALF with Bi-ConvLSTM | **28.75** | **0.85** | **147.54** | **0.82** | **34.42** | **0.94** | **112.56** | **0.97** | **37.44** | **0.97** | **66.16** | **0.94** |
| **6 slice** | | | | | | | | | | | | |
| **Comparison** | **Brain data** | | | | | | | | | | | |
| HP-ALF with 2D CNN | 31.33 | 0.93 | 97.49 | 0.88 | 38.77 | 0.98 | 66.74 | 0.97 | 42.9 | 0.99 | 37.77 | 0.99 |
| HP-ALF with 3D CNN | 31.65 | 0.93 | 94.31 | 0.90 | 39.41 | 0.98 | 62.26 | 0.97 | 43.78 | 0.99 | 31.45 | 0.99 |
| HP-ALF with 2.5D U-net | 32.02 | 0.93 | 92.36 | 0.89 | 39.52 | 0.98 | 63.11 | 0.98 | 44.08 | 0.99 | 28.11 | 0.99 |
| HP-ALF with Bi-ConvLSTM | **32.22** | **0.94** | **96.41** | **0.90** | **40.25** | **0.98** | **58.88** | **0.98** | **45.24** | **0.99** | **28.45** | **0.99** |
| | **Cardiac data** | | | | | | | | | | | |
| HP-ALF with 2D CNN | 30.24 | 0.89 | 115.40 | 0.85 | 36.37 | 0.95 | 70.42 | 0.95 | 40.19 | 0.97 | 51.49 | 0.96 |
| HP-ALF with 3D CNN | 30.51 | 0.89 | 108.89 | 0.87 | 37.05 | 0.95 | 68.76 | 0.95 | 42.60 | 0.98 | 53.12 | 0.96 |
| HP-ALF with 2.5D U-net | 30.02 | 0.89 | 97.20 | 0.87 | 36.16 | 0.95 | 75.43 | 0.95 | 42.01 | 0.97 | 50.49 | 0.96 |
| HP-ALF with Bi-ConvLSTM | **31.30** | **0.90** | **96.44** | **0.89** | **37.39** | **0.96** | **58.10** | **0.96** | **43.51** | **0.98** | **44.41** | **0.97** |
| | **Knee data** | | | | | | | | | | | |
| HP-ALF with 2D CNN | 28.31 | 0.82 | 159.74 | 0.82 | 33.29 | 0.92 | 119.96 | 0.97 | 36.60 | 0.97 | 74.33 | 0.92 |
| HP-ALF with 3D CNN | 28.40 | 0.82 | 153.69 | 0.82 | 34.05 | 0.92 | 118.47 | 0.97 | 37.02 | 0.97 | 70.11 | 0.93 |
| HP-ALF with 2.5D U-net | 28.29 | 0.83 | 155.48 | 0.82 | 33.25 | 0.92 | 121.21 | 0.97 | 36.43 | 0.96 | 77.90 | 0.92 |
| HP-ALF with Bi-ConvLSTM | **28.72** | **0.85** | **145.63** | **0.82** | **34.29** | **0.94** | **115.49** | **0.97** | **37.30** | **0.97** | **68.63** | **0.94** |
| **7 slice** | | | | | | | | | | | | |
| **Comparison** | **Brain data** | | | | | | | | | | | |
| HP-ALF with 2D CNN | 31.25 | 0.93 | 101.22 | 0.88 | 38.74 | 0.98 | 72.91 | 0.97 | 42.63 | 0.99 | 39.51 | 0.99 |
| HP-ALF with 3D CNN | 31.69 | 0.93 | 97.55 | 0.90 | 39.36 | 0.98 | 65.79 | 0.97 | 43.77 | 0.99 | 32.67 | 0.99 |
| HP-ALF with 2.5D U-net | 31.88 | 0.93 | 96.71 | 0.89 | 39.50 | 0.98 | 67.81 | 0.98 | 44.01 | 0.99 | 28.92 | 0.99 |
| HP-ALF with Bi-ConvLSTM | **32.19** | **0.94** | **92.31** | **0.90** | **40.21** | **0.98** | **61.73** | **0.98** | **45.19** | **0.99** | **27.06** | **0.99** |
| | **Cardiac data** | | | | | | | | | | | |
| HP-ALF with 2D CNN | 30.10 | 0.89 | 118.70 | 0.85 | 36.28 | 0.95 | 84.36 | 0.95 | 40.07 | 0.97 | 59.96 | 0.96 |
| HP-ALF with 3D CNN | 30.68 | 0.89 | 110.44 | 0.87 | 37.02 | 0.95 | 71.28 | 0.95 | 42.40 | 0.98 | 55.69 | 0.96 |
| HP-ALF with 2.5D U-net | 30.94 | 0.89 | 99.40 | 0.87 | 36.05 | 0.95 | 77.79 | 0.95 | 41.22 | 0.97 | 57.25 | 0.96 |
| HP-ALF with Bi-ConvLSTM | **31.28** | **0.90** | **96.79** | **0.88** | **37.31** | **0.96** | **58.98** | **0.96** | **43.24** | **0.98** | **51.44** | **0.97** |
| | **Knee data** | | | | | | | | | | | |
| HP-ALF with 2D CNN | 28.27 | 0.82 | 163.79 | 0.82 | 33.23 | 0.92 | 127.67 | 0.97 | 36.54 | 0.97 | 77.80 | 0.92 |
| HP-ALF with 3D CNN | 28.34 | 0.82 | 159.88 | 0.82 | 34.01 | 0.92 | 125.43 | 0.97 | 36.98 | 0.97 | 75.01 | 0.93 |
| HP-ALF with 2.5D U-net | 28.25 | 0.83 | 169.48 | 0.82 | 33.21 | 0.92 | 129.77 | 0.97 | 36.25 | 0.96 | 80.66 | 0.92 |
| HP-ALF with Bi-ConvLSTM | **28.71** | **0.85** | **150.83** | **0.82** | **34.24** | **0.94** | **117.66** | **0.97** | **37.21** | **0.97** | **71.47** | **0.94** |

