Article

SelfCoLearn: Self-Supervised Collaborative Learning for Accelerating Dynamic MR Imaging

Juan Zou [1,2], Cheng Li [2], Sen Jia [2], Ruoyou Wu [2], Tingrui Pei [1,3,*], Hairong Zheng [2] and Shanshan Wang [2,4,*]

[1] School of Physics and Optoelectronics, Xiangtan University, Xiangtan 411105, China
[2] Paul C. Lauterbur Research Center for Biomedical Imaging, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China
[3] College of Information Science and Technology, Jinan University, Guangzhou 510631, China
[4] Guangdong Provincial Key Laboratory of Artificial Intelligence in Medicial Image Analysis and Application, Shenzhen 518055, China
[*] Correspondence: peitr@163.com (T.P.); ss.wang@siat.ac.cn (S.W.)

Abstract: Lately, deep learning technology has been extensively investigated for accelerating dynamic magnetic resonance (MR) imaging, with encouraging progresses achieved. However, without fully sampled reference data for training, the current approaches may have limited abilities in recovering fine details or structures. To address this challenge, this paper proposes a self-supervised collaborative learning framework (SelfCoLearn) for accurate dynamic MR image reconstruction from undersampled k-space data directly. The proposed SelfCoLearn is equipped with three important components, namely, dual-network collaborative learning, reunderampling data augmentation and a special-designed co-training loss. The framework is flexible and can be integrated into various model-based iterative un-rolled networks. The proposed method has been evaluated on an in vivo dataset and was compared to four state-of-the-art methods. The results show that the proposed method possesses strong capabilities in capturing essential and inherent representations for direct reconstructions from the undersampled k-space data and thus enables high-quality and fast dynamic MR imaging.

Keywords: dynamic MR imaging; self-supervised learning; collaborative learning; reunderampling data augmentation; co-training loss

1. Introduction

Deep learning-based dynamic magnetic resonance (MR) imaging has attracted substantial attention in recent years. It draws knowledge from big datasets via network training and then uses the trained network to reconstruct a dynamic image from the undersampled k-space data. Compared to the classical compressed sensing methods [1–7], deep learning-based methods have made encouraging performances and progresses.

Based on the reliance on the fully sampled dataset or not, existing methods for dynamic MR imaging can be roughly classified into two types [8–10]: fully-supervised methods and unsupervised methods. For the fully-supervised methods, data pairs are needed for the training of the neural networks between the corrupted/ undersampled data and the ground truth/fully sampled data [11–18]. In this category, different network structures and prior knowledge have been explored [19–26]. For example, Schlemper et al. proposed a cascade network architecture composed of an intermediate de-aliasing convolutional neural network (CNN) module and a data consistency layer [22]. Chen et al. applied bidirectional convolutional recurrent neural network (CRNN) with interleaved data consistency to accelerate MR imaging [23]. Chen et al. designed a parallel framework, including a time-frequency domain CRNN and an image domain CRNN to simultaneously exploit spatiotemporal correlations [24]. Wang et al. applied both k-space and spatial prior

knowledge to accelerate MR imaging [25]. Ke et al. exploited the low rank priors (SLR-Net) [26]. The aforementioned methods have made great progress in accelerating dynamic MRI reconstruction. However, one major challenge of the above methods is that, in many practical imaging scenarios, obtaining high-quality fully sampled dynamic MR data is infeasible due to various factors, such as the physiological motions of patients and imaging speed restriction. Therefore, the requirement for fully sampled reference data of network training limits the wide application of supervised learning methods.

To address this problem, researchers have developed unsupervised learning methods to train models without fully sampled reference data [27–30]. For example, Jin et al. extended the framework of deep image prior [31] to dynamic non-Cartesian MRI [29]. Recently, Yaman et al. proposed a classical self-supervised learning strategy (SSDU) for static MR imaging [32], which divides the acquired undersampled data into two parts, of which one is treated as input data, and another is used as the supervisory signals [33]. Subsequently, Acar et al. applied SSDU to reconstruct dynamic MR images [30]. The above-mentioned works have made great contributions to unsupervised dynamic MR image reconstruction. Nevertheless, since the undersampled data have incomplete inherent representation compared to the fully sampled data, these works still have room to improve in recovering fine details or structures.

To boost the performances for accelerating dynamic MR imaging without fully sampled reference data, this paper proposes a self-supervised collaborative learning framework named the SelfCoLearn. The SelfCoLearn is based on the assumption that the latent representation of network predictions is consistent under different reundersampling data augmentation from the same data. The SelfCoLearn performs collaborative training of a dual-network using reundersampling data augmentation to explore more sufficient prior knowledge compared to a single network. Specifically, from undersampled k-space data, the reundersampling data augmentation operations are implemented to obtain two reundersampling inputs for a dual-network. In addition, dual networks are trained collaboratively with a special-designed co-training loss in an end-to-end manner. By using this collaborative training strategy, the proposed framework can possess strong capabilities in capturing essential and inherent representations from the undersamled k-space data in a self-supervised learning manner. Moreover, the proposed framework is flexible and can be integrated with various model-based iterative un-rolled networks [34] for dynamic MR imaging. In summary, the main contributions of this work can be expressed as follows:

1. We present a self-supervised collaborative learning framework with reundersampling data augmentation for accelerating dynamic MR imaging. The proposed framework is flexible and can be integrated with various model-based iterative un-rolled networks;
2. A co-training loss, including both undersampled consistency loss term and a contrastive consistency loss term, is designed to guide the end-to-end framework to capture essential and inherent representations from undersamled k-space data;
3. Extensive experiments are conducted to evaluate the effectiveness of the proposed SelfCoLearn with different model-based iterative un-rolled networks, with more promising results obtained compared to self-supervised methods.

The remainder of this paper is organized as follows: Section 2 states the dynamic MR imaging problem and the proposed SelfCoLearn with different backbone networks. Section 3 summarizes the comparison experiments and results to demonstrate the effectiveness of SelfCoLearn. Section 4 presents discussions about the impact of different backbone networks and loss functions. Section 5 concludes the work.

2. Methodology

2.1. Dynamic MR Imaging Formulation

The problem of dynamic MR imaging aims to estimate dynamic MR image sequences $\mathbf{x} \in \mathbb{C}^N$ from undersampled measurements $\mathbf{y} \in \mathbb{C}^M (M \ll N)$ in k-space. $N = N_h N_W T$ is a

vector. N_h and N_W are the height and width of the frame, respectively. T represents the number of frames in each sequence. Thus, the imaging model is described as follows:

$$\mathrm{y} = \mathbf{Ax} + \mathrm{e} \tag{1}$$

where $\mathrm{e} \in \mathbb{C}^M$ is noise and $\mathbf{A} = \mathbf{PF}$ is an undersampled Fourier encoding operator, $\mathbf{F}$ is 2D Fourier transform to each frame in the image sequence and $\mathbf{P}$ is the undersampled mask for each frame. In general, the reconstruction problem is formulated as the following unconstrained optimization problem:

$$\mathrm{x}^* = \arg \min_{\mathrm{x}} \frac{1}{2} \|\mathbf{Ax} - \mathrm{y}\|_2^2 + \lambda \mathcal{R}(\mathbf{x}) \tag{2}$$

where $\mathcal{R}(\mathbf{x})$ represents a prior regularization item on $\mathbf{x}$, and λ is the weight of the regularization. $\frac{1}{2} \|\mathbf{Ax} - \mathrm{y}\|_2^2$ is the data fidelity item, which guarantees the reconstruction result to be consistent with the raw undersampled measurements.

For fully-supervised deep learning methods, it typically uses a CNN $f_{CNN}(\mathbf{y} \mid \theta)$ as a regularization term $\mathcal{R}(\mathbf{x})$, by learning the mapping between corrupted/undersampled data and their corresponding fully sampled data with parameters θ. Its mathematical description can be given as:

$$\theta^* = \arg \min_{\theta} \sum_{i=1}^{S} \mathcal{L}\left(f_{CNN}(\mathbf{y}_i \mid \theta), \mathbf{x}_i^{ref} \right) \tag{3}$$

where i is the index of the subject in the training dataset, and S is its total number. $\mathbf{x}_i^{ref}$ is the ground truth (fully sampled reference data) of the subject data i. $\mathcal{L}(\cdot)$ denotes the loss function between the predicted output and the ground truth, which typically adopts the l_1-norm or l_2-norm.

