Coarse–Super-Resolution–Fine Network (CoSF-Net): A Unified End-to-End Neural Network for 4D-MRI with Simultaneous Motion Estimation and Super-Resolution

Shaohua Zhi, Yinghui Wang, Haonan Xiao, Ti Bai, Bing Li, Yunsong Tang, Chenyang Liu, Wen Li, Tian Li, Hong Ge* and Jing Cai*

***Abstract*— Four-dimensional magnetic resonance imaging (4D-MRI) is an emerging technique for tumor motion management in image-guided radiation therapy (IGRT). However, current 4D-MRI suffers from low spatial resolution and strong motion artifacts owing to the long acquisition time and patients' respiratory variations. If not managed properly, these limitations can adversely affect treatment planning and delivery in IGRT. In this study, we developed a novel deep learning framework called the coarse–super-resolution–fine network (CoSF-Net) to achieve simultaneous motion estimation and super-resolution within a unified model. We designed CoSF-Net by fully excavating the inherent properties of 4D-MRI, with consideration of limited and imperfectly matched training datasets. We conducted extensive experiments on multiple real patient datasets to assess the feasibility and robustness of the developed network. Compared with existing networks and three state-of-the-art conventional algorithms, CoSF-Net not only accurately estimated the deformable vector fields between the respiratory phases of 4D-MRI but also simultaneously improved the spatial resolution of 4D-MRI, enhancing anatomical features and producing 4D-MR images with high spatiotemporal resolution.**

***Index Terms*— Coarse-to-fine registration, Deep learning, Four-dimensional magnetic resonance Imaging, Super-resolution.**

Shaohua Zhi, Yinghui Wang, Haonan Xiao, Chenyang Liu, Wen Li, Tian Li and Jing Cai are with the Department of Health Technology and Informatics, The Hong Kong Polytechnic University, Hong Kong SAR, China. (e-mail: jing.cai@polyu.edu.hk)

Ti Bai is with the Medical Artificial Intelligence and Automation (MAIA) Laboratory, Department of Radiation Oncology, University of Texas Southwestern Medical Center, Dallas, Texas 75239, USA;

Bing Li and Hong Ge are with the Department of Radiation Oncology, the Affiliated Cancer Hospital of Zhengzhou University & Henan Cancer Hospital, Zhengzhou 450008, China.

This research was partly supported by research grants from the Centrally Funded Postdoctoral Fellowship (P0040712), The Hong Kong Polytechnic University, General Research Fund (GRF 15102118, GRF 15102219, GRF 15104822), the University Grants Committee, and the Health and Medical Research Fund (HMRF 06173276), the Health Bureau, the Government of the Hong Kong Special Administrative Regions, the Henan provincial Medical Science and Technology Research Project (SBGJ202103038 and SBGJ202102056), Henan Province Key R&D and Promotion Project (Science and Technology Research) (222102310015), Natural Science Foundation of Henan Province of China (232300420231).

* Prof. Jing Cai is the first corresponding author, Prof. Hong Ge is the second corresponding author.

I. Introduction

IMAGE-guided radiation therapy (IGRT) has been widely adopted in clinic for precision radiotherapy in patients with cancer [1]. In the past decade, magnetic resonance imaging (MRI) has gained much attention in IGRT because of its superior soft-tissue contrast and zero radiation hazard as compared to X-ray imaging techniques, such as computed tomography (CT) and cone-beam CT [2], [3]. In particular, MRI plays an important role in IGRT for abdominal cancers as it provides images with excellent anatomical details for accurate tumor volume delineation and possesses dynamic imaging capacity for tumor motion management [4], [5].

During radiotherapy treatment of abdominal and lung cancer patients, physiological motion (respiratory motion) is one of the major factors influencing treatment precision. However, respiratory motion can cause significant treatment errors if not managed properly. Thus, it is critical to manage respiratory motion when performing radiotherapy for abdominal cancers [6], [7]; this becomes particularly crucial when using stereotactic body radiation therapy (SBRT) [8], a modern radiotherapy technology that precisely delivers radiation treatment using a much higher radiation dose (10× higher) than conventional radiotherapy. Improper management of tumor motion when using SBRT can adversely affect patient treatment to a much greater degree than when it occurs in conventional radiotherapy.

Four-dimensional magnetic resonance imaging (4D-MRI) is an emerging technique for motion management in the radiotherapy of mobile abdominal tumors. To date, various 4D-MRI techniques have been developed, and their promises have been well demonstrated [9]–[13]. One important imaging approach of 4D-MRI is fast volumetric MRI [14], in which the volume of interest is imaged at a sub-second speed, yielding real-time 4D-MR images. Furthermore, deformable image registration (DIR) can be performed on 4D-MR images to generate patient-specific motion models that depict voxel-wise motion patterns at different respiratory phases. The derived motion models combined with 4D-MR images are of great value in aiding precise radiotherapy, including 4D treatment planning, internal target volume determination, tumor tracking [15], 4D dose calculation, and organs at risk sparing [16].

Currently, 4D-MRI is still under investigation and development. There are a number of challenges to overcome before 4D-MRI can be fully adapted to the clinical setting. First, 4D-MR images suffer from limited spatial resolution; i.e., the temporal resolution of real-time 4D-MRI is approximately 1 s, whereas its voxel size is isotropically approximately 3 mm [9]. Owing to its insufficient image quality, as evidenced by a relatively low signal-to-noise ratio (SNR) and image artifacts caused by breathing variations, 4D-MRI may fail to display detailed anatomical structures. Second, in the abdominal region, it is difficult to calculate and model deformable vector fields (DVFs) from 4D-MR images for tumor tracking, primarily due to the extensive respiratory-related deformations, complicated soft anatomy variations, and low-resolution images. These deficiencies of 4D-MRI can adversely affect its applications and diminish its values in IGRT.

With the development of the deep learning (DL) technique, data-driven approaches have been proposed and widely applied to medical imaging fields for quick calculation times and powerful learning ability. Numerous potential DL-based solutions have been broadly divided into two categories to help address the deficiencies of the current 4D-MRI techniques: First, image enhancement [17] or super-resolution (SR) [18]–[22] models may assist on directly improving the MR image quality. The main idea is to learn the mapping from low-resolution (LR) images to high-resolution (HR) images, thus restoring high-frequency structures as much as possible. However, most magnetic resonance imaging (MRI)-related super-resolution (SR) studies so far have focused on three-dimensional (3D)-MR images or 2D+time MR sequence [17], while the spatiotemporal correlation of 4D-MR images remains little exploited. Second, the category of DL-based DIR methods seeks to estimate DVFs for motion compensation or multi-contrast registration [23] of dynamic MRI. This strategy has been successfully applied in the field of cardiac motion estimation in image domain [24]–[26] or with k-space data guidance [27], [28]. For respiratory motion estimation of abdominal 4D-MRI, a DL-based DIR study has also been recently explored to learn the accurate motion trajectory and apply it to synthesis ultra-quality 4D-MR images [29].

Although DL-based models have resulted in breakthroughs, there is still much room for improvement of 4D-MRI. For example, it is known that DVF estimation accuracy is highly sensitive to image quality in DIR algorithms. For 4D-MRI, its voxel values at different respiratory phases may vary due to imaging factors and image artifacts. As a result, the compromised and inconsistent quality of 4D-MRI can lead to registration errors. Although coarse-to-fine registration mechanism was proven successful for many medical imaging applications [30]–[32] by predicting DVF progressively in multi-image levels, it has limitations when applied to 4D-MRI, such as the loss of subtle structures and misalignments. Thus, in this study, we were motivated to develop an upgraded registration architecture to improve the performance of coarse-to-fine registration with respect to 4D-MR images.

The data preparation and pre-processing for 4D-MRI training also require careful consideration. First, obtaining the reference DVFs (training labels) for 4D-MRI registration is difficult, and the labeling process is time-consuming and laborious. Second, 4D-MRI is still under the investigational stage and is not being routinely used in clinical practice yet. 4D-MRI studies generally involve a small patient sample size ($\leq$ 20 patients) [19], [33], [34], posing a great challenge for DL-based analysis because a small sample size can cause over-fitting during DL network training and subsequently affect model robustness. Second, it is common in radiotherapy that 3D T1-/T2-weighted MR images of the same patient are always available together with their 4D-MR images, which can be regarded as prior knowledge to promote image quality. It is clear that we need to develop DL networks tailored for 4D-MR images to overcome the mentioned limitations.

In this study, our aim was to develop an end-to-end network capable of simultaneously achieving accurate DVF estimation and image resolution enhancement. The concept of joint motion correction and SR is derived from conventional regularized optimization algorithms for 3D cardiac cine MRI reconstruction [35], precise cardiac segmentation [36], quantitative MRI parameter mapping improvement [37], and brain MRI for fetal posterior fossa measurements via novel acquisition technique [38]. However, to date, there has been limited investigation into the development of a DL-based unified motion modeling and SR approach for abdominal 4D-MRI tasks. To address this, we developed a novel DL-based framework for 4D-MRI called the CoSF-Net, which effectively exploits the inherent prior information in 4D-MRI. CoSF-Net comprises three individual modules, including an unsupervised coarse DVF estimation module, an SR improvement module, and a refined DVF estimation module associated with the initial predictions from the previous two modules.