$\cos(\theta) = 0$ and if two feature maps are linearly correlated, then $\cos(\theta) = 1$ (the dark colors correspond to be orthogonal whereas bright colors correspond to being linearly correlated). When the values of the cosine distance are smaller, these feature maps are less similar. This means the feature maps learned from a context-aware learning block are diverse and complex. This experiment obtain 32 feature maps for each baseline context-aware learning block in the generator (i.e., 2D CNN, 3D CNN and Bi-ConvLSTM). The generator with the 2.5D U-net cannot the similarity between the feature maps because it does not contain the context-aware learning block. The results show that our Bi-ConvLSTM has more dark pixels than the other baseline blocks. Especially at epoch 150, Bi-ConvLSTM obtains the most dark pixels. It also obtains more dark pixels than the other baseline blocks when the epoch is the same. Then, the mean values of the feature map for the Bi-ConvLSTM, 3D CNN and 2D CNN are 0.43, 0.53 and 0.66, respectively. These results indicate that the Bi-ConvLSTM from the context-aware learning generators learning rich feature information leads to the small average value and complex feature representation of the feature map.

## V. Conclusion

In this paper, we propose a hierarchical perception adversarial learning framework to reconstruct high-quality MRI images. This framework is implemented by the proposed GAN architecture, including the multilevel perspective discrimination, the global and local coherent discriminator and the context-aware learning generator. Specifically, the multilevel perspective discrimination provides the information for adversarial learning from overall and regional perspectives. Then, the global and local coherent discriminator enables both global and local feedback to the generator. Finally, the context-aware generator builds relationships among successive MRI slices to improve the reconstruction performance. Extensive experiments on three datasets (brain, heart and knee) show the effectiveness of our framework on high-quality MRI reconstruction.