2.2. The Overall Framework

This work proposes a simple but effective self-supervised training framework for dynamic MR imaging, whose paradigm is shown in Figure 1. The proposed framework simultaneously trains two independent reconstruction networks, which have different inputs and different weight parameters. The backbone network can adopt various iterative un-rolled network, such as CRNN [23], k-t NEXT [21], and SLR-Net [26]. Based on the consistency between two networks' prediction results, the network provides complementary information for the to-be-reconstructed dynamic MR images in its peer partner. The two networks will finally realize consistent reconstruction in the training process. Specifically, given a raw undersampled k-space data sequence $\Omega = \left\{ \mathbf{y}_\Omega^t \right\}_{t=1}^{T}$, the original k-space data $\mathbf{y}_\Omega^t$ are reundersampled to construct a partial data points sequence $\left\{ \mathbf{y}_u^t \right\}_{t=1}^{T}$ as follows:

$$\mathbf{y}_u^t = P_u^t \left(\mathbf{y}_\Omega^t \right), t = 1, \dots, \mathrm{T}, u = \Theta, \Lambda \tag{4}$$

where t is the sequence index, u denotes the index of the two training sequences and P_u^t is the undersampled mask for frame t. To achieve full use of all data points in $\mathbf{y}_\Omega^t$ to learn representation, and ensure that each network can provide complementary information for the to-be-reconstructed dynamic MR images in its peer network, these training sequences are generated to adhere to the following data augmented principles: (1) The union of data points in two training sequences must be equal to the data $\mathbf{y}_\Omega^t$, i.e., $\mathbf{y}_\Omega^t = \mathbf{y}_\Theta^t \cup \mathbf{y}_\Lambda^t$. (2) The data points in two training sequences should be different, i.e., $\mathbf{y}_\Theta^t \neq \mathbf{y}_\Lambda^t$. (3) The training sequences should include most of the low frequency signals and part of the high frequency signals. Low frequency signals correspond to data points in the k-space center or close to it and high frequency signals to the outer parts of the k-space. Following these principles, the two training sequences contain different points in the high frequency region, and similar data points in the low frequency region. It should be noted that data reundersampling is

necessary only during training, whereas the reconstructed images can be inferred from the test data directly.
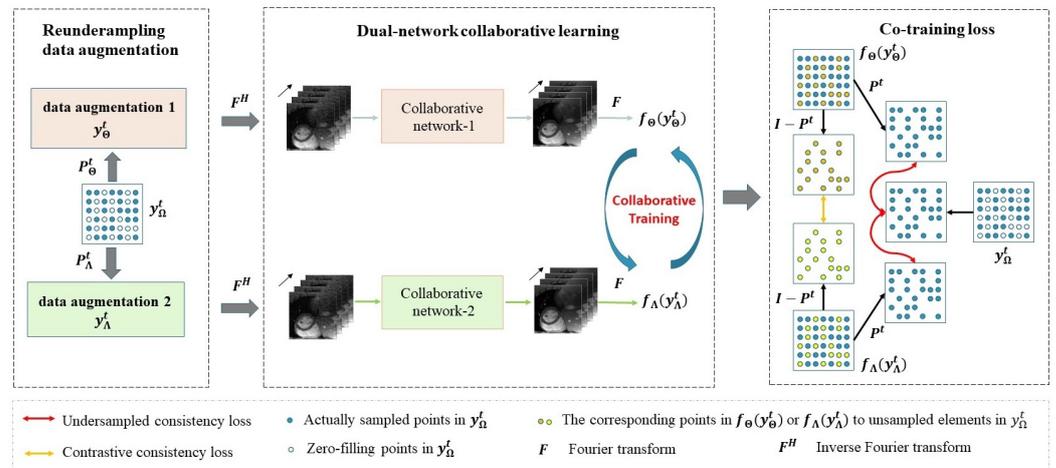
Figure 1. An overview of the proposed self-supervised collaborative training framework. A raw undersampled k-space data sequence $\mathbf{y}_\Omega^t$ is undersampled from the fully sampled data using an undersampled mask P^t retrospectively, and then two k-space data sequences $\mathbf{y}_\Theta^t$ and $\mathbf{y}_\Lambda^t$ are augmented from $\mathbf{y}_\Omega^t$. In the considered scenario, $\mathbf{y}_\Theta^t$ and $\mathbf{y}_\Lambda^t$ are reundersampled from $\mathbf{y}_\Omega^t$ using reundersampled mask $\mathbf{P}_\Theta^t$ and $\mathbf{P}_\Lambda^t$, respectively. Next, the two networks received inputs from zero-filling image sequences of $\mathbf{y}_\Theta^t$ and $\mathbf{y}_\Lambda^t$. The predicted image sequences of networks are transformed to the k-space data $f_\Theta(y_\Theta^t)$ and $f_\Lambda(y_\Lambda^t)$ by two-dimensional Fourier transform. Afterward, a co-training loss is calculated using $\mathbf{y}_\Omega^t$, $f_\Theta(y_\Theta^t)$ and $f_\Lambda(y_\Lambda^t)$. The backbone reconstruction network can flexibly adopt different iterative un-rolled network, such as CRNN, k-t NEXT and SLR-Net. Collaborative network-1 and collaborative network-2 have the same network structure but different weight parameters θ_Θ and θ_Λ respectively.

2.3. Network Architectures

2.3.1. Model-Driven Deep Learning with Image-Domain Regularization

In these settings, the common practice is to decouple Equation (2) into a regularization term and a data fidelity term via utilizing the variable splitting technique [22,23]. By introducing an auxiliary variable $\mathbf{z} = \mathbf{x}$, Equation (2) can be re-formulated as a penalty function [23], which can be expressed as follows:

$$\arg \min_{\mathbf{x},\mathbf{z}} \lambda \mathcal{R}(\mathbf{z}) + \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \mu\|\mathbf{x} - \mathbf{z}\|_2^2 \tag{5}$$

where μ denotes a penalty parameter. Equation (5) can then be solved iteratively via alternating minimization over $\mathbf{z}$ and $\mathbf{x}$:

$$\mathbf{z}^n = \arg \min_{\mathbf{z}} \lambda \mathcal{R}(\mathbf{z}) + \mu\|\mathbf{x}^{n-1} - \mathbf{z}\|_2^2 \tag{6}$$

$$\mathbf{x}^n = \arg \min_{\mathbf{x}} \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \mu\|\mathbf{x} - \mathbf{z}^n\|_2^2 \tag{7}$$

where $n \in \{1, 2, \ldots, N\}$ is the nth iteration, $\mathbf{x}^0$ is the zero-filling image transformed from original undersampled measurement, $\mathbf{z}^n$ denotes the intermediate reconstruction sequence, and $\mathbf{x}^n$ denotes the final reconstruction sequence at each iteration. In Equation (7), the operation on the intermediate reconstruction sequence $\mathbf{z}^n$ is a data consistency step [22]. The iterative optimization process in Equations (6) and (7) is unrolled into a neural network.

The CRNN [23] is a typical model-driven deep learning method with image-domain regularization for dynamic MR imaging [35]. A single iteration of the CRNN can be expressed as follows:

$$\mathbf{x}_{rnn}^{(n)} = \mathbf{x}_{rec}^{(n-1)} + \text{CRNN}\left(\mathbf{x}_{rec}^{(n-1)}\right) \tag{8}$$

$$\mathbf{x}_{rec}^{(n)} = \text{DC}\left(\mathbf{x}_{rnn}^{(n)}; y, \lambda\right) \tag{9}$$

where $\mathbf{x}_{rnn}^{(n)}$ is the intermediate reconstruction sequence analogous to $\mathbf{z}^n$ in Equation (6), and $\mathbf{x}_{rec}^{(n)}$ denotes the final predicted result at each iteration analogous to $\mathbf{x}^n$ in Equation (7). The regularization subproblem in Equation (6) is solved by using a convolutional recurrent neural network. The data consistency subproblem in Equation (7) is treated as a data consistency network layer, which uses the original sampled k-space data points to replace the corresponding data points in the reconstructed k-space data [22]. More details of CRNN layers can be found in Ref. [23].

2.3.2. Model-Driven Deep Learning with Complementary Regularization

The complementary regularization is also an effective method for dynamic MR imaging. The k-t NEXT [21] is a typical model-driven deep learning method with complementary regularization [35], which exploits prior information in both combined spatial and temporal Fourier (x-f) domain and spatiotemporal image (x-t) domain. A single iteration of the k-t NEXT can be expressed as the following process:

$$\rho^{(n)} = \text{DC}(\mathbf{y}_{base}) + \text{xf} - \text{CNN}\left(\mathbf{y}_{rec}^{(n-1)} - \mathbf{y}_{base}\right), \tag{10}$$

$$\mathbf{x}_{rec}^{(n)} = \text{CRNN}\left(\mathbf{F}_f^H \rho^{(n)}; y_0\right), \quad \mathbf{y}_{rec}^{(n)} = \mathbf{F}_{xy}\mathbf{x}_{rec}^{(n)} \tag{11}$$

where $\rho^{(n)}$ denotes the intermediate reconstruction results in the x-f domain from xf-CNN at nth iteration, $\mathbf{x}_{rec}^{(n)}$ denotes the reconstruction image sequence in the x-t domain at nth iteration, $\mathbf{y}_{base}$ is the corresponding baseline signal, and $\mathbf{F}_{xy}$ and $\mathbf{F}_f^H$ denote, respectively, the Fourier transform in x-t domain and the inverse Fourier transform in x-f domain.