The main contributions of CoSF-Net are summarized as follows:

1) CoSF-Net integrates three sub-models in an end-to-end fashion. To the best of our knowledge, it is the first DL framework provides a comprehensive solution for simultaneous motion estimation and SR in 4D-MRI applications.

2) We developed and embedded an SR model between the two registration sub-models to construct a coarse–SR–fine architecture. This architecture can boost the registration performance, particularly when the input 4D-MRI pairs have a low spatial resolution. To address the challenges of the limited and imperfect matched training pairs in 4D-MRI, we designed a 2.5-dimensional conditional generative adversarial network (2.5D-cGAN). This model leverages spatial information from neighboring slices, which mitigates blurring of structures due to through-plane motion. Particularly, we employed a U-Net based discriminator, which offers more detailed per-pixel feedback to the generator compared to a standard GAN discriminator.

3) Both the coarse and fine DIR modules use unsupervised training based on the VoxelMorph (VM) model [39]. The integration of the coarse DIR CNN into CoSF-Net framework allows us to provide a warm start for DVF estimation with the input of original 4D-MRI, which simplifies and accelerates the subsequent fine DIR and SR stages. Additionally, in the fine DIR model, we incorporated an independent feature extraction pathway for the prior MR image to enhance detailed DVF estimation. We also adopted a residual DVF estimation

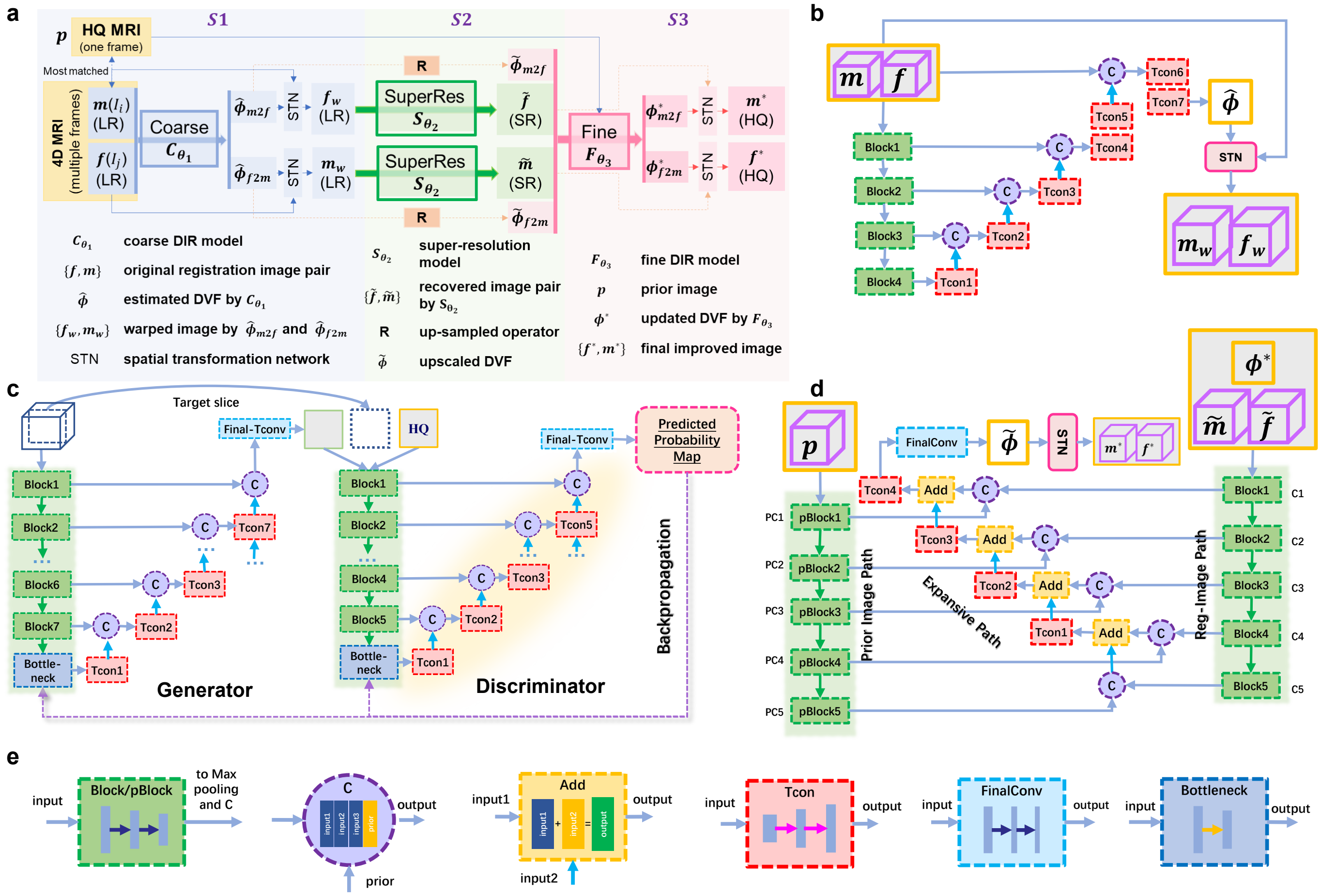

Fig. 1. The detailed architecture of the proposed network. (a) A schematic illustration depicting the process of training CoSF-Net. The network involves the three cascaded sub-models in different colors: the coarse registration model ($\boldsymbol{C}_{\theta_1}$, blue), the SR model ($\boldsymbol{S}_{\theta_2}$, green), and the fine registration model ($\boldsymbol{F}_{\theta_3}$, pink); (b) Stage 1: A coarse DIR CNN to predict a rough DVF between the phases in 4D-MRI at the low-resolution level; (c) Stage 2: A 2.5D-cGAN-based SR model to enhance the image quality and resolution of the deformed 4D-MR images; (d) Stage 3: A Fine DIR CNN to further compute the DVF residue for improved 4D-MR images from stage 2 at a high-resolution level, same as for the normal MRI; (e) individual convolutional blocks in CoSF-Net, from left to right are convolutional unit (Block/pBlock) in every layer of three sub-networks, concatenation unit (C), add unit (Add), transposed convolutional unit (Tconv), and final convolutional unit (FinalConv).

mechanism to update the refined DVF.

4) We conducted extensive real-patient experiments with both visual comparison and quantitative evaluation to validate the effectiveness of CoSF-Net.

The rest of this article is organized as follows. In Section II, we describe the framework and implementation of the proposed network. Section III presents the experimental setup, data arrangement, evaluation metrics, and competitive algorithms adopted in this study. In Section IV, we report and analyze the results of our experiments conducted using real-patient data. In Section V, we discuss the relevant problems and conclude the study.

II. Methods and Materials

A. Problem Definition and Notations

Figure 1(a) presents a schematic illustration of the problem definition and notations of 4D-MRI. We denote $\boldsymbol{Q} \supset \{I^1, I^2, \ldots, I^K\}$, the original 4D-MRI sequence comprising K respiratory-correlated phases/frames. For simplicity, a pair of arbitrary phases $\{I^i, I^j\}$ $(I^i, I^j \in \boldsymbol{Q}, i \neq j)$ in 4D-MRI is denoted as the moving image $\boldsymbol{m}$ and fixed image $\boldsymbol{f}$, respectively. The purpose of our method is to develop a unified model $\boldsymbol{U}_{\boldsymbol{\Theta}}$ parameterized by $\boldsymbol{\Theta}$ for enhancing the image quality of 4D-MRI while simultaneously estimating the DVF between the enhanced frames. The objective function can be described as follows:

$$\widehat{\boldsymbol{\Theta}} = \arg\min_{\boldsymbol{\Theta}} \boldsymbol{U}_{\boldsymbol{\Theta}}(\boldsymbol{f}, \boldsymbol{m}, \boldsymbol{p}), \tag{1}$$

where $\boldsymbol{p}$ denotes a clinical T1-weighted MRI scan from the same patient, designated as a prior MR image. As shown in Fig. 1(a), the unified model is split into three cascaded sub-models, including the coarse registration model ($\boldsymbol{C}_{\theta_1}$, blue), the SR model ($\boldsymbol{S}_{\theta_2}$, green), and the refined registration model ($\boldsymbol{F}_{\theta_3}$, pink). The coarse DIR model $\boldsymbol{C}_{\theta_1}$ is used to calculate the coarse DVF using DL-based model $\boldsymbol{\phi} = \boldsymbol{C}_{\theta_1}(\boldsymbol{f}, \boldsymbol{m})$, considering the registration pair $\{\boldsymbol{f}, \boldsymbol{m}\}$ as the input. The optimization problem can be modeled as follows:

$$\widehat{\boldsymbol{\theta}_1} = \arg\min_{\boldsymbol{\theta}_1} \mathcal{L}_1\left(\boldsymbol{m}, \boldsymbol{f}, \boldsymbol{C}_{\boldsymbol{\theta}_1}(\boldsymbol{f}, \boldsymbol{m})\right), \tag{2}$$

where $\widehat{\boldsymbol{\theta}_1}$ denotes the learnable parameters of $\boldsymbol{C}_{\boldsymbol{\theta}_1}$ and $\mathcal{L}_1$ denotes the loss function.