## Appendix I
## Details of Multilevel Perspective Discrimination

The theoretical analysis for multilevel perspective discrimination is shown as follows:

*Theorem 1:* When $G$ is fixed, for any outcome $\mathbf{v}$ and input sample $x$, the optimal discriminator $D$ satisfies

$$D_G^\star(\boldsymbol{x}, v) = \frac{p_{\text{data}}(\boldsymbol{x})}{p_{\text{data}}(\boldsymbol{x}) + p_g(\boldsymbol{x})} + \frac{\mathcal{R}_1(v)\, p_{\text{data}}(\boldsymbol{x}) + \mathcal{R}_0(v)\, p_g(\boldsymbol{x})}{p_{\text{data}}(\boldsymbol{x}) + p_g(\boldsymbol{x})}$$

*Proof*: Given a fixed $G$, the objective of $D$ is:

$$\begin{aligned}\max_G \min_D V(G, D) &= \mathbb{E}_{\boldsymbol{x}\sim p_{\text{data}}}\left[\mathcal{D}_{\text{KL}}\left(\mathcal{R}_1(\mathbf{v}) \| D(\boldsymbol{x})\right) + \log(D(\boldsymbol{x}))\right] \\ &+ \mathbb{E}_{\boldsymbol{x}\sim p_g}\left[\mathcal{D}_{\text{KL}}\left(\mathcal{R}_0(\mathbf{v}) \| D(\boldsymbol{x})\right) + \log(1 - D(\boldsymbol{x}))\right] \\ &= -\int_{\boldsymbol{x}} \left(p_{\text{data}}(\boldsymbol{x}) h\left(\mathcal{R}_1\right) + p_g(\boldsymbol{x}) h\left(\mathcal{R}_0\right)\right) dx \\ &- \int_{\boldsymbol{x}}\int_{v} \left(p_{\text{data}}(\boldsymbol{x})\mathcal{R}_1(v) + p_g(\boldsymbol{x})\mathcal{R}_0(v)\right) \log D(\boldsymbol{x}, v) dv dx,\end{aligned}$$

where $h(\mathcal{R}_1)$ and $h(\mathcal{R}_0)$ are their entropies. However, the first term as $C_1$ in the above equation is irrelevant to $D$, the objective thus is equivalent to

$$\begin{aligned}\min_D V(G, D) &= -\int_{\boldsymbol{x}} \left(p_{\text{data}}(\boldsymbol{x}) + p_g(\boldsymbol{x})\right) \\ &\int_v \frac{p_{\text{data}}(\boldsymbol{x})\mathcal{R}_1(v) + p_g(\boldsymbol{x})\mathcal{R}_0(v)}{p_{\text{data}}(\boldsymbol{x}) + p_g(\boldsymbol{x})} \log D(\boldsymbol{x}, v) dv dx + C_1 \\ &+ \int_{\boldsymbol{x}} \left(p_{\text{data}}(\boldsymbol{x}) + p_g(\boldsymbol{x})\right) \\ &\times \int_v \frac{p_{\text{data}}(\boldsymbol{x})}{p_{\text{data}}(\boldsymbol{x}) + p_g(\boldsymbol{x})} \log D(\boldsymbol{x}, v) dv dx,\end{aligned}$$

where $p_{\boldsymbol{x}}(v) = \frac{p_{\text{data}}(\boldsymbol{x})(\mathcal{R}_1(v)+1) + p_g(\boldsymbol{x})\mathcal{R}_0(v)}{p_{\text{data}}(\boldsymbol{x}) + p_g(\boldsymbol{x})}$ is a distribution defined on $\Omega$. The third term as $C_2$ is divided from the $C_1$. Let $C_3 = p_{\text{data}}(\boldsymbol{x}) + p_g(\boldsymbol{x})$, we then have

$$\begin{aligned}\min_D V(G, D) &= C_1 + C_2 \\ &+ \int_{\boldsymbol{x}} C_3 (-\int_v p_{\boldsymbol{x}}(v) \log D(\boldsymbol{x}, v) dv + h\left(p_{\boldsymbol{x}}\right) - h\left(p_{\boldsymbol{x}}\right)) dx\end{aligned}$$

For any valid $x$ in the above equation, when $\mathcal{D}_{\text{KL}}\left(p_{\boldsymbol{x}} \| D(\boldsymbol{x})\right)$ achieves its minimum, $D$ obtains its optimal $D^\star$, leading to $D^\star(\boldsymbol{x}) = p_{\boldsymbol{x}}(v)$, which concludes the proof.