2.3.3. Model-Driven Deep Learning with Low-Rank Regularization

Another widely-used prior regularization is low-rank based dynamic MR imaging, which applies low-rank priors as regularized terms. The SLR-Net [26] is a typical example of a model-driven deep learning method with low-rank regularization. In the SLR-Net, by introducing an auxiliary variable $\mathbf{M}$, Equation (2) can be decoupled as the fidelity term, sparse regularization term, and the low rank regularization term:

$$\arg\min_{\text{x,M}} \frac{1}{2}\|\mathbf{A}\mathbf{x} - y\|_2^2 + \lambda_1\|D\mathbf{x}\|_1 + \lambda_2\|M\|_* \tag{12}$$

where D is a sparse transform in a certain sparse domain. $M = R\mathbf{x}$ is a matrix (with size $(N_{\mathbf{h}} \times N_{\mathbf{w}}, T)$), in which each column corresponds to one frame in dynamic MR image sequence. R is a reshaping operator. $\|M\|_*$ is the nuclear norm. Previous works have proven that nuclear norm minimization is effective in low-rank matrix recovery [36]. More details of the iterative process in SLR-Net can be found in Ref. [26].

2.4. The Proposed Co-Training Loss

In this study, a co-training loss is defined to promote accurate dynamic MR image reconstruction in a self-supervised manner. The main idea of the co-training loss is to enforce the consistency not only between the reconstruction results and the original undersampled k-space data, but also between two network predictions. Compared with existing self-supervised methods with single network, the consistency between two network predictions is an additional regularization, which guides the dual-network to narrow the divergence

and learn more correct information. Specifically, the co-training loss in SelfCoLearn, including an undersampled consistency loss term and a contrastive consistency loss term, is calculated to optimize the proposed framework.

Let $f_{SelfCoLearn}(\mathbf{y}_\Omega^t)$ denote SelfCoLearn, $\mathbf{y}_\Omega^t$ is the original undersampled k-space data. During training, two training sequences $\mathbf{y}_\Theta^t$ and $\mathbf{y}_\Lambda^t$ are generated from $\mathbf{y}_\Omega^t$ following the data augmented principles in Section 2.2 as follows:

$$\mathbf{y}_\Theta^t = \mathbf{P}_\Theta^t \mathbf{y}_\Omega^t, \quad \mathbf{y}_\Lambda^t = \mathbf{P}_\Lambda^t \mathbf{y}_\Omega^t, \tag{13}$$

where $\mathbf{P}_\Theta^t$ and $\mathbf{P}_\Lambda^t$ are the reundersampled mask for $\mathbf{y}_\Omega^t$. The undersampled consistency loss is mainly referred to the actually sampled k-space points in $\mathbf{y}_\Omega^t$, which ensures that the corresponding sampled points in network prediction are consistent with the actually sampled k-space points in $\mathbf{y}_\Omega^t$. The actually sampled points in these two network predictions are denoted as $y_{\Theta\to\Omega}^t$ and $y_{\Lambda\to\Omega}^t$, respectively. $y_{\Theta\to\Omega}^t$ and $y_{\Lambda\to\Omega}^t$ in these two network predictions can be written as:

$$y_{\Theta\to\Omega}^t = \mathbf{P}^t f(\mathbf{y}_\Theta^t), \quad y_{\Lambda\to\Omega}^t = \mathbf{P}^t f(\mathbf{y}_\Lambda^t), \tag{14}$$

where k-space data $f(\mathbf{y}_\Theta^t)$ and $f(\mathbf{y}_\Lambda^t)$ are transformed from the predicted image sequences of two networks, respectively. $\mathbf{P}^t$ is the undersampled mask, which is applied to generate the raw undersampled k-space data $\mathbf{y}_\Omega^t$ from the fully sampled data.

The Undersampled Consistency loss term is used to calculate the MSE between the actually sampled k-space points in $\mathbf{y}_\Omega^t$ and those predicted by the network as follows:

$$\mathcal{L}_{UC} = \left\| \mathbf{y}_{\Theta\to\Omega}^t - \mathbf{y}_\Omega^t \right\|_2^2 + \left\| \mathbf{y}_{\Lambda\to\Omega}^t - \mathbf{y}_\Omega^t \right\|_2^2. \tag{15}$$

In the ideal case, when different reundersampled k-space data from the same data are set as inputs of the two networks, the networks' predictions should approximate the fully-sampled reference data after network optimization. However, when fully sampled reference data are unavailable, these two networks can be trained only using the undersampled consistency loss, and they will be likely to generate different prediction results, which will lead to different reconstruction performances. As mentioned above, a contrastive consistency loss is defined to compute the MSE between two network predictions obtained by using different reundersampling inputs generated from the same data. Specially, the proposed contrastive consistency loss term mainly refers to the points in network predictions corresponding to unsampled k-space points in $\mathbf{y}_\Omega^t$. Points $\bar{\mathbf{y}}_{\Theta\to\Omega}^t$ and $\bar{\mathbf{y}}_{\Lambda\to\Omega}^t$ in two network predictions $f(\mathbf{y}_\Theta^t)$ and $f(\mathbf{y}_\Lambda^t)$ can be expressed as follows:

$$\bar{\mathbf{y}}_{\Theta\to\Omega}^t = (\mathbf{I} - \mathbf{P}^t) f(\mathbf{y}_\Theta^t), \quad \bar{\mathbf{y}}_{\Lambda\to\Omega}^t = (\mathbf{I} - \mathbf{P}^t) f(\mathbf{y}_\Lambda^t), \tag{16}$$

therefore, the Contrastive Consistency loss term is formulated as:

$$\mathcal{L}_{CC} = \left\| \bar{\mathbf{y}}_{\Theta\to\Omega}^t - \bar{\mathbf{y}}_{\Lambda\to\Omega}^t \right\|_2^2. \tag{17}$$

combining the two loss terms, the final co-training loss function can be defined as follows:

$$\mathcal{L}_{co} = \mathcal{L}_{UC} + \gamma \mathcal{L}_{CC}, \tag{18}$$

where γ is used to balance the weight parameter of the undersampled consistency loss and the contrastive consistency loss. During the testing phase, the undersampled data is used as input of the collaborative network-1 or collaborative network-2 to obtain the final reconstruction result.

3. Experimental Results

Extensive experiments have been performed to evaluate the effectiveness of Self-CoLearn. SelfCoLearn is compared with fully-supervised and self-supervised learning

methods at different acceleration factors. Besides, SelfCoLearn with different backbone networks for dynamic MR imaging have been experimented. Then, the results of the ablation studies are reported to investigate the impacts of the undersampled consistency loss term and contrastive consistency loss term. Finally, reconstruction results with a different co-training loss calculated in different domains are reported to further evaluate the proposed SelfCoLearn.

3.1. Experimental Setup

3.1.1. Dataset

The dataset includes fully sampled 2D+t complex-valued short-axis cardiac cine MR data collected on a 3T Siemens Magnetom Trio scanner from 101 healthy volunteers. T1-weighted FLASH sequence is utilized. Each scan includes single-slice FLASH acquisition from the volunteer with retrospectively electrocardiogram ECG-gating. Each volunteer needed to breath-hold for 15–20 s on each slice. The parameters of data acquisition include 24 receiving coils, FOV of 330 mm × 330 mm, acquisition matrix of 192 × 192, slice thickness of 6 mm, repetition time of 50 ms, and echo time of 3 ms. Each scan with a single slice covers the entire cardiac dynamic process with 25 temporal frames. This retrospective study was approved by local ethics committee and the informed consent was obtained from all of the involved volunteers. In the experiments, the set of scanned multi-coil MR data for each frame is transformed to a single-channel MRI by the adaptive reconstruction technique [37]. The corresponding k-space data to the single-channel MRI can be viewed as a fully sampled single-coil data. To enlarge the training dataset, we implement data augmentation strategies by shearing the single-channel complex-valued image along the dimensions of x, y, and t. After data augmentation, the dataset includes 6214 complex-valued data sequences of size 128 × 128 × 14. A total of 5950 cardiac MR data sequences were selected as the training dataset, 50 cardiac sequences were used as the validation dataset, and the remaining sequences were used for testing.