We also denote the SR model $\boldsymbol{S_{\theta_2}}$ parameterized by $\boldsymbol{\theta_2}$ as a HR image ($\boldsymbol{I}_{\text{HR}}$) generation procedure based on observed LR counterpart ($\boldsymbol{I}_{\text{LR}}$), the solution for which can be expressed by:

$$\widehat{\boldsymbol{\theta}_2} = \arg\min_{\boldsymbol{\theta}_2} \mathcal{L}_2\left(\boldsymbol{S_{\theta_2}}\left(\boldsymbol{I}_{\text{LR}}\right), \boldsymbol{I}_{\text{HR}}\right), \tag{3}$$

where $\boldsymbol{S_{\theta_2}}$:$\boldsymbol{I}_{\text{LR}} \rightarrow \boldsymbol{I}_{\text{HR}}$ can be replaced by the cGAN structure with a generator $\boldsymbol{G}$ and a discriminator $\boldsymbol{D}$. Both $\boldsymbol{G}$ and $\boldsymbol{D}$ can be optimized in an alternative manner to solve the adversarial min-max problem as follows [40]:

$$\begin{aligned} \boldsymbol{S}_{\boldsymbol{\theta}_2} &= \{\boldsymbol{G}, \boldsymbol{D}\} \\ \boldsymbol{G}^* &= \arg\min_{\boldsymbol{G}} \max_{\boldsymbol{D}} \mathcal{L}_2(\boldsymbol{G}, \boldsymbol{D}). \end{aligned} \tag{4}$$

Finally, the fine DIR model $\boldsymbol{F_{\theta_3}}$ parameterized by $\boldsymbol{\theta_3}$ feeds the enhanced HR image pair $\{\tilde{\boldsymbol{f}}, \tilde{\boldsymbol{m}}\}$ through $\boldsymbol{S_{\theta_2}}$, the up-sampled DVF $\widehat{\boldsymbol{\phi}} = \boldsymbol{R}(\tilde{\boldsymbol{\phi}})$, together with the prior MRI image $\boldsymbol{p}$ to estimate a finer DVF $\boldsymbol{\phi}^*$. $R(.)$ indicates a trilinear interpolation operator and is fixed, non-trainable operator. It interpolates the coarsely predicted DVF in low resolution to a higher resolution for further DVF fine-tuning. The optimization function can be written as:

$$\widehat{\boldsymbol{\theta}_3} = \arg\min_{\boldsymbol{\theta}_3} \mathcal{L}_3\left(\tilde{\boldsymbol{f}}, \tilde{\boldsymbol{m}}, \boldsymbol{F}_{\boldsymbol{\theta}_3}(\tilde{\boldsymbol{f}}, \tilde{\boldsymbol{m}}, \boldsymbol{p}, \tilde{\boldsymbol{\phi}})\right) \tag{5}$$

To sum up, CoSF-Net is a cascade of three individual sub-models; this can be denoted uniformly using the following equation $\boldsymbol{U_\Theta} = \{\boldsymbol{C_{\theta_1}}; \boldsymbol{S_{\theta_2}}; \boldsymbol{F_{\theta_3}}\}$.

B. Network Architecture

The overall workflow of the proposed CoSF-Net is outlined as follows. The first stage estimates a coarse DVF $\widehat{\boldsymbol{\phi}}$ of the input pair $\{\boldsymbol{f}, \boldsymbol{m}\}$ with a coarse DIR model $\boldsymbol{C_{\theta_1}}$, deforming $\{\boldsymbol{f}, \boldsymbol{m}\}$ through a spatial transformation network (STN) to $\{\boldsymbol{f_w}, \boldsymbol{m_w}\}$. In the second stage, the HR images $\{\tilde{\boldsymbol{f}}, \tilde{\boldsymbol{m}}\}$ are recovered from $\{\boldsymbol{f_w}, \boldsymbol{m_w}\}$ through the SR model slice by slice. In the final stage, the recovered image pair $\{\tilde{\boldsymbol{f}}, \tilde{\boldsymbol{m}}\}$, the up-sampled DVF $\tilde{\boldsymbol{\phi}}$, and the prior MR image $\boldsymbol{p}$ are fed together into the fine DIR CNN $\boldsymbol{F_{\theta_3}}$ to calculate an updated residual DVF $\boldsymbol{v}$ and finally to obtain a finer DVF $\boldsymbol{\phi}^*$ and the corresponding deformed HR images $\{\boldsymbol{f^*}, \boldsymbol{m^*}\}$. A combination of three cascaded modules is considered the coarse-SR-fine structure.

1) Stage 1: Coarse DIR CNN:

Fig. 1 (b) shows the coarse DIR CNN, which inputs a concatenation of two arbitrary phases of MRI to predict the DVF between them. The network architecture was inspired by VM and trained in an unsupervised manner. To be specific, U-Net [41] comprises a contracting path of four 3D-convolutional blocks, a bottleneck connection, an expansive path of four 3D-convolutional blocks, and a final output layer. The STN [42] is used for calculating the deformed volume $\boldsymbol{M} \circ \widehat{\boldsymbol{\phi}}$ with the symbol $\circ$ denoting the deformable transformation operation based on the STN. In particular, we added an inverse-consistency penalty to render the DVF bidirectional $\widehat{\boldsymbol{\phi}} = \{\widehat{\boldsymbol{\phi}}_{m2f}, \widehat{\boldsymbol{\phi}}_{f2m}\}$, which means that $\boldsymbol{f}$ and $\boldsymbol{m}$ can deform each other. Accordingly, the loss function $\mathcal{L}_1(.)$ in Eq. (2) contains two similarity terms: $\mathcal{L}_{sim}(.)$, measuring the image difference between the target and warped images, and a diffusion regularizer $\mathcal{L}_{\text{smooth}}(.)$, encouraging a smooth DVF.

$$\begin{aligned} &\mathcal{L}_1\left(\boldsymbol{m}, \boldsymbol{f}, \boldsymbol{C_{\theta_1}}(\boldsymbol{f}, \boldsymbol{m})\right) \\ &\quad = \mathcal{L}_{\text{sim}}\left(\boldsymbol{f}, \boldsymbol{m} \circ \widehat{\boldsymbol{\phi}}_{\boldsymbol{m2f}}\right) + \mathcal{L}_{\text{sim}}\left(\boldsymbol{m}, \boldsymbol{f} \circ \widehat{\boldsymbol{\phi}}_{f2m}\right) \\ &\quad\quad + \lambda_1 \mathcal{L}_{\text{smooth}}(\widehat{\boldsymbol{\phi}}) \end{aligned} \tag{6}$$

Negative normalized cross-correlation (NCC) is employed in $\mathcal{L}_{\text{sim}}(.)$ instead of the L1 or L2 norm calculation owing to varying intensities among phases in 4D-MRI. The regularization parameter λ_1 is used to control the trade-off between the fidelity and regularization terms.

2) Stage 2: SR Network:

In stage 2, the SR model uses the cGAN to synthesize improved MR images from the 4D-MR image counterparts. Fig. 1(c) illustrates the detailed architecture of the SR model, also known as the 2.5D-cGAN. The proposed model inputs the 2.5D images by considering the presence of limited volumetric MRI training data and the high inter-slice correlation within the volume. Hence, five consecutive 2D transversal slices along the plane dimension are integrated and fed into the 2.5D-cGAN for predicting the central slice. The proposed GAN model is a modified version of pixel2pixel [43] network for image-to-image translation tasks. We updated the discriminator module as U-Net instead of a simple classification network. Specifically, U-Net based discriminator [44], [45] is an extension of patchGAN, by reusing convolutional layers of original patchGAN in pixel2pixel as the encoder part and adding upsampling convolutions to build the decoder part.

For the loss function, apart from the adversarial loss ($\mathcal{L}_{\text{GAN}}(\boldsymbol{G}, \boldsymbol{D})$) and the pixel-wise L1-based intensity loss ($\mathcal{L}_{\text{L1}}(\boldsymbol{G})$), a multi-scale structural similarity index (MS-SSIM) ($\mathcal{L}_{\text{SSIM}}(\boldsymbol{G})$) [46] is incorporated into the total loss function of the 2.5D-cGAN, restoring structural information effectively as follows:

$$\begin{aligned} \mathcal{L}_2(\boldsymbol{G}, \boldsymbol{D}) &= \mathcal{L}_{GAN}(\boldsymbol{G}, \boldsymbol{D}) + \lambda_2 \mathcal{L}_{L1}(\boldsymbol{G}) + \lambda_3 \mathcal{L}_{\text{SSIM}}(\boldsymbol{G}) \\ &= \mathbb{E}_{(I_{\text{LR}}, I_{\text{SR}})}\left[\log \boldsymbol{D}\left(I_{\text{LR}}, I_{\text{SR}}\right)\right] \\ &\quad + \mathbb{E}_{I_{\text{LR}}}\left[\log\left(1 - \boldsymbol{D}\left(I_{\text{LR}}, \boldsymbol{G}\left(I_{\text{LR}}\right)\right)\right)\right] \\ &\quad + \lambda_2 \mathbb{E}_{(I_{\text{LR}}, I_{\text{SR}})}\left[\left\|I_{\text{SR}} - \boldsymbol{G}\left(I_{\text{LR}}\right)\right\|_1\right] \\ &\quad + \lambda_3\left(1 - \text{SSIM}\left(I_{\text{SR}}, \boldsymbol{G}\left(I_{\text{SR}}\right)\right)\right), \end{aligned} \tag{7}$$

where $\mathbb{E}$ indicates the expected value, and the MS-SSIM index is computed using the same parameters proposed in a previous study [47]. Both λ_2 and λ_3 are weighting factors that control the relative importance of L1 and SSIM losses, respectively. For technical details, the generator $\boldsymbol{G}$ comprises seven blocks of convolution-BN-ReLU operations in both the encoder and decoder, while the discriminator $\boldsymbol{D}$ contains five identical blocks.