*Theorem 2:* When $D = D_G^\star$, and there exists an outcome $v \in \Omega$ such that $\mathcal{R}_1(v) \neq \mathcal{R}_0(v)$, the maximum of $V\left(G, D_G^\star\right)$ is achieved if and only if $p_g = p_{\text{data}}$

*Proof*: When $p_g(\boldsymbol{x}) = p_{\text{data}}(\boldsymbol{x})$, $D_G^\star(\boldsymbol{x}, v) = \frac{(\mathcal{R}_1(v)+1)+\mathcal{R}_0(v)}{2}$, we have

$$V^\star\left(G, D_G^\star\right) = \mathcal{D}_{\text{KL}}\left(\mathcal{R}_1 \| D^\star(\boldsymbol{x})\right) + \mathcal{D}_{\text{KL}}\left(\mathcal{R}_0 \| D^\star(\boldsymbol{x})\right).$$

Subtracting $V^\star\left(G, D_G^\star\right)$ from $V\left(G, D_G^\star\right)$ gives

$$\begin{aligned}V'\left(G, D_G^\star\right) &= V\left(G, D_G^\star\right) - V^\star\left(G, D_G^\star\right) \\ &= -2\mathcal{D}_{\text{KL}}\Big(\frac{p_{\text{data}}(\boldsymbol{x})\mathcal{R}_1 + p_g(\boldsymbol{x})\mathcal{R}_0}{2} \\ &\quad \times \| \frac{\left(p_{\text{data}}(\boldsymbol{x}) + p_g(\boldsymbol{x})\right)\left(\mathcal{R}_1 + \mathcal{R}_0 - 1\right)\frac{1}{2}}{4}\Big)\end{aligned}$$

Since $V^\star\left(G, D_G^\star\right)$ is a constant with respect to G, maximising $V\left(G, D_G^\star\right)$ is equivalent to maximising $V'\left(G, D_G^\star\right)$. The optimal $V'\left(G, D_G^\star\right)$ is achieved if and only if the KL divergence reaches its minimum, where

$$\frac{p_{\text{data}}(\boldsymbol{x})\mathcal{R}_1 + p_g(\boldsymbol{x})\mathcal{R}_0}{2} = \frac{\left(p_{\text{data}}(\boldsymbol{x}) + p_g(\boldsymbol{x})\right)\left(\mathcal{R}_1 + \mathcal{R}_0\right)\frac{1}{2}}{4}$$

$$\left(p_{\text{data}}(\boldsymbol{x}) - p_g(\boldsymbol{x})\right)\left(\mathcal{R}_1 - \mathcal{R}_0\right)\frac{1}{2} = 0$$

for any valid $\boldsymbol{x}$ and $\mathbf{v}$. Hence, as long as there exists a valid $\mathbf{v}$ that $\mathcal{R}_1(v) \neq \mathcal{R}_0(v)$, we have $p_{\text{data}}(\boldsymbol{x}) = p_g(\boldsymbol{x})$ for any valid $\boldsymbol{x}$. If one views the above equation as a cost function to minimise, when $p_{\text{data}}(\boldsymbol{x}) \neq p_g(\boldsymbol{x})$, the larger the difference between $\mathcal{R}_1(v)$ and $\mathcal{R}_0(v)$ is, the stronger the constraint on $G$ becomes.