3.1.2. Reundersampling K-Space Data Augmentation

In the proposed method, the fully sampled data are only used to generate the original undersampled k-space data $\mathbf{y}_\Omega^t$ with a Cartesian retrospective undersampled mask $\mathbf{P}^t$. Following the principles of training data augmentation in Section 2.2, $\mathbf{y}_\Omega^t$ is augmented to two training sequences $\mathbf{y}_\Theta^t$ and $\mathbf{y}_\Lambda^t$ with two Cartesian reundersampled masks $\mathbf{P}_\Theta^t$ and $\mathbf{P}_\Lambda^t$. $\mathbf{P}_\Theta^t$ with 2-fold acceleration is used for collaborative network-1, and $\mathbf{P}_\Lambda^t$, which combines the complementary set of $\mathbf{P}_\Theta^t$ with some low-frequency data points of $\mathbf{P}^t$, is used for collaborative network-2.

3.1.3. Evaluation Metrics

Reconstruction performances are evaluated by calculating mean-squared-error (MSE), peak-signal-to-noise ratio (PSNR), and structural similarity index (SSIM) [38] on magnitude images. The evaluation metrics are measured between the reconstruction image sequence Rec with the reference image sequence Ref as follows:

$$\text{MSE} = \|\text{Ref} - \text{Rec}\|_2^2 \tag{19}$$

$$\text{PSNR} = 20 \log_{10} \frac{MAX_{Ref}}{\sqrt{MSE}} \tag{20}$$

$$\text{SSIM} = \frac{\left(2\mu_{Ref}\mu_{Rec} + c_1\right)\left(2\sigma_{Ref,Rec} + c_2\right)}{\left(\mu_{Ref}^2 + \mu_{Rec}^2 + c_1\right)\left(\sigma_{Ref}^2 + \sigma_{Rec}^2 + c_2\right)} \tag{21}$$

where MAX_{Ref} is the maximum possible value in the image. μ_{Ref} and μ_{Rec} are the averaged intensity values of the corresponding images. σ_{Ref} and σ_{Rec} are the variances. c_1 and c_2

are adjustable constants. $\sigma_{Ref,Rec}$ is the covariance. (details of SSIM index can be found in Ref. [38]).

3.1.4. Model Configuration and Implementation Details

The proposed framework is flexible and can be integrated with various iterative unrolled networks, such as CRNN, k-t NEXT and SLR-Net. Most of our experiments adopt CRNN as the backbone network. In detail, the network is composed of a bidirectional CRNN layer, three CRNN layers, a 2D CNN layer, a residual connection and a DC layer. For the bidirectional CRNN and CRNN layer, the convolution filter is set as 64 and the kernel size is set as 3. The 2D CNN layer has kernel size $k = 3$ and convolution filter $N_f = 2$. We use $stride = 1$ and the padding is set to half of the filter size (rounded down). The DC layer is followed by the 2D CNN layer, which forces the actually sampled points in the predicted k-space data to be consistent with that in the input data.

For model training, the number of iteration steps is set to $N = 5$. The batch size is set to 1. All training data and test data are normalized to the range of [0, 1]. The SelfCoLearn framework with CRNN and k-t NEXT is implemented in PyTorch 1.8.1, and that with SLR-Net is implemented in Tensorflow 2.2.0. The experiments are performed on an Nvidia Titan Xp GPU, with 12GB memory. SelfCoLearn is trained by Adam optimizer [39] with parameters $\beta_1 = 0.5$ and $\beta_2 = 0.999$. The learning rate is set to 10^{-4}. The weight parameter γ in co-training loss is set to 0.01. It takes 52 h to train SelfCoLearn with CRNN and each cardiac MR data sequence takes roughly 0.5 s to get the reconstructed result.

3.2. Comparisons to State-of-the-Art Unsupervised Methods

To evaluate the proposed SelfCoLearn, we compared it with two self-supervised methods, SS-DCCNN and SS-CRNN, at different acceleration factors. It is worth noting that the state-of-the-art self-supervised method SSDU [32] was developed for static MR imaging. Ref. [30] adopted a similar self-supervised training manner as SSDU for dynamic MR imaging. They evaluated several backbone architectures for dynamic MR imaging including DCCNN and CRNN, whereas SSDU adopted ResNet as the backbone network. We choose two self-supervised learning methods SS-DCCNN and SS-CRNN [30] for comparison. In this experiment, the proposed SelfCoLearn selects the CRNN as the backbone network.

Figure 2 plots the reconstruction results of different self-supervised methods at 4-fold acceleration, 8-fold acceleration, and 12-fold acceleration, respectively. The first row and fourth row show the ground truth (fully sampled image) and the reconstruction images of the respective methods in the diastolic and systolic at different accelerations, respectively (display range [0, 1]). The second row and fifth row show their corresponding enlarged images in the heart regions. The third row and sixth row plot the error images of the corresponding methods (display range [0, 0.2]). The y-t images at the 40th slice along the dimensions of y and t are shown in the seventh row. The corresponding error images of y-t images are plotted in the last row. From the visualization results, the proposed SelfCoLearn generates better reconstruction results than the two self-supervised methods, SS-DCCNN and SS-CRNN, at all acceleration factors. The reconstruction images of SelfCoLearn show finer structural details and more precise heart borders with fewer artifacts.

The quantitative results of these self-supervised methods are listed in Table 1. Similar conclusions can be obtained, showing that the SelfCoLearn achieves better quantitative performance than these self-supervised learning methods. Therefore, our collaborative learning strategy can effectively capture essential and inherent representations from undersampled k-space data directly.
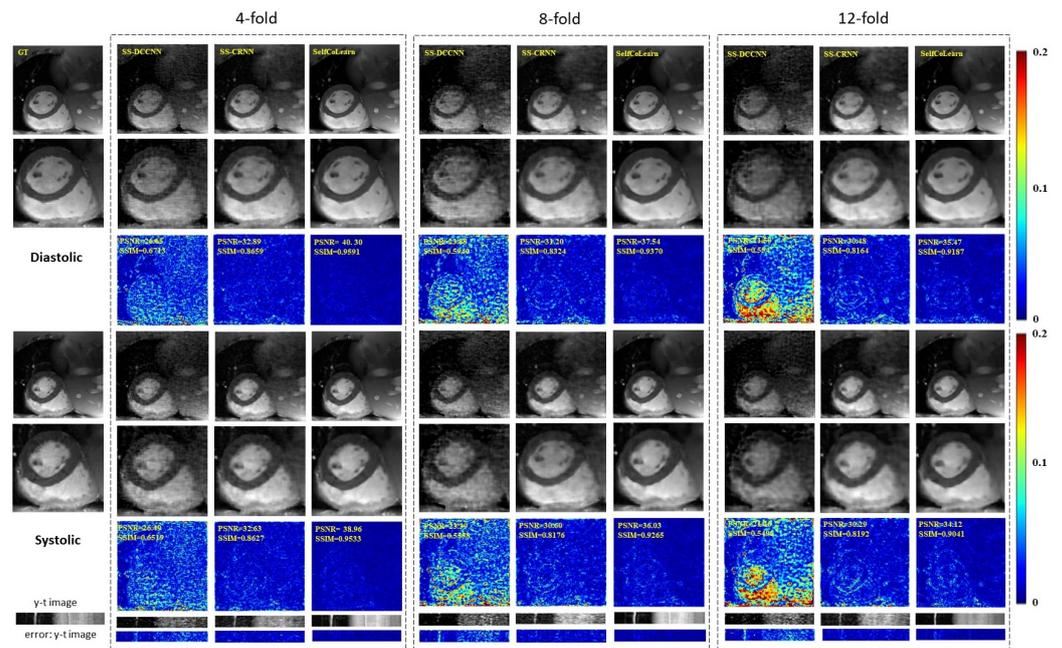
Figure 2. Reconstruction results of different self-supervised methods (SS-DCCNN, SS-CRNN, and SelfCoLearn) at 4-fold acceleration, 8-fold acceleration, and 12-fold acceleration. The first row and fourth row show the ground truth (fully sampled image) and the reconstruction images of the respective methods in the diastolic (the 10th frame of image sequence) and systolic (the 5th frame of image sequence), respectively. The second row and fifth row show their corresponding enlarged images in the heart regions. The third row and sixth row plot the error images of corresponding methods. The last two rows show y-t images (the 40th slice along the dimensions of y and t) and the corresponding error images.