3) Stage 3: Fine DIR CNN:

As mentioned in Section II-A, a combination of the three sub-models can be regarded as a coarse-SR-fine structure. The architecture of the fine DIR CNN is displayed in Fig. 1(d). It contains two independent paths in the encoder module for extracting multi-level features; one is called Reg-Image-Path for the registration pair $\{\tilde{\boldsymbol{f}}, \tilde{\boldsymbol{m}}\}$, whereas the other is called Prior-Image-Path for $\boldsymbol{p}$. A concatenation of both features at the same scale is then delivered into the decoder module. It is worth noting that the prior MR image $\boldsymbol{p}$ is not perfectly matched with the moving image. Before integrating $\boldsymbol{p}$ into the network, it was pre-aligned to be in the same phase as the enhanced moving image $\tilde{\boldsymbol{m}}$. Instead of explicitly calculating $\boldsymbol{\phi}^*$, we adopted a residual DVF calculation strategy. This strategy utilizes $\tilde{\boldsymbol{\phi}}$ as an initial guess to obtain a warped volume using the function $\tilde{\boldsymbol{m}} \circ \tilde{\boldsymbol{\phi}}_{m2f}$. For back-propagation, the loss function of the fine DIR CNN is accordingly modified as follows:

$$\begin{aligned}&\mathcal{L}_3\left(\tilde{\boldsymbol{f}}, \tilde{\boldsymbol{m}}, \boldsymbol{F}_{\boldsymbol{\theta}_3}(\tilde{\boldsymbol{f}}, \tilde{\boldsymbol{m}}, \boldsymbol{p}, \tilde{\boldsymbol{\phi}})\right)\\ &= \alpha \mathcal{L}_{\text{sim}}\left(\tilde{\boldsymbol{f}}, \tilde{\boldsymbol{m}} \circ \widetilde{\boldsymbol{\phi}}_{m2f}\right) + (1-\alpha)\mathcal{L}_{\text{sim}}\left(\tilde{\boldsymbol{f}}, \boldsymbol{p} \circ \widetilde{\boldsymbol{\phi}}_{m2f}\right)\\ &\qquad + \lambda_4 \mathcal{L}_{\text{smooth}}(\widetilde{\boldsymbol{\phi}}),\end{aligned} \tag{8}$$

where the predicted $\widetilde{\boldsymbol{\phi}}_{m2f}$ is constrained by two similarity forms: one is controlled by $\tilde{\boldsymbol{f}}$ with improved resolution, whereas the other one is attributed to the prior MRI. The parameter α is adopted to adjust the contribution factors of both similarity metrics. The $\mathcal{L}_{\text{sim}}(.)$ and $\mathcal{L}_{\text{smooth}}(.)$ in Eq. (8) for fine DIR CNN are the same as those in Eq. (6) for coarse DIR CNN. Using the proposed fine DIR CNN, we could then estimate a residual DVF $\boldsymbol{v}$ between $\tilde{\boldsymbol{m}} \circ \tilde{\boldsymbol{\phi}}_{m2f}$ and $\tilde{\boldsymbol{f}}$. Residual DVF $\boldsymbol{v}$ not only reflects a more accurate DVF but also contains detailed anatomic changes. In doing so, the updated DVF can be obtained with $\boldsymbol{\phi}^* = \boldsymbol{v} + \tilde{\boldsymbol{\phi}}$.

III. EXPERIMENTS

A. Dataset Preparation

The MRI data used in this study were acquired from patients with liver tumors undergoing radiotherapy using a 3T scanner (Skyra, Siemens, Erlangen, Germany). The study protocol was approved by the institutional review board (IRB). The patients underwent regular 3D MRI scans (T1- and T2-weighted), which were designated as “prior MRI”. In addition to regular 3D MRI, each patient also underwent 4D-MRI using a TWIST-VIBE (time-resolved imaging of contrast kineticS-volumetric interpolated breath-hold examination) MRI sequence by Siemens. This specific MRI sequence is used to acquire multiple 3D volumes within seconds. In reality, all patients underwent free-breathing imaging without employing a breath-hold technique. For each patient, it continuously generated 72 frames covering several respiratory cycles. Ten frames covering a breathing cycle were then selected from the original 4D-MRI using the body area method [48]; these frames represent the respiratory-correlated phases used in this study. The dimension of each volume of the 4D-MRI is 160 × 128 × 64, with a voxel size of 2.7 mm × 2.7 mm × 3.0 mm; the dimensions of the prior MRI is 320 × 320 × 72, with a smaller voxel size of 1.2 mm × 1.2 mm × 3.0 mm.

Particularly, we implemented a pre-processing procedure to generate LR/HR image pairs for SR model training as shown in Fig. 2. In our study, we utilized clinical 3D MRI scans as the ground truth and a specific phase of 4D-MRI as the input, both acquired using the same sequence. First, we conducted frame selection by computing cross-correlation between all 4D-MRI frames and the corresponding 3D MRI to determine the frame with the highest value, which corresponds to the closest breathing phase to the 3D-MRI. Next, we resized the selected frame to match the same scale as the 3D-MRI with higher resolution. Then, we employed the Elastix [49] toolbox to register the original 3D-MRI to the selected frame to address residual mismatches. After these pre-processing procedures, the aligned 3D-MRI was considered the corresponding HR image with regard to the LR image.

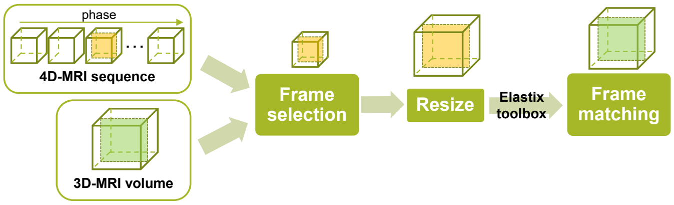

Fig. 2. The generation workflow of LR/HR image pairs of SR training dataset.

For network training and evaluation, we retrospectively included a total of twenty-seven MRI patients, all of whom underwent 4D-MRI scans and prior 3D MRI scans. Twenty patients were used for network training and validation, while the remaining seven patients were employed for testing and evaluation. To address the issue of limited training data, we performed data augmentation by expanding the dataset by 90°,180°, and 270° rotations. This yielded 2,480 2D pairs of transverse MR images for the 2.5D-cGAN training. Although rotations beyond 90° may not resemble realistic scenarios, these augmented images help the model to learn various spatial transformations and to generalize better, especially for DIR CNNs. Moreover, we set the respiratory phase number to ten and quantified the degree of deformation change into four grades (termed phase interval) according to the breathing amplitude for further data augmentation. Finally, 3,200 volumetric pairs were obtained for coarse and fine DIR CNNs training. The intensity of all the images was normalized to 0-1. Regarding the seven testing datasets, four representative cases were displayed for visualization, and all seven patients were included in the quantitative analysis.

B. Implementation Details

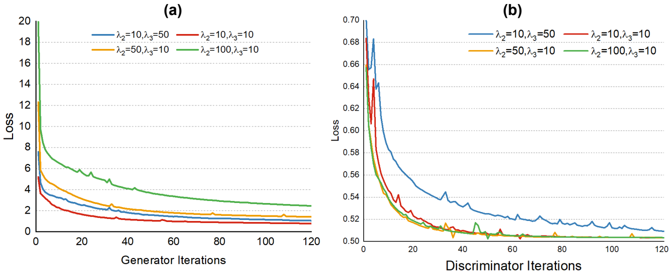

Fig. 3. Loss curves of Generator (a) and Discriminator (b) in 2.5D-cGAN with different values of the weighting factor $\boldsymbol{\lambda_2}$ and $\boldsymbol{\lambda_3}$

1) Training procedure/strategy: To ensure that the three sub-models in CoSF-Net play their expected roles, we trained the framework in two stages. First, pre-trained weights of coarse DIR CNN and 2.5D-cGAN are frozen while the fine DIR CNN was cascaded with these two pre-trained sub-models, ensuring a minimum and stable output. All the sub-models in CoSF-Net used ADAM optimizer for network training.

2) Hyper-parameter settings: We presented the details of our experimental setup in Table I. In coarse DIR CNN, a total of 300 epochs were obtained for it. λ_1 in Eq. (6) presents a tradeoff between the image similarity term and DVF smoothness term and we set λ_1=4 based on the original VoxelMorph[40] results. The learning rates of coarse DIR CNN was set to 4e-4 initially and decreased to 90% after every 30 epochs. In the 2.5D-cGAN training process, the network converged after 120 epochs with a batch size of 55. For the configuration of regularization parameters, the loss curves of both G and D with different parameter values were plotted in a controlled setting, as shown in Fig. 3. Based on our observations and controlled settings, we chose λ_2=10 and λ_3=10 in Eq. (7), which provided a stale output and led to a well-balanced training between G and D. The learning rate of 2.5D-cGAN was set to 4e-4 initially and decreased to 90% after every 30 epochs.