## REFERENCES


[1] O. N. Jaspan, R. Fleysher, and M. L. Lipton, "Compressed sensing MRI: A review of the clinical literature," *Brit. J. Radiol.*, vol. 88, no. 1056, Dec. 2015, Art. no. 20150487.

[2] L. W. Mann et al., "Accelerating MR imaging liver steatosis measurement using combined compressed sensing and parallel imaging: A quantitative evaluation," *Radiology*, vol. 278, no. 1, pp. 247–256, Jan. 2016.

[3] R. Zhou et al., "Simple motion correction strategy reduces respiratory-induced motion artifacts for *k*-t accelerated and compressed-sensing cardiovascular magnetic resonance perfusion imaging," *J. Cardiovascular Magn. Reson.*, vol. 20, no. 1, pp. 1–13, Dec. 2018.

[4] M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly, "Compressed sensing MRI," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 72–82, Mar. 2008.

[5] S. Mönch, N. Sollmann, A. Hock, C. Zimmer, J. S. Kirschke, and D. M. Hedderich, "Magnetic resonance imaging of the brain using compressed sensing—Quality assessment in daily clinical routine," *Clin. Neuroradiol.*, vol. 30, no. 2, pp. 279–286, Jun. 2020.

[6] J. Y. Cheng et al., "Comprehensive motion-compensated highly accelerated 4D flow MRI with ferumoxytol enhancement for pediatric congenital heart disease," *J. Magn. Reson. Imag.*, vol. 43, no. 6, pp. 1355–1368, Jun. 2016.

[7] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Jan. 2006.

[8] B. M. A. Delattre, S. Boudabbous, C. Hansen, A. Neroladaki, A.-L. Hachulla, and M. I. Vargas, "Compressed sensing MRI of different organs: Ready for clinical daily practice?" *Eur. Radiol.*, vol. 30, no. 1, pp. 308–319, Jan. 2020.

[9] R. L. Robertson, S. Silk, K. Ecklund, S. D. Bixby, S. D. Voss, and C. D. Robson, "Imaging optimization in children," *J. Amer. College Radiol.*, vol. 15, no. 3, pp. 440–443, 2018.

[10] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magn. Reson. Med.*, vol. 58, no. 6, pp. 1182–1195, 2007.

[11] G. Yang et al., "DAGAN: Deep de-aliasing generative adversarial networks for fast compressed sensing MRI reconstruction," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1310–1321, Jun. 2017.

[12] M. Mardani et al., "Deep generative adversarial neural networks for compressive sensing MRI," *IEEE Trans. Med. Imag.*, vol. 38, no. 1, pp. 167–179, Jan. 2019.

[13] Y. Chen et al., "AI-based reconstruction for fast MRI—A systematic review and meta-analysis," *Proc. IEEE*, vol. 110, no. 2, pp. 224–245, Feb. 2022.

[14] K. T. Block, M. Uecker, and J. Frahm, "Undersampled radial MRI with multiple coils. Iterative image reconstruction using a total variation constraint," *Magn. Reson. Med.*, vol. 57, no. 6, pp. 1086–1098, Jun. 2007.

[15] H. Jung, K. Sung, K. S. Nayak, E. Y. Kim, and J. C. Ye, "*k*-t FOCUSS: A general compressed sensing framework for high resolution dynamic MRI," *Magn. Reson. Med.*, vol. 61, no. 1, pp. 103–116, 2009.

[16] C. Feng, Y. Yan, H. Fu, L. Chen, and Y. Xu, "Task transformer network for joint MRI reconstruction and super-resolution," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*, 2021, pp. 307–317.

[17] R. Shaul, I. David, O. Shitrit, and T. R. Raviv, "Subsampled brain MRI reconstruction by generative adversarial neural networks," *Med. Image Anal.*, vol. 65, Oct. 2020, Art. no. 101747.

[18] C. Qin, J. Schlemper, J. Caballero, A. N. Price, J. V. Hajnal, and D. Rueckert, "Convolutional recurrent neural networks for dynamic MR image reconstruction," *IEEE Trans. Med. Imag.*, vol. 38, no. 1, pp. 280–290, Jan. 2019.