Table 1. Quantitative reconstruction results of different self-supervised methods (SS-DCCNN, SS-CRNN, and SelfCoLearn) at 4-fold, 8-fold, and 12-fold acceleration factors (mean ± std).

AF	Methods	Training Pattern	PSNR (dB)	SSIM	MSE ($\times 10^{-4}$)
	SS-DCCNN	Self-supervised	25.81 ± 2.86	0.6409 ± 0.0739	32.81 ± 24.85
4-fold	SS-CRNN	Self-supervised	32.49 ± 1.79	0.8383 ± 0.0387	6.14 ± 2.62
	SelfCoLearn	Self-supervised	40.34 ± 2.69	0.9536 ± 0.0239	1.11 ± 0.72
	SS-DCCNN	Self-supervised	22.56 ± 2.71	0.5615 ± 0.0732	67.87 ± 49.27
8-fold	SS-CRNN	Self-supervised	30.81 ± 1.77	0.8015 ± 0.0427	9.02 ± 3.75
	SelfCoLearn	Self-supervised	37.27 ± 2.40	0.9243 ± 0.0338	2.17 ± 1.22
	SS-DCCNN	Self-supervised	22.17 ± 2.76	0.5270 ± 0.0702	74.89 ± 54.96
12-fold	SS-CRNN	Self-supervised	30.14 ± 1.78	0.7943 ± 0.0444	10.54 ± 4.40
	SelfCoLearn	Self-supervised	35.19 ± 2.24	0.8985 ± 0.0399	3.44 ± 1.78

Figure 3 shows the box plots displaying the median and interquartile range (25th–75th percentile) of the reconstruction results of different self-supervised methods on the test cardiac cine data at 4-fold acceleration, 8-fold acceleration, and 12-fold acceleration, respectively. The results in Figure 3 show that, for all dynamic cine sequences, the SelfCoLearn outperforms the two self-supervised learning methods (SS-DCCNN and SS-CRNN) at all three acceleration factors.
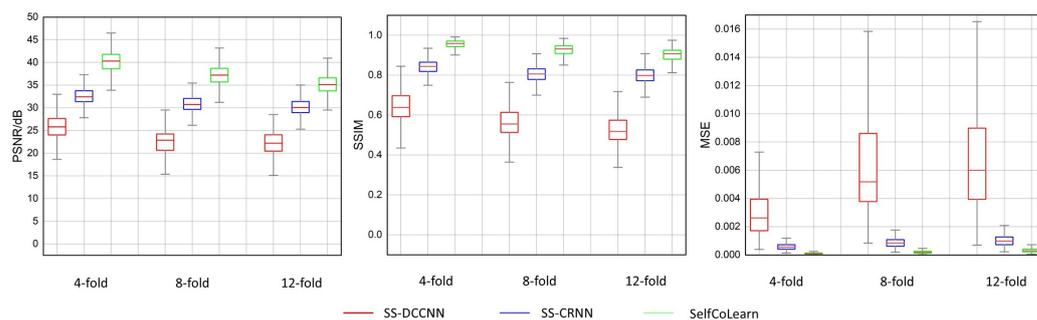
Figure 3. Box plots of different methods (SS-DCCNN, SS-CRNN, and SelfCoLearn) at 4-fold, 8-fold, and 12-fold accelerations are presented, which show the median and interquartile range of the PSNR, SSIM, and MSE on the cardiac cine test dataset.

3.3. Comparisons to State-of-the-Art Supervised Methods

We further compare our SelfCoLearn with different supervised methods, including supervised U-Net and supervised CRNN [23], at different acceleration factors. Figure 4 plots the reconstruction images of different methods at 4-fold acceleration, 8-fold acceleration, and 12-fold acceleration, respectively. The error images of SelfCoLearn indicate minor reconstruction errors than those of supervised U-Net.
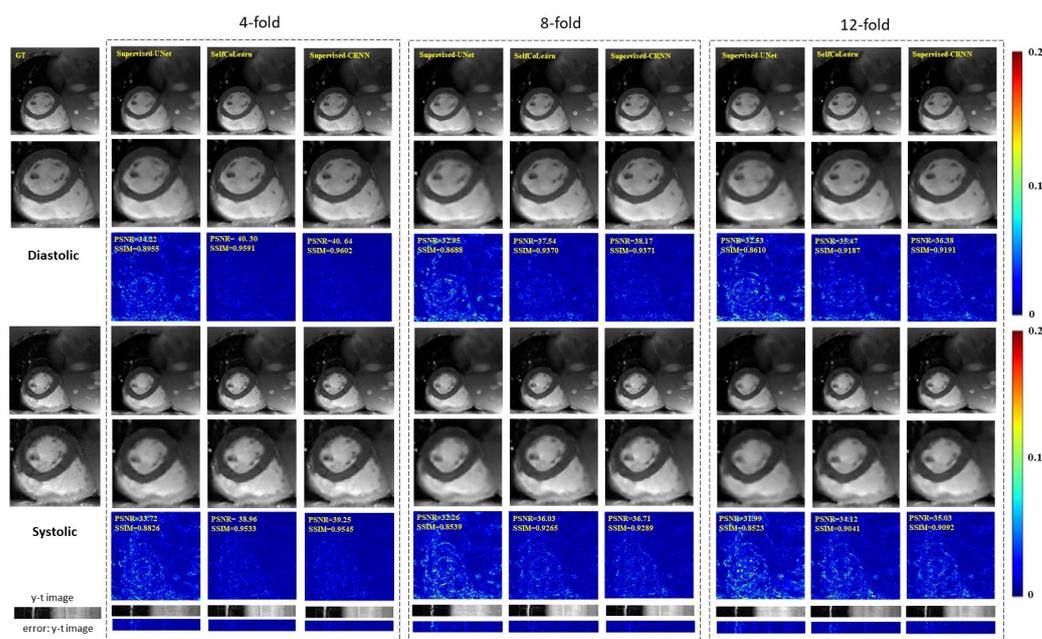
Figure 4. Reconstruction results of different methods (Supervised U-Net, SelfCoLearn, and Supervised CRNN) at 4-fold acceleration, 8-fold acceleration, and 12-fold acceleration. The first row and fourth row show the ground truth (fully sampled image) and the reconstruction images of respective methods in the diastolic (the 10th frame of the image sequence) and systolic (the 5th frame of the image sequence), respectively. The second row and fifth row show their corresponding enlarged images in the heart regions. The third row and sixth row plot the error images of the corresponding methods. The last two rows show y-t images (the 40th slice along the dimensions of y and t) and the corresponding error images.

In addition, the reconstruction results generated by SelfCoLearn are close to those of supervised CRNN at low acceleration factors. From the quantitative results in Table 2, the PSNR and SSIM of SelfCoLearn present 1.3% and 0.17% lower than those of supervised CRNN at 4-fold acceleration factors, respectively. At higher acceleration factors, such as 12-fold acceleration, the reconstructed images of SelfCoLearn become slightly

blurred. Nevertheless, most of the structural details in the heart regions are still successfully restored by SelfCoLearn. The PSNR and SSIM of SelfCoLearn present 3.2% and 0.69% lower than those of supervised CRNN at 12-fold acceleration factors, respectively. Therefore, SelfCoLearn can achieve comparable reconstruction performance with baseline fully-supervised methods via self-supervised dual-network collaborative learning.

Table 2. Quantitative reconstruction results of different methods (Supervised U-Net, Supervised CRNN and SelfCoLearn) at 4-fold, 8-fold, and 12-fold acceleration factors (mean ± std).