In the second stage of model fine-tuning, the regularization parameter α in the fine DIR CNN, was set to 0.35 to adjust the relative weights between the prior image and original image fidelity. It is learned from the parameter setting of PICCS [50], a conventional dynamic imaging reconstruction algorithm. The fine-tuning learning rate was set to 5e-5 and adjusted to 90% after every 10 epochs. Meanwhile, the regularization parameter was set as λ_4=6 in Eq. (8) empirically.

TABLE I

HYPERPARAMETER CONFIGURATION AND MODEL PERFORMANCE METRICS

Configuration		Coarse DIR CNN	GAN	Fine DIR CNN
Memory footprint		22048MB	23955MB	21574MB

No. of weights		396457	54412481(G)
			54409409(D)
				3093922

Batch size		30	55	1
Learning rate		4e-4	4e-4	5e-5
Epochs		300	120	100
Regularization parameter		$\lambda_1 = 4$	$\lambda_2 = 10$ $\lambda_3 = 10$	$\alpha = 0.35$ $\lambda_4 = 5$
Run-time	Training (seconds)	27360	13061	69600
	Inference (ms)	18.23	1048.75	1148.35

Note: All experiments were conducted on a single NVIDIA GTX3090GPU (24576MB). The corresponding training and testing codes will be available on the authors' website once the paper is published.

3) Model performance metrics: We present the details of our experimental setup and the performance metrics. Table I also reports the run-time of training and inference, memory footprint, number of weights, batch size for our experiments.

C. Model Evaluation

We evaluated the proposed CoSF-Net both qualitatively and quantitatively in the following aspects: 1) We analyzed the intermediate results at individual stages to evaluate the effectiveness of the proposed coarse-SR-fine framework; 2) We compared CoSF-Net with existing DL-based neural networks and conventional optimization-based algorithms; 3) We conducted an ablation study to investigate the impact of the designed components on the network performance; 4) We analyzed the tumor localization and feature recovery in 4D-MR images using the proposed CoSF-Net.

1) Comparison with existing conventional DIR algorithms and DL models: Since the proposed CoSF-Net is a cascade of three sub-models, we first reported the intermediate results sequentially to show the performance of each sub-model, whose results are shown in Section IV-A. Second, we carried out an ablation study to investigate the contribution of each sub-model in CoSF-Net by ablating each sub-model from the entire network, including the coarse DIR CNN (CoSF-Net(w/o coarse)), GAN-based SR component (CoSF-Net(w/o SR)), a prior MRI component in fine DIR CNN by using U-Net instead of N-Net (CoSF-Net (w/o prior p)). Both visualization results and quantitative evaluation were reported in Section IV-B.

Third, existing registration-related conventional algorithms DL models and SR-related models are implemented to fairly compare with the proposed CoSF-Net. To be specific, we compared the proposed network with the following three classical registration algorithms: pTV algorithm [51], Elastix [49], and Demons [52]. Two DL-based DIR models were employed to assess the effectiveness of CoSF-Net, including the supervised VM (sVM) and single-scale unsupervised VM (uVM). In implementing sVM, the reference DVFs were generated from pTV algorithm. This part of results is reported in Section IV-C For the SR recovery ability, both a classical CNN, namely enhanced deep super-resolution network (EDSR) [18] and a GAN-based image-to-image synthesis network pixel2pixel [43] were used to compare with the proposed 2.5D-cGAN. In EDSR, it consists of sixteen layers of Conv+ReLU blocks without BN function. In pixel2pixel, its generator follows the same architecture of 2.5D-cGAN, while the discriminator is four layers of convolutional blocks (Conv+ReLU) with a receptive field of 70×70. The analysis of SR-related results is reported and discussed in Section IV-D. All the parameters were carefully selected to ensure a fair comparison.

2) Evaluation metrics: For quantitative evaluation, we firstly evaluated the DVF estimation accuracy by calculating end-point error (EPE) $EPE = \|\mu - \mu_{ref}\|_2$ and end-angulation error (EAE) $EAE = arg\left(\mu, \mu_{ref}\right)$ between the predicted DVF and the reference DVF obtained from pTV algorithm.

Furthermore, we measured the error distance between the warped and the fixed images using the rooted mean square error (RMSE), the structural similarity index metric (SSIM) [53], the peak signal-to-noise ratio (PSNR), and normalized mutual information (NMI) [54], respectively. RMSE calculates the absolute difference between the restored image and the ground truth; SSIM evaluates the preserving ability of structural information; PSNR represents the noise suppression ability; and NMI evaluates the correlation between the ground truth and the reconstructed images. A larger NMI value indicates a higher similarity with the ground truth, whereas a smaller value indicates a lower similarity.

IV. RESULTS

A. Intermediate Results Analysis

Figure 4 presents the intermediate results for Patient I (a) and II (b). Particularly, we selected the registration pair with the maximum phase interval to test the robustness of the network, that is, the moving image is at the end-of-inhale (EOI)

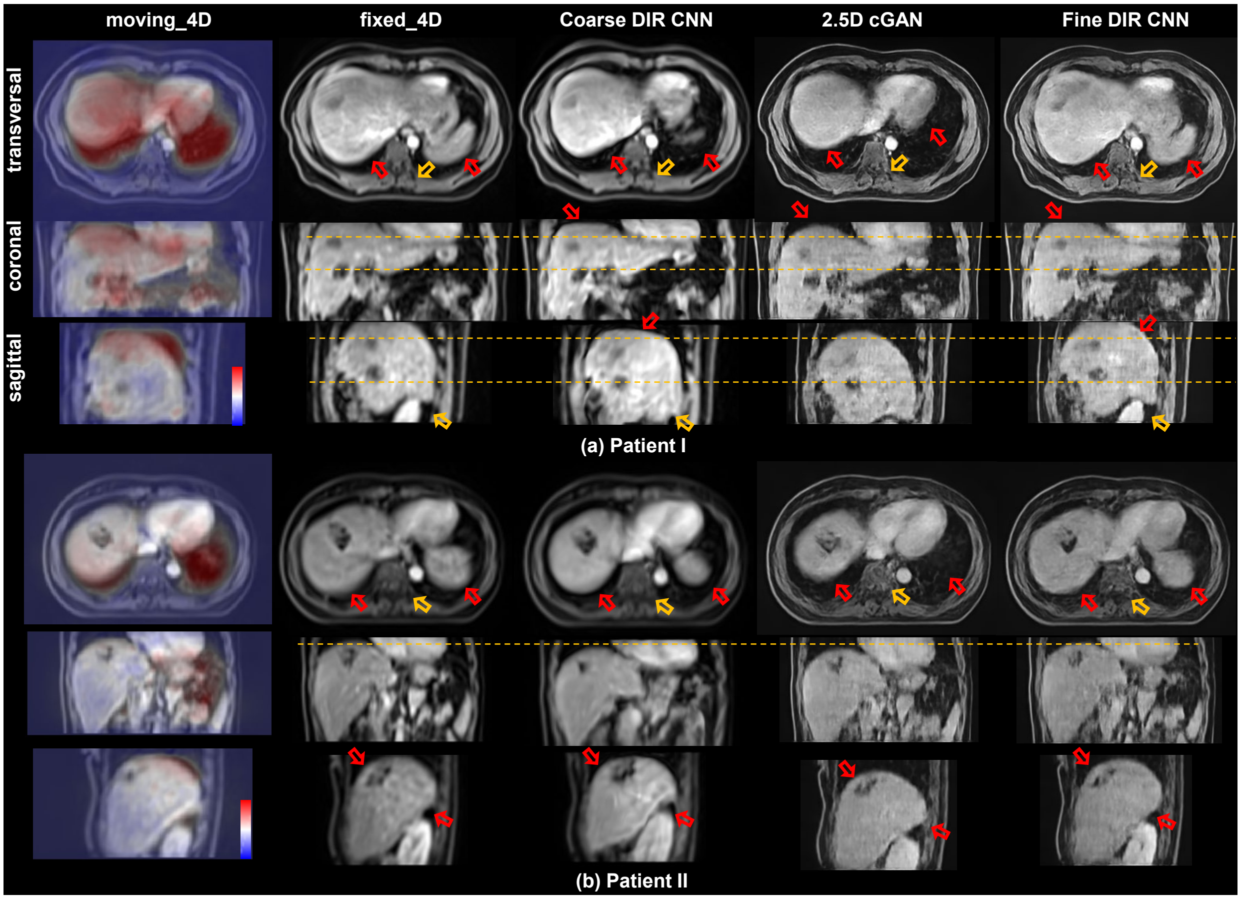

Fig. 4. The results of 4D-MRI for (a) Patient I and (b) Patient II. From left to right: moving image in original 4D-MRI (moving), fixed image in original 4D-MRI (fixed), intermediate warped image from the moving image by coarse DIR CNN, the enhanced moving image via 2.5D-cGAN, and final warped image by fine DIR CNN. Especially, the color-coded DVF predicted by CoSF-Net overlaps the moving image, showing the variation degree of DVF. All the volumes are displayed in the transversal, coronal, and sagittal views. The display window is provided at a grayscale window of C = 0.4, W = 0.8.

phase (first column), whereas the fixed image is at the end-of-exhale (EOE) phase (second column). As observed in Fig. 4, considering the registration image pair with LR as the input, the warped volume by coarse DIR CNN (third column) is similar to the fixed image in the second column. However, misalignments still exist, such as that in the shape of the diaphragm (right arrows in the three visual views), which indicates that the DVF by coarse DIR CNN roughly depicts the contour changes of the registration pair but fails to estimate the detailed deformations. Using the SR model, the image quality of the generated image has improved; the artifacts are suppressed and small structures and sharp contour shapes are recovered, as compared with the original 4D-MR images obtained at the EOI phase. Lastly, by feeding the initial DVF using a coarse DIR CNN and an image pair with better resolution, the fine DIR CNN could predict a more accurate DVF (red arrows) and retain more detailed anatomic structures (yellow arrows) in the resulting images of both patients. In addition to the intermediate results, a moving image with a color-coded updated DVF is displayed in the first column to show the deformation changes, which are discussed in Section V.