[19] J. Huang et al., "Swin transformer for fast MRI," *Neurocomputing*, vol. 493, pp. 281–304, Jul. 2022.

[20] Y. Blau and T. Michaeli, "The perception-distortion tradeoff," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6228–6237.

[21] I. Goodfellow et al., "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, 2014, pp. 2672–2680.

[22] T. M. Quan, T. Nguyen-Duc, and W.-K. Jeong, "Compressed sensing MRI reconstruction using a generative adversarial network with a cyclic loss," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1488–1497, Jun. 2018.

[23] Y. Xiangli, Y. Deng, B. Dai, C. C. Loy, and D. Lin, "Real or not real, that is the question," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2020, pp. 1–12.

[24] Y. Guo, C. Wang, H. Zhang, and G. Yang, "Deep attentive Wasserstein generative adversarial networks for MRI reconstruction with recurrent context-awareness," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*, 2020, pp. 167–177.

[25] Y. Lou, T. Zeng, S. Osher, and J. Xin, "A weighted difference of anisotropic and isotropic total variation model for image processing," *SIAM J. Imag. Sci.*, vol. 8, no. 3, pp. 1798–1823, 2015.

[26] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[27] L. Ma, L. Moisan, J. Yu, and T. Zeng, "A dictionary learning approach for Poisson image deblurring," *IEEE Trans. Med. Imag.*, vol. 32, no. 7, pp. 1277–1289, Jul. 2013.

[28] Z. Gao et al., "Learning physical properties in complex visual scenes: An intelligent machine for perceiving blood flow dynamics from static CT angiography imaging," *Neural Netw.*, vol. 123, pp. 82–93, Mar. 2020.

[29] A. Güngör, B. Askin, D. A. Soydan, E. U. Saritas, C. B. Top, and T. Çukur, "TranSMS: Transformers for super-resolution calibration in magnetic particle imaging," *IEEE Trans. Med. Imag.*, vol. 41, no. 12, pp. 3562–3574, Dec. 2022.

[30] S. Guo, L. Xu, C. Feng, H. Xiong, Z. Gao, and H. Zhang, "Multi-level semantic adaptation for few-shot segmentation on cardiac image sequences," *Med. Image Anal.*, vol. 73, Oct. 2021, Art. no. 102170.

[31] A. B. Szczotka, D. I. Shakir, M. J. Clarkson, S. P. Pereira, and T. Vercauteren, "Zero-shot super-resolution with a physically-motivated downsampling kernel for endomicroscopy," *IEEE Trans. Med. Imag.*, vol. 40, no. 7, pp. 1863–1874, Jul. 2021.

[32] Z. Gao et al., "Privileged modality distillation for vessel border detection in intracoronary imaging," *IEEE Trans. Med. Imag.*, vol. 39, no. 5, pp. 1524–1534, May 2020.

[33] P. Guo, J. M. J. Valanarasu, P. Wang, J. Zhou, S. Jiang, and V. M. Patel, "Over-and-under complete convolutional RNN for MRI reconstruction," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*, 2021, pp. 13–23.

[34] C. Hu, C. Li, H. Wang, Q. Liu, H. Zheng, and S. Wang, "Self-supervised learning for MRI reconstruction with a parallel network training framework," in *Proc. Int. Conf. Image Comput. Comput. Assist. Intervent. (MICCAI)*. Cham, Switzerland: Springer, 2021, pp. 382–391.

[35] C. Wu et al., "Vessel-GAN: Angiographic reconstructions from myocardial CT perfusion with explainable generative adversarial networks," *Future Gener. Comput. Syst.*, vol. 130, pp. 128–139, May 2022.