AF	Methods	Training Pattern	PSNR (dB)	SSIM	MSE ($\times 10^{-4}$)
4-fold	U-Net	Supervised	33.77 ± 1.96	0.8698 ± 0.0391	4.66 ± 2.22
	SelfCoLearn	Self-supervised	40.34 ± 2.69	0.9536 ± 0.0239	1.11 ± 0.72
	CRNN	Supervised	40.89 ± 2.90	0.9553 ± 0.0237	1.01 ± 0.68
8-fold	U-Net	Supervised	32.63 ± 1.97	0.8329 ± 0.0456	6.06 ± 2.88
	SelfCoLearn	Self-supervised	37.27 ± 2.40	0.9243 ± 0.0338	2.17 ± 1.22
	CRNN	Supervised	38.09 ± 2.52	0.9269 ± 0.0342	1.83 ± 1.07
12-fold	U-Net	Supervised	31.96 ± 1.88	0.8315 ± 0.0478	6.99 ± 3.03
	SelfCoLearn	Self-supervised	35.19 ± 2.24	0.8985 ± 0.0399	3.44 ± 1.78
	CRNN	Supervised	36.32 ± 2.29	0.9048 ± 0.0392	2.67 ± 1.42

4. Discussion

4.1. Network Backbone Architectures

In this section, we explore the reconstruction results of the proposed self-supervised learning strategy with different backbone networks for dynamic MR imaging. The experiments are conducted using SLR-Net [26], k-t NEXT [21], and CRNN [23] at 8-fold acceleration. The reconstruction results with different backbone networks are exhibited in Figure 5 and Table 3. Compared with SS-CRNN [11], the proposed SelfCoLearn can achieve better results regardless of the utilized backbone network. Among the three different backbone networks, SLR-Net generates worse results than k-t NEXT and CRNN. The reason for this phenomenon may be that SLR-Net needs to learn a singular value threshold, and the absence of the fully sampled reference data causes the learned singular value threshold to be suboptimal. However, the proposed self-supervised learning strategy with SLR-Net still obtains acceptable reconstruction results. The qualitative results in Figure 5 clearly show that SelfCoLearn can better restore the structural details and achieve clearer reconstructed MR images (especially in the heart regions around the red and yellow arrows) than SS-CRNN. The quantitative results also indicate more accurate reconstructions achieved by the proposed SelfCoLearn. These results indicate that our proposed self-supervised learning framework is flexible, and it can achieve promising reconstruction results with various iterative un-rolled networks for dynamic MR imaging.

Table 3. Quantitative results of SS-CRNN and SelfCoLearn with different backbone networks at 8-fold acceleration (mean ± std).

Methods	Training Pattern	PSNR (dB)	SSIM	MSE ($\times 10^{-4}$)
SS-CRNN	Self-supervised	30.81 ± 1.77	0.8015 ± 0.0427	9.02 ± 3.75
SelfCoLearn with SLR-Net	Self-supervised	33.58 ± 2.24	0.9001 ± 0.0369	5.57 ± 10.48
SelfCoLearn with k-t Next	Self-supervised	36.95 ± 2.39	0.9226 ± 0.0343	2.34 ± 1.32
SelfCoLearn with CRNN	Self-supervised	37.27 ± 2.40	0.9243 ± 0.0338	2.17 ± 1.22
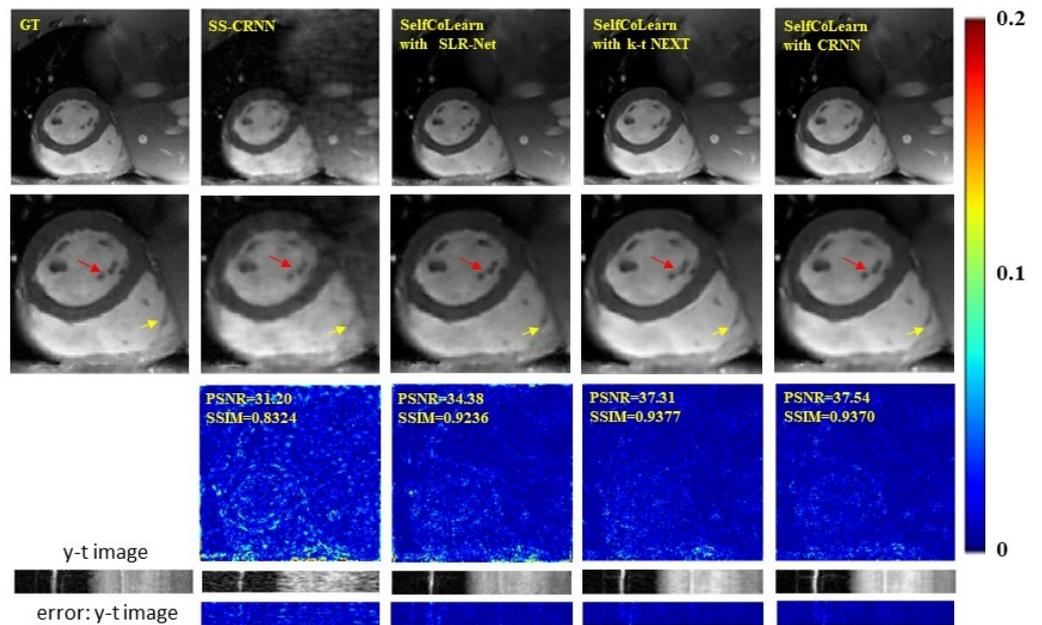
Figure 5. Reconstruction results of SS-CRNN and the proposed SelfCoLearn with SLR-Net, k-t NEXT, and CRNN backbone networks at 8-fold acceleration. The first row shows ground truth (fully sampled image), the reconstruction images of SS-CRNN and the proposed self-supervised learning strategy with SLR-Net, k-t NEXT, and CRNN (10th frame). The second row shows their enlarged images in the heart regions. The third row plots the error images of these two methods. The last two rows show the y-t images (the 40th slice along the dimensions of y and t) and the corresponding error images.

4.2. Co-Training Loss Function

In this section, we investigate the utility of the designed co-training loss function. The backbone network in these experiments adopts CRNN. Different training strategies at 8-fold acceleration are utilized. Strategy B-I: a single reconstruction network is trained in self-supervised manner. Only the loss function between the output $f\left(\mathbf{y}_\Theta^t\right)$ of network and $\mathbf{y}_\Lambda^t$ is used to train the network. Strategy B-II: a strategy similar to B-I but the loss function here is calculated between the output $f\left(\mathbf{y}_\Theta^t\right)$ of the network and the original undersampled k-space data $\mathbf{y}_\Omega^t$. SelfCoLearn: two networks are trained collaboratively with $\mathcal{L}_{UC}$ and $\mathcal{L}_{CC}$, and the two collaborative networks adopt the same backbone network as that in strategy B-I. Reconstruction images of methods utilizing the different training strategies are plotted in Figure 6. Quantitative results are listed in Table 4. From both qualitative and quantitative results, we can observe that SelfCoLearn (training two networks collaboratively with both loss terms) achieves the best performance (especially in the heart regions around the red and yellow arrows). In particular, the contrastive consistency loss term results in a large reconstruction performance improvement. For example, PSNR is improved from 31.04 dB (Strategy B-II) to 37.27 dB (SelfCoLearn).

Table 4. Quantitative results of reconstruction models utilizing different training strategies at 8-fold acceleration (mean ± std).

Methods	Single-Net	Parallel-Net	L_{UC}	L_{CC}	PSNR (dB)	SSIM	MSE ($\times 10^{-4}$)
Strategy B-I	√	×	×	×	30.81 ± 1.77	0.8015 ± 0.0427	9.02 ± 3.75
Strategy B-II	√	×	√	×	31.04 ± 1.74	0.8102 ± 0.0411	8.53 ± 3.50
SelfCoLearn	×	√	√	√	37.27 ± 2.40	0.9243 ± 0.0338	2.17 ± 1.22
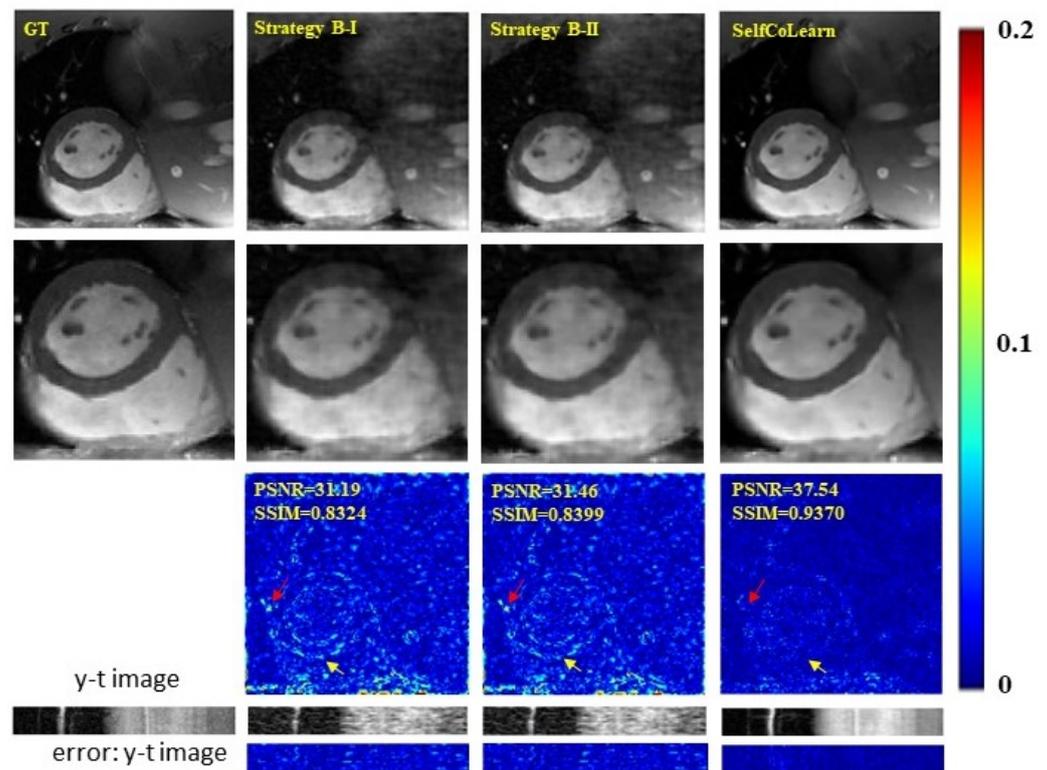
Figure 6. Ablation studies utilizing different training strategies at 8-fold acceleration. The first row shows the ground truth (fully sampled image), and the reconstruction images of strategy B-I, strategy B-II, and proposed SelfCoLearn (10th frame). The second row shows their enlarged images in the heart regions. The third row plots the error images of respective methods. The last two rows show y-t images (the 40th slice along the dimensions of y and t) and the corresponding error images.