B. Ablation Study

Figure 5 presents the results of patient III by removing different components in CoSF-Net, revealing that each component plays a complementary role in the final output. When using only a coarse DIR CNN, as shown in Fig. 5(d), an initial DVF guess is obtained, but noticeable DVF misalignment is observed at the diaphragm margin, indicated by the red arrow, when compared to the fixed image in 4D-MRI in Fig. 5(c). After removing both the SR model and the prior MRI, the remaining structure constitutes a classical coarse-to-fine architecture, referred to as coarse–fine in Fig. 5(e). The image quality by coarse-fine model remains unimproved, although it achieves better DVF estimation than the results in Fig. 5(d). Furthermore, without the guidance of prior MRI component, the fine DIR CNN architecture is simplified to a U-Net. As can be seen in Fig. 5(f), although the image quality improves to some extent, some detailed structures in the warped image appear unrealistic. Compared with Fig. 5(d) and (e), the image generated by the complete CoSF-Net demonstrates superior performance, suggesting that the SR model and prior MRI contribute to preserving the anatomic features and topology of real patients.

The quantitative analysis for the ablation studies, in terms of DVF estimation accuracy (EPE and EAE) and image similarity

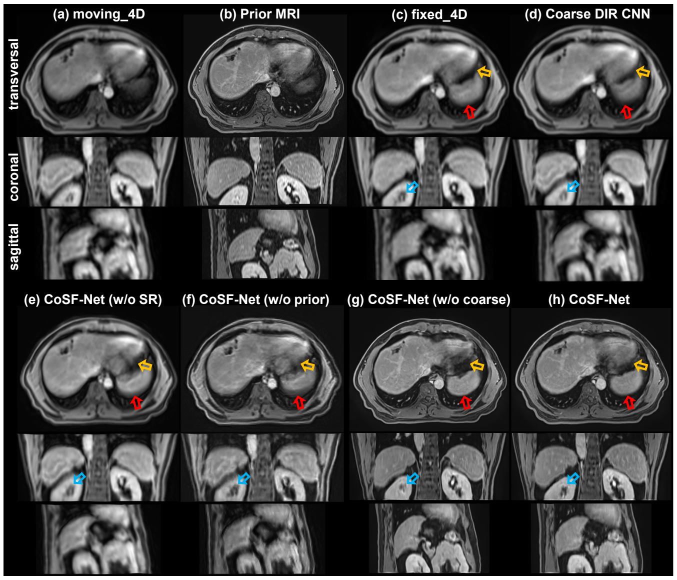

Fig. 5. Visual comparison of Patient III in the three views for investigating the contribution of the SR model, the prior MRI component, and the coarse DIR CNN in the architecture of CoSF-Net. From left to right: (a) presents the moving image in 4D-MRI and (b) is the corresponding prior image of (a); (c) the fixed image in 4D-MRI with low resolution; (d) the warped image by sole coarse DIR CNN; (e) the results by a classical coarse-to-fine model without an initial coarse DIR CNN; (f) the result by CoSF-Net without a prior MRI; (g) the result by CoSF-Net without an initial coarse DIR CNN; and (h) the result by the complete CoSF-Net. Three arrows with different colors point to the corresponding region reflecting respiratory motion (red arrow) and shape declination (yellow and blue arrows).

between the deformed image and its corresponding fixed image (PSNR and SSIM), is presented in Table II. All evaluation metrics are divided into two categories: with or without the SR module. The baseline method is pTV-based results through both LR image pair in 4D-MRI and predicted HR images, respectively. First, comparing the coarse DIR CNN and CoSF-Net (w/o SR) models, both of which do not involve the SR module, the latter achieves lower EPE and EAE values and higher PSNR and SSIM scores. These metrics demonstrate the enhanced accuracy of our proposed CoSF-Net (w/o SR) framework in estimating DVF compared to the coarse DIR alone. When incorporating the SR module, the complete CoSF-Net exhibits largely improved EPE and EAE and the highest PSNR and SSIM compared to CoSF-Net (w/o coarse) and CoSF-Net (w/o prior). It should be noted that EPE and EAE values of the full CoSF-Net are slightly higher than those of CoSF-Net (w/o SR), and the PSNR and SSIM values are similar. This difference can be attributed to the use of different DVFs at low and high resolution by pTV as a reference. Nonetheless, the overall performance improvement brought by the complete CoSF-Net highlights the importance of incorporating the SR model and other components in the proposed framework for generating high-quality 4D-MR images.

C. Comparison with Existing Methods

Figure 6 shows the qualitative results for Patient II using CoSF-Net and other comparison methods. Three slices profiled along the superior-inferior dimension in the same respiratory phase were chosen for visualization. In terms of the quantitative analysis, Fig. 7 lists the statistical analysis of the registration performance using different methods across all seven testing patient datasets. Different phase intervals (∼4) were also considered to test the robustness of the models.

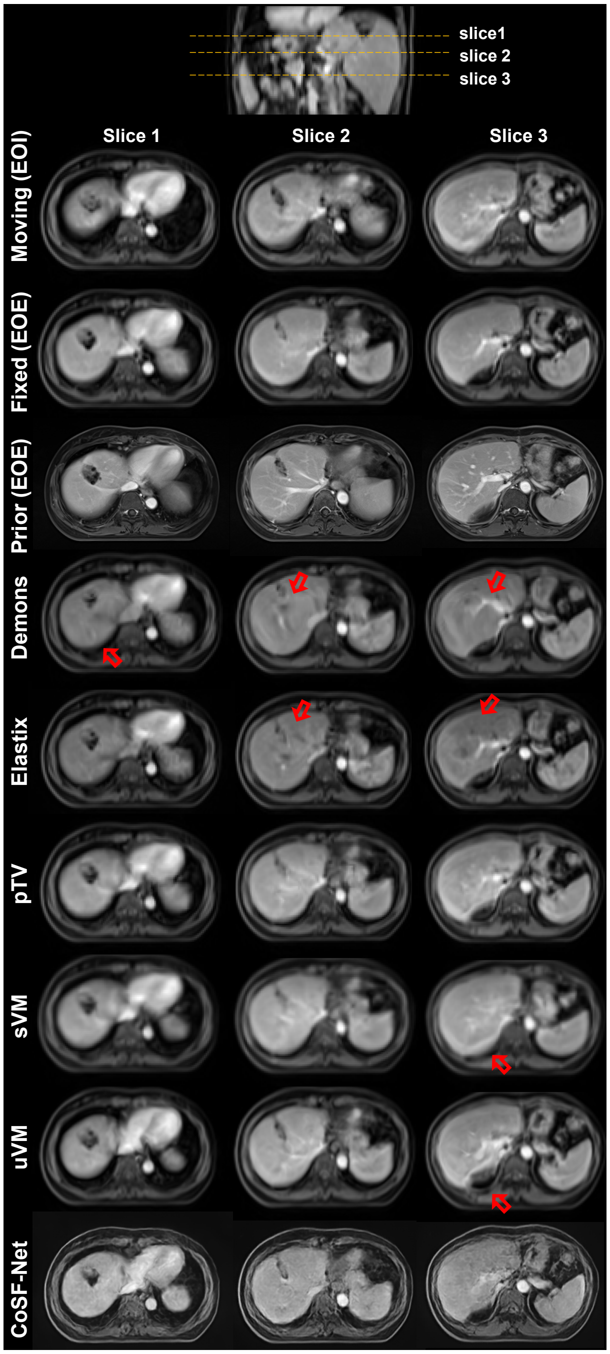

Fig. 6. Visual results of Patient I with the largest displacement from EOI to EOE. From top to bottom: the moving images in original 4D-MRI; the fixed image at EOE in original 4D-MRI; the prior reconstructed MRI image at EOE; deformation results obtained by warping the moving image by the predicted DVF using the following models: Demons, Elastix, pTV, supervised VM(sVM), coarse DIR CNN, and CoSF-Net. Three individual slices were selected in the transversal view for comparison. All the images are displayed in the window of level C = 0.4 and window W= 0.8.