[36] Y. Korkmaz, S. U. H. Dar, M. Yurt, M. Özbey, and T. Cukur, "Unsupervised MRI reconstruction via zero-shot learned adversarial transformers," *IEEE Trans. Med. Imag.*, vol. 41, no. 7, pp. 1747–1763, Jul. 2022.

[37] H. Wei, Z. Li, S. Wang, and R. Li, "Undersampled multi-contrast MRI reconstruction based on double-domain generative adversarial network," *IEEE J. Biomed. Health Informat.*, vol. 26, no. 9, pp. 4371–4377, Sep. 2022.

[38] E. Schonfeld, B. Schiele, and A. Khoreva, "A U-Net based discriminator for generative adversarial networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 8207–8216.

[39] J. Zbontar et al., "FastMRI: An open dataset and benchmarks for accelerated MRI," 2018, *arXiv:1811.08839*.

[40] C. M. Sandino et al., "Compressed sensing: From research to clinical practice with deep neural networks: Shortening scan times for magnetic resonance imaging," *IEEE Signal Process. Mag.*, vol. 37, no. 1, pp. 117–127, Jan. 2020.

[41] J. Yang, Y. Zhang, and W. Yin, "A fast alternating direction method for TVL1-L2 signal reconstruction from partial Fourier data," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 288–297, Apr. 2010.

[42] Y. Yang, J. Sun, H. Li, and Z. Xu, "Deep ADMM-Net for compressive sensing MRI," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, 2016, pp. 10–18.

[43] J. Lv, C. Wang, and G. Yang, "PIC-GAN: A parallel imaging coupled generative adversarial network for accelerated multi-channel MRI reconstruction," *Diagnostics*, vol. 11, no. 1, p. 61, Jan. 2021.

[44] J. Lv et al., "Transfer learning enhanced generative adversarial networks for multi-channel MRI reconstruction," *Comput. Biol. Med.*, vol. 134, Jul. 2021, Art. no. 104504.

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

[46] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "GANs trained by a two time-scale update rule converge to a local Nash equilibrium," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, 2017, pp. 6626–6637.

[47] K. Gu, L. Li, H. Lu, X. Min, and W. Lin, "A fast reliable image quality predictor by fusing micro-and macro-structures," *IEEE Trans. Ind. Electron.*, vol. 64, no. 5, pp. 3903–3912, May 2017.

[48] S. V. M. Sagheer and S. N. George, "A review on medical image denoising algorithms," *Biomed. Signal Process. Control*, vol. 61, Aug. 2020, Art. no. 102036.

[49] S. Aja-Fernández and G. Vegas-Sánchez-Ferrero, *Statistical Noise Models for MRI*. Cham, Switzerland: Springer, 2016, pp. 31–71.

[50] R. M. Henkelman, "Measurement of signal intensities in the presence of noise in MR images," *Med. Phys.*, vol. 12, no. 2, pp. 232–233, 1985.

[51] D. You, J. Zhang, J. Xie, B. Chen, and S. Ma, "COAST: Controllable arbitrary-sampling network for compressive sensing," *IEEE Trans. Image Process.*, vol. 30, pp. 6066–6080, 2021.

[52] R. Liu, Y. Zhang, S. Cheng, Z. Luo, and X. Fan, "A deep framework assembling principled modules for CS-MRI: Unrolling perspective, convergence behaviors, and practical modeling," *IEEE Trans. Med. Imag.*, vol. 39, no. 12, pp. 4150–4163, Dec. 2020.

[53] P. Deora, B. Vasudeva, S. Bhattacharya, and P. M. Pradhan, "Structure preserving compressive sensing MRI reconstruction using generative adversarial networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 522–523.

[54] X. Liu, M. Tanaka, and M. Okutomi, "Single-image noise level estimation for blind denoising," *IEEE Trans. Image Process.*, vol. 22, no. 12, pp. 5226–5237, Dec. 2013.

[55] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein GAN," 2017, *arXiv:1701.07875*.

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

[57] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, Z. Wang, and S. P. Smolley, "Least squares generative adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2794–2802.