4.3. Loss Functions

In this section, we inspect the effects of loss functions. The backbone network in these experiments adopts CRNN. Reconstruction results at 8-fold acceleration are given in Figure 7 and Table 5. Three strategies utilizing different loss function settings are investigated. In Strategy C-I, two networks are trained collaboratively with $\mathcal{L}_{UC}$ and $\mathcal{L}_{CC}$, in which $\mathcal{L}_{UC}$ is calculated in the x-t domain, and $\mathcal{L}_{CC}$ is calculated in the k-space domain. In Strategy C-II, both $\mathcal{L}_{UC}$ and $\mathcal{L}_{CC}$ are calculated in the x-t domain. In Strategy C-III, both $\mathcal{L}_{UC}$ and $\mathcal{L}_{CC}$ are calculated in the k-space domain. From both qualitative and quantitative results, we can observe that the influence of utilizing different loss function settings on the reconstruction performance is insignificant. All the other experiments in this work adopt the setting of strategy C-III.

Table 5. Quantitative results of methods utilizing different loss function strategies at 8-fold acceleration (mean ± std).

Methods	L_{UC}	L_{CC}	PSNR (dB)	SSIM	MSE ($\times 10^{-4}$)
Strategy C-I	x-t domain	k-space	37.00 ± 2.35	0.9230 ± 0.0344	2.30 ± 1.29
Strategy C-II	x-t domain	x-t domain	37.20 ± 2.37	0.9235 ± 0.0343	2.20 ± 1.22
Strategy C-III	k-space	k-space	37.27 ± 2.40	0.9243 ± 0.0338	2.17 ± 1.22
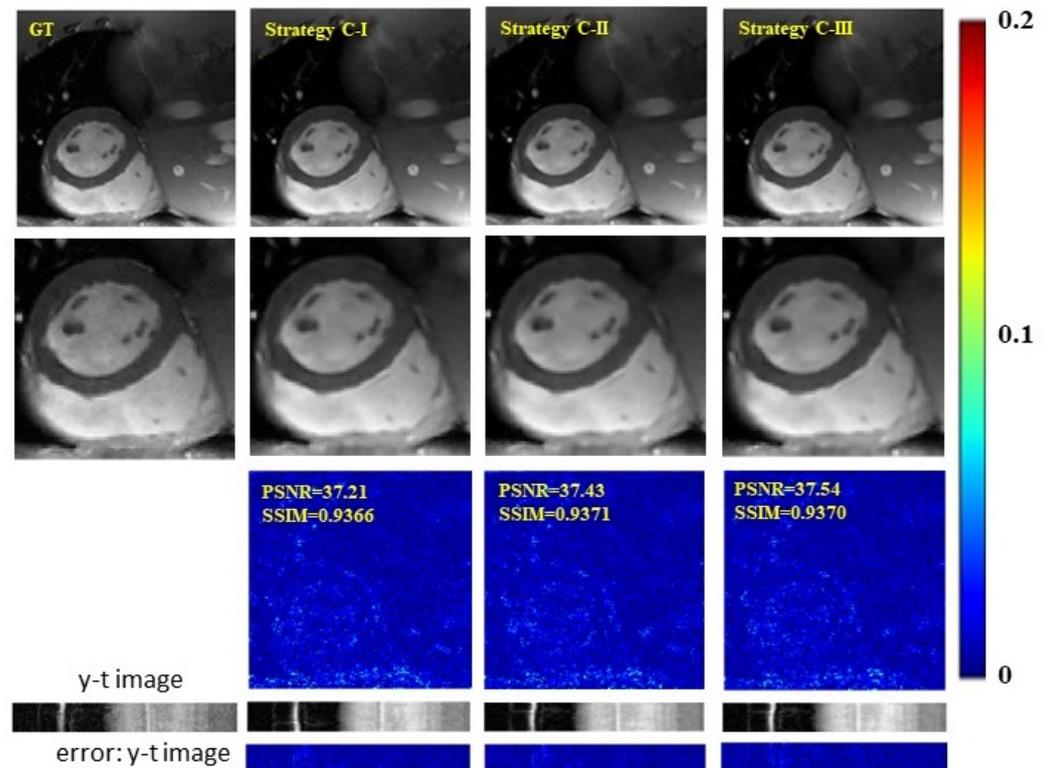
Figure 7. Effects of loss functions calculated in different domains on the reconstruction results at 8-fold acceleration. The first row shows ground truth (fully sampled image), the reconstruction results of models utilizing Strategy C-I, C-II and C-III (10th frame). The second row shows their enlarged images in the heart regions. The third row plots their error images of respective strategies. The last two rows show y-t images (the 40th slice along the dimensions of y and t) and the corresponding error images.

5. Conclusions

In our work, we propose a self-supervised collaborative training framework to boost the image reconstruction performance for accelerating dynamic MR imaging. Specifically, two independent reconstruction networks are trained collaboratively with different inputs, which are augmented from the same k-space data. To guide the dual-network in capturing the detailed structural features and spatiotemporal correlations in dynamic image sequences, a co-training loss function is designed to promote the consistency between network predictions to provide complementary information for the to-be-reconstructed dynamic MR images. The proposed framework is flexible and can be integrated with various iterative un-rolled networks. In addition, the proposed method has been comprehensively evaluated on a cardiac cine dataset. The quantitative and qualitative results indicate that SelfCoLearn possesses strong capabilities in capturing essential and inherent representations directly from the undersampled k-space data and thus enable high-quality and fast dynamic MR imaging.

Author Contributions: Methodology, S.W. and J.Z.; software, J.Z.; validation, J.Z. and R.W.; investigation, J.Z. and R.W.; data curation, J.Z. and R.W.; writing—original draft preparation, J.Z. and C.L.; writing—review and editing, J.Z., C.L., S.W., T.P. and S.J.; supervision, S.W. and T.P.; project administration, S.W.; funding acquisition, S.W. and H.Z. All authors have read and agreed to the published version of the manuscript.