As depicted in Fig. 6, the results obtained by Demons and Elastix display unsatisfactory deformation. Both methods

TABLE II
ABLATION STUDY OF EACH COMPONENT IN COSF-NET

model	M1	M2	M3		DVF		deformed image	
			U-Net	N-Net	EPE	EAE	PSNR	SSIM
coarse DIR CNN	✓	✗	✗	✗	2.4151 ± 1.2960	23.61° ± 17.72°	30.7457 ± 0.2352	0.9245 ± 0.0019
CoSF-Net (w/o SR)	✓	✗	✗	✓	0.4622 ± 0.1625	12.74° ± 7.82°	31.3286 ± 0.2578	0.9377 ± 0.0015
CoSF-Net (w/o coarse)	✗	✓	✗	✓	0.9393 ± 0.6495	13.54° ± 6.79°	29.5480 ± 0.3045	0.8721 ± 0.0024
CoSF-Net (w/o prior)	✓	✓	✓	✗	1.2456 ± 0.8795	16.78° ± 8.49°	26.8715 ± 0.2479	0.8016 ± 0.0018
CoSF-Net	✓	✓	✗	✓	0.6373 ± 0.2314	10.32° ± 6.97°	31.6831 ± 0.2608	0.9128 ± 0.0022

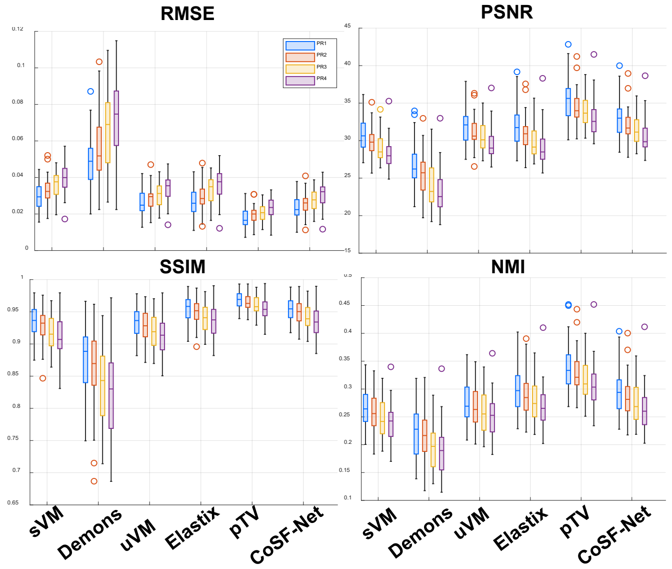

Fig. 7. Statistical analysis results for Patient I. Four evaluation metrics (RMSE, PSNR, SSIM, and NMI) were calculated in different phase intervals as represented by the blue, red, yellow, and purple box plots, respectively. Note that the moving image used in this experiment is up-sampled directly without using the SR model.

introduce inaccurate features or fail to maintain consistent intensities, as indicated by the red arrows, likely because both methods are sensitive to distorted image quality. The results using pTV maintain the correct physical topology and outperform those obtained using Demons and Elastix in terms of four metrics, as illustrated in Fig. 7. Regarding the DL-based DIR methods, the sole coarse DIR CNN is limited in estimating sufficiently accurate DVF due to its self-learning property, when compared to sVM. The bottom two rows display the deformed image from EOI to EOE using DVF predicted by CoSF-Net and the resultant image by CoSF-Net, respectively. Overall, CoSF-Net can predict an excellent 4D-MR image with accurate DVF estimation and high image quality.

In the quantitative evaluation, there are no ground truth images for the real patient dataset. To obtain a fair comparison, we used the fixed image (EOE phase) as the baseline for comparative methods. We also used prior MRIs (EOE phase) as the baseline for calculating the metrics for CoSF-Net. Our results demonstrate that the registration performance of CoSF-Net surpasses that of the other DL-based models and ranks second in all four quantitative metrics [Fig. 7]. The pTV algorithm exhibits the best performance, with its results being only slightly better than those of CoSF-Net. Furthermore, benefiting from a combination of DIR and SR, CoSF-Net achieves a considerable quality improvement in detailed feature recovery, whereas the other models fall short in this regard.

TABLE III
COMPARISON OF INFERENCE TIME FOR COSF-NET AND TRADITIONAL REGISTRATION ALGORITHMS

	pTV	Demons	Elastix	CoSF-Net
Low resolution	42.205079	7.557094	45.8407	\
High resolution	159.869495	125.285015	123.3749	2.2013

Table III also reports the inference time of different traditional registration algorithms for both low and high-resolution images. From the results, we can observe that Demons achieved the fastest prediction time of less than 10 seconds under LR images. By feeding a LR pair of 4D-MR images, CoSF-Net estimated DVFs and improved their image resolution simultaneously, with an inference time of 2.2013 seconds. This significantly outperforms the other traditional methods. These results demonstrate that CoSF-Net is not only efficient but also highly competitive in comparison to well-established image registration techniques.

D. Analysis of Tumor Localization and Detailed Structure Recovery Ability

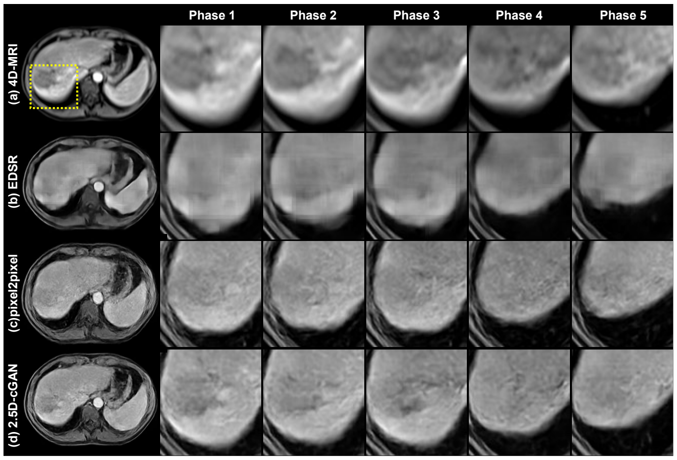

Fig. 8. Anatomical feature recovery results for Patient IV: (a) Original 4D-MRI as baseline, (b) EDSR, (c) pixel2pixel, and (d) 2.5D-cGAN. An ROI is selected in the transversal view (yellow rectangle) to represent the detailed features affected by respiratory motion. The display window is provided at a grayscale window of C = 0.4 and W = 0.7.

The accuracy of tumor positioning and the recoverability of detailed structures in MR images mainly depends on the performance of the SR network. To demonstrate the effectiveness of CoSF-Net, we selected two patient cases

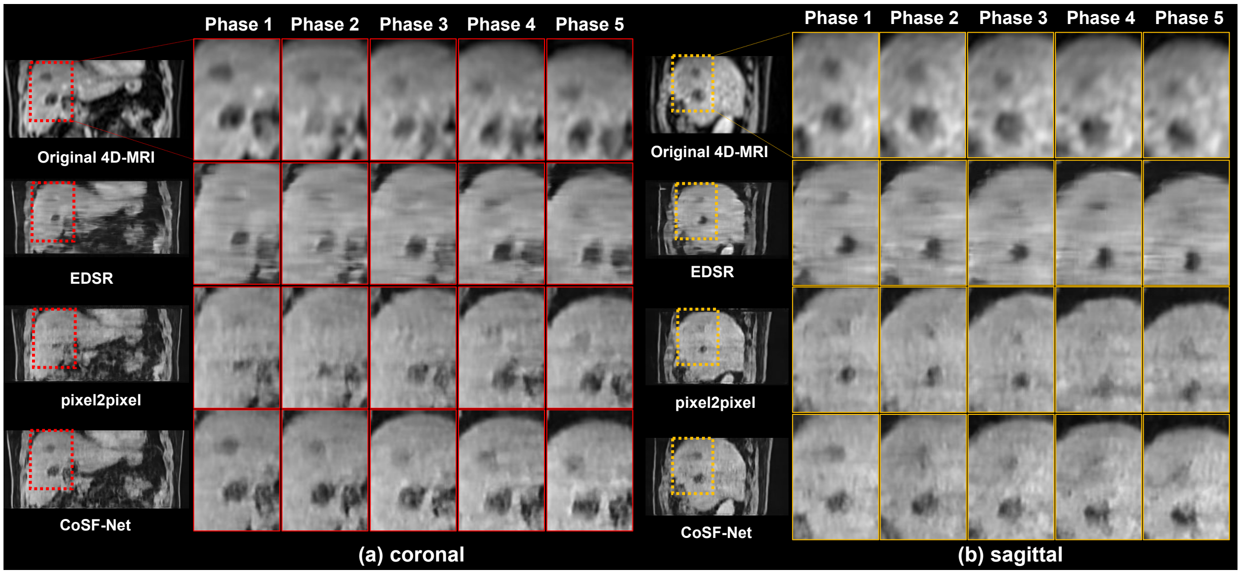

Fig. 9. Analysis of tumor localization and motion trajectory for patient I by 4D-MRI, EDSR, pixel2pixel, and CoSF-Net, respectively. To identify the variation of tumor location and shape, two ROIs were chosen in the coronal (red rectangle) and sagittal (yellow rectangle) views, with successive zoom-in images from the EOE (Phase 1) to EOI (Phase 5). The display window is C = 0.35 and W = 0.70.

(Patient IV in Fig. 8 and Patient I in Fig. 9) that have apparent tumors close to the diaphragm, which move with respiratory motion.

In Fig. 8, five consecutive phases of zoom-in region of interests (ROIs) at the transversal view (yellow rectangles) are shown to illustrate the detailed feature recovery performance and tumor localization ability of the proposed 2.5D-cGAN and its competitors. Since real-patient datasets lack an HR ground truth, we used the original 4D-MR images as the baseline for visual comparison. It can be observed that although the original 4D-MR images display the tumor's motion trajectory near the diaphragm, they suffer from limited image resolution and poor PSNR. Applying EDSR to 4D-MR images can effectively reduce the noise; however, the anatomical features could not be recovered, and over-smoothing is also observed because EDSR does not account for the non-uniqueness property of 4D-MRI. The results achieved by pixel2pixel show improved resolution compared to EDSR, but the contrast of the lesion's contour shape cannot be clearly detected. The bottom row in Fig. 8 by 2.5D-cGAN displays higher tumor contrast than the pixel2pixel method and restores tiny structures to the maximum possible extent, providing a satisfactory spatiotemporal resolution. In addition to the transversal view, Fig. 9 reports the coronal and sagittal views of five consecutive phases of Patient I to describe the movement of tumors affected by respiratory motion. Two enlarged ROIs covering the tumors and the diaphragm margin are selected with red and yellow rectangles, respectively. It is observed that the results of EDSR display poor resolution caused by intra-slice artifacts. This phenomenon may be due to the 2D-based image processing and inconsistent pixel value accuracy of EDSR. With the pixel2pixel process, intra-slice artifacts are eliminated, but the contrast of the two tumors is not recovered well compared to the reference image of the original 4D-MRI. Finally, the resulting ROIs obtained using CoSF-Net at various phases yield an improved contour depiction of the tumor and diaphragm and recover an accurate motion trajectory. By considering the correlation between different slices, CoSF-Net achieves better intra-slice resolution and ensures pixel consistency between neighboring pixels.

V. Discussions and Conclusion

In this study, we developed a unified DL framework to reconstruct a sequence of 4D-MR images with enhanced spatiotemporal resolution and image quality for application in radiotherapy. The proposed model enables simultaneous motion estimation and image resolution enhancement in 4D-MR using a cascade of three sub-models, including a coarse DIR CNN, an SR model, and a fine DIR CNN. Extensive experiments have demonstrated that the proposed CoSF-Net can predict accurate DVFs between the respiratory phases of 4D-MR images and effectively enhance the image resolution. Ultimately, the 4D-MRI images generated using CoSF-Net have a much higher spatial resolution and depict more detailed anatomical features than the original 4D-MR images. To the best of our knowledge, CoSF-Net is the first DL model that considers both motion estimation and image enhancement in 4D-MRI, and no such model for 4D-MRI exists thus far.

In CoSF-Net, a cascade of three sub-models, each of which is essential and indispensable, is well-conceived to comprehensively exploit the inherent image dynamics of 4D-MRI. We derive the following considerations concerning the design of CoSF-Net. First, the purpose of a coarse DIR CNN is to provide a warm start of DVF estimation, which helps obtain an initial alignment of the images while reducing the complexity of the problem. This initial alignment allows the subsequent fine DIR CNN to focus on refining the registration and correcting residual errors. In the ablation study, we conducted CoSF-Net (w/o coarse) to demonstrate the effectiveness of the coarse-SR-fine mechanism. As observed in Fig. 5(g), the complete CoSF-Net outperforms CoSF-Net (w/o coarse) in terms of

EPE, EAE, PSNR, and SSIM. This indicates that the coarse DIR CNN plays a significant role in improving the overall registration performance.

Second, we constructed a GAN-based SR model to address the challenges in 4D-MRI reconstruction, such as varying intensity distributions and breathing-related motion patterns [55]. Conventional CNNs with L-1 or L-2 norm-based loss functions may be inappropriate for this task. Benefiting from the mechanism of the discriminator module, GANs help learn the perceptual loss between image pairs, resulting in outputs more consistent with the actual perception of human vision. As presented in Fig. 8, it demonstrates that GAN is effective in restoring the detailed features of 4D-MRI. In our study, we chose a U-Net based discriminator architecture for its proven success in handling images and maintaining spatial information. This architecture is an extension of patchGAN, with convolutional layers from the original patchGAN in pixel2pixel serving as the encoder part. This design facilitates the integration of both global and local information into the output feature map, which shares the same resolution as the generated image. As a result, the U-Net based discriminator provides more detailed per-pixel feedback to the generator than a standard GAN discriminator. This design choice enables the model to generate high-quality images while effectively preserving fine structural details, thus contributing to improved registration performance in the overall CoSF-Net architecture. A comparative study between the U-Net discriminator and other architectures, such as patchGAN, could be a direction that motivates us to explore the potential improvements in the performance of our proposed method for future work.

Third, previous studies have demonstrated that incorporating prior images can improve the 4D-cone beam CT image quality in both classical algorithms and DL networks [54], [56], [57]. In this study, the prior MR image is used to facilitate DVF estimation. The prior MR image is integrated into both architecture design and back-propagation calculation, as described in Section II-B.3. Hence, CoSF-Net outperformed the other controlled settings when guided by prior MRI, as illustrated in Fig. 5 and Table II. In Section II-B.2, we also calculated the residual DVF to refine the updated DVF based on residual flow networks [30], rather than making a direct DVF prediction. To highlight the merit of residue DVF estimation by fine DIR CNN, the flow displacements of different stages are depicted in Fig. 10 using the motion color wheel. By estimating the residual DVF, the final DVF [Fig. 10(c)] is found to effectively extract more detailed information compared to the predicted DVF through coarse DIR CNN [Fig. 10(a)].

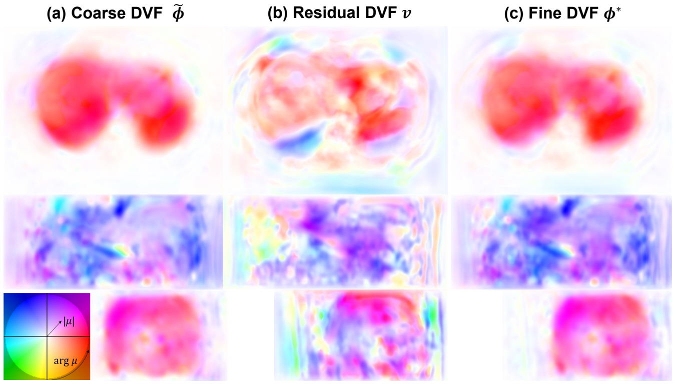

Fig. 10. DVF estimation in Patient I at different stages. (a) Coarse DVF estimated by coarse DIR CNN; (b) residual DVF by fine DIR CNN; (c) fine DVF by adding (a) and (b). The estimated DVF is represented by motion color wheel, in which the angle with the x-axis denotes the motion direction, while the color intensity expresses the displacement magnitude.

We also draw up the prospects for our study as follows. First, we aim to improve the through-plane resolution of enhanced 4D-MR images, as clinical MRI often exhibits lower through-plane resolution compared to in-plane resolution. In future studies, we plan to upgrade our SR model to achieve isotropic resolution, potentially by enlarging the effective GPU memory size and utilizing 3D or 3D+time volumes for training. In addition, a comprehensive evaluation of the trade-offs between model efficiency and memory requirements will be included. Second, given that MRI can perform functional imaging in addition to anatomical imaging, it serves as a powerful tool for treatment response assessment and outcome prediction. This encourages us to develop a contrast-variant 4D-MRI (5D) model based on the current 4D-MRI network. One possible approach involves extending the pair-wise registration to group-wise registration, perhaps by using long short-term memory (LSTM)-based architecture [58] to incorporate the temporal correlation property in network design. Third, several studies have explored the possibility of combining DL-based models (e.g., DIR and SR networks) with k-space data consistency penalties to develop unrolled deep reconstruction models. For example, DeepComplexMRI [59] has been developed for accelerated parallel MRI imaging, while SelfCoLearn model [60] and DIMENSION model [61] for conducting dynamic MR imaging via exploring the k-space and spatial space properties and enhancing the image details and textures. Besides, CINENet [62] integrates ADMM-based data consistency blocks with enhancement blocks. Additionally, modified DIR models have been proposed for both the image domain [63] and k-space data domain [64]. These advances motivate us to harness the potential of DL-based models to develop a comprehensive method that directly generates high spatiotemporal 4D-MRI series from k-space data.

Lastly, quantitatively evaluating the performance of a CNN model for real-world patient data poses an inherent challenge due to the absence of a reference or gold standard. In our study, this makes it difficult to directly assess DVF estimation accuracy and quality improvement of the generated images by CoSF-Net. As an alternative, we adopted DVFs and deformed images from the pTV method as a baseline for quantitative ablation experiments and comparisons with existing algorithms. As demonstrated in Fig. 5 and Fig. 6, CoSF-Net outperformed networks with ablated modules, classical CNNs, and existing optimization-based DIR algorithms, achieving performance comparable to pTV method. To address the difficulty in quantitative evaluation, we suggest three possible solutions for our future work. 1) Projection/k-space domain evaluation: By comparing the similarity metrics of the forward projection of the predicted image with its original projection/k-space data without reconstruction, we can quantitatively assess the method's effectiveness in preserving essential information. 2) Clinical evaluations with the assistance of experienced doctors: Evaluating the images in terms of diagnostic value, visual quality, and interpretability by clinicians or radiologists can provide a better understanding of the practical benefits offered by our proposed method in real-world clinical settings. 3) Physical phantom study: Conducting experiments using physical phantoms can serve as a valuable benchmark for evaluating the accuracy, precision, and robustness of the proposed method in a controlled environment.

In conclusion, we proposed an innovative DL model capable of