References

1. Gamper, U.; Boesiger, P.; Kozerke, S. Compressed sensing in dynamic MRI. *Magn. Reson. Med. Off. J. Int. Soc. Magn. Reson. Med.* **2008**, *59*, 365–373. [CrossRef] [PubMed]
2. Zhao, B.; Haldar, J.P.; Christodoulou, A.G.; Liang, Z.P. Image reconstruction from highly undersampled (k, t)-space data with joint partial separability and sparsity constraints. *IEEE Trans. Med. Imaging* **2012**, *31*, 1809–1820. [CrossRef] [PubMed]
3. Jung, H.; Ye, J.C.; Kim, E.Y. Improved k–t BLAST and k–t SENSE using FOCUSS. *Phys. Med. Biol.* **2007**, *52*, 3201. [CrossRef] [PubMed]
4. Wang, Y.; Ying, L. Compressed sensing dynamic cardiac cine MRI using learned spatiotemporal dictionary. *IEEE Trans. Biomed. Eng.* **2013**, *61*, 1109–1120. [CrossRef]
5. Caballero, J.; Price, A.N.; Rueckert, D.; Hajnal, J.V. Dictionary learning and time sparsity for dynamic MR data reconstruction. *IEEE Trans. Med. Imaging* **2014**, *33*, 979–994. [CrossRef]
6. Jung, H.; Sung, K.; Nayak, K.S.; Kim, E.Y.; Ye, J.C. k-t FOCUSS: A general compressed sensing framework for high resolution dynamic MRI. *Magn. Reson. Med. Off. J. Int. Soc. Magn. Reson. Med.* **2009**, *61*, 103–116. [CrossRef]
7. Otazo, R.; Candes, E.; Sodickson, D.K. Low-rank plus sparse matrix decomposition for accelerated dynamic MRI with separation of background and dynamic components. *Magn. Reson. Med.* **2015**, *73*, 1125–1136. [CrossRef]
8. Wang, S.; Xiao, T.; Liu, Q.; Zheng, H. Deep learning for fast MR imaging: A review for learning reconstruction from incomplete k-space data. *Biomed. Signal Process. Control* **2021**, *68*, 102579. [CrossRef]
9. Wang, S.; Cao, G.; Wang, Y.; Liao, S.; Wang, Q.; Shi, J.; Li, C.; Shen, D. Review and Prospect: Artificial Intelligence in Advanced Medical Imaging. *Front. Radiol.* **2021**, *1*, 781868. [CrossRef]
10. Li, C.; Li, W.; Liu, C.; Zheng, H. Artificial intelligence in multiparametric magnetic resonance imaging: A review. *Med. Phys.* **2022**, *49*, e1024–e1054. [CrossRef]
11. Wang, S.; Su, Z.; Ying, L.; Peng, X.; Zhu, S.; Liang, F.; Feng, D.; Liang, D. Accelerating magnetic resonance imaging via deep learning. In Proceedings of the 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI), Prague, Czech Republic, 13–16 April 2016; pp. 514–517.
12. Zhang, J.; Ghanem, B. ISTA-Net: Interpretable optimization-inspired deep network for image compressive sensing. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 1828–1837.
13. Eo, T.; Jun, Y.; Kim, T.; Jang, J.; Lee, H.J.; Hwang, D. KIKI-net: Cross-domain convolutional neural networks for reconstructing undersampled magnetic resonance images. *Magn. Reson. Med.* **2018**, *80*, 2188–2201. [CrossRef] [PubMed]
14. Sun, J.; Li, H.; Xu, Z.; Yang, Y. Deep ADMM-Net for compressive sensing MRI. *Adv. Neural Inf. Process. Syst.* **2016**, *29*.
15. Aggarwal, H.K.; Mani, M.P.; Jacob, M. MoDL: Model-based deep learning architecture for inverse problems. *IEEE Trans. Med. Imaging* **2018**, *38*, 394–405. [CrossRef] [PubMed]
16. Hammernik, K.; Klatzer, T.; Kobler, E.; Recht, M.P.; Sodickson, D.K.; Pock, T.; Knoll, F. Learning a variational network for reconstruction of accelerated MRI data. *Magn. Reson. Med.* **2018**, *79*, 3055–3071. [CrossRef] [PubMed]
17. Akçakaya, M.; Moeller, S.; Weingärtner, S.; Uğurbil, K. Scan-specific robust artificial-neural-networks for k-space interpolation (RAKI) reconstruction: Database-free deep learning for fast imaging. *Magn. Reson. Med.* **2019**, *81*, 439–453. [CrossRef] [PubMed]
18. Mardani, M.; Gong, E.; Cheng, J.Y.; Vasanawala, S.S.; Zaharchuk, G.; Xing, L.; Pauly, J.M. Deep generative adversarial neural networks for compressive sensing MRI. *IEEE Trans. Med. Imaging* **2018**, *38*, 167–179. [CrossRef]
19. Huang, Q.; Xian, Y.; Yang, D.; Qu, H.; Yi, J.; Wu, P.; Metaxas, D.N. Dynamic MRI reconstruction with end-to-end motion-guided network. *Med. Image Anal.* **2021**, *68*, 101901. [CrossRef]
20. Seegoolam, G.; Schlemper, J.; Qin, C.; Price, A.; Hajnal, J.; Rueckert, D. Exploiting motion for deep learning reconstruction of extremely-undersampled dynamic MRI. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Shenzhen, China, 13–17 October 2019; pp. 704–712.
21. Qin, C.; Schlemper, J.; Duan, J.; Seegoolam, G.; Price, A.; Hajnal, J.; Rueckert, D. k-t NEXT: Dynamic MR image reconstruction exploiting spatio-temporal correlations. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Shenzhen, China, 13–17 October 2019; pp. 505–513.

22. Schlemper, J.; Caballero, J.; Hajnal, J.V.; Price, A.N.; Rueckert, D. A deep cascade of convolutional neural networks for dynamic MR image reconstruction. *IEEE Trans. Med. Imaging* **2017**, *37*, 491–503. [CrossRef]
23. Qin, C.; Schlemper, J.; Caballero, J.; Price, A.N.; Hajnal, J.V.; Rueckert, D. Convolutional recurrent neural networks for dynamic MR image reconstruction. *IEEE Trans. Med. Imaging* **2018**, *38*, 280–290. [CrossRef]
24. Qin, C.; Duan, J.; Hammernik, K.; Schlemper, J.; Küstner, T.; Botnar, R.; Prieto, C.; Price, A.N.; Hajnal, J.V.; Rueckert, D. Complementary time-frequency domain networks for dynamic parallel MR image reconstruction. *Magn. Reson. Med.* **2021**, *86*, 3274–3291. [CrossRef]
25. Wang, S.; Ke, Z.; Cheng, H.; Jia, S.; Ying, L.; Zheng, H.; Liang, D. DIMENSION: Dynamic MR imaging with both k-space and spatial prior knowledge obtained via multi-supervised network training. *NMR Biomed.* **2022**, *35*, e4131. [CrossRef] [PubMed]
26. Ke, Z.; Huang, W.; Cui, Z.X.; Cheng, J.; Jia, S.; Wang, H.; Liu, X.; Zheng, H.; Ying, L.; Zhu, Y.; et al. Learned low-rank priors in dynamic MR imaging. *IEEE Trans. Med. Imaging* **2021**, *40*, 3698–3710. [CrossRef] [PubMed]
27. Hu, C.; Li, C.; Wang, H.; Liu, Q.; Zheng, H.; Wang, S. Self-supervised learning for mri reconstruction with a parallel network training framework. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Strasbourg, France, 27 September–1 October 2021; pp. 382–391.
28. Wang, S.; Wu, R.; Li, C.; Zou, J.; Zhang, Z.; Liu, Q.; Xi, Y.; Zheng, H. PARCEL: Physics-based Unsupervised Contrastive Representation Learning for Multi-coil MR Imaging. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2022**. [CrossRef]
29. Yoo, J.; Jin, K.H.; Gupta, H.; Yerly, J.; Stuber, M.; Unser, M. Time-dependent deep image prior for dynamic MRI. *IEEE Trans. Med. Imaging* **2021**, *40*, 3337–3348. [CrossRef] [PubMed]
30. Acar, M.; Çukur, T.; Öksüz, İ. Self-supervised Dynamic MRI Reconstruction. In Proceedings of the International Workshop on Machine Learning for Medical Image Reconstruction, Strasbourg, France, 1 October 2021; pp. 35–44.
31. Ulyanov, D.; Vedaldi, A.; Lempitsky, V. Deep image prior. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 9446–9454.
32. Yaman, B.; Hosseini, S.A.H.; Moeller, S.; Ellermann, J.; Uğurbil, K.; Akçakaya, M. Self-supervised learning of physics-guided reconstruction neural networks without fully sampled reference data. *Magn. Reson. Med.* **2020**, *84*, 3172–3191. [CrossRef]
33. Akçakaya, M.; Yaman, B.; Chung, H.; Ye, J.C. Unsupervised Deep Learning Methods for Biological Image Reconstruction and Enhancement: An overview from a signal processing perspective. *IEEE Signal Process. Mag.* **2022**, *39*, 28–44. [CrossRef]
34. Liang, D.; Cheng, J.; Ke, Z.; Ying, L. Deep Magnetic Resonance Image Reconstruction: Inverse Problems Meet Neural Networks. *IEEE Signal Process. Mag.* **2020**, *37*, 141–151. [CrossRef]
35. Qin, C.; Rueckert, D. Artificial Intelligence-Based Image Reconstruction in Cardiac Magnetic Resonance. In *Artificial Intelligence in Cardiothoracic Imaging*; Springer: Cham, Switzerland, 2022; pp. 139–147.
36. Candès, E.J.; Recht, B. Exact matrix completion via convex optimization. *Found. Comput. Math.* **2009**, *9*, 717–772. [CrossRef]
37. Lee, K.; Bresler, Y. Admira: Atomic decomposition for minimum rank approximation. *IEEE Trans. Inf. Theory* **2010**, *56*, 4402–4416. [CrossRef]
38. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [CrossRef]
39. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. In Proceedings of the ICLR (Poster), San Diego, CA, USA, 7–9 May 2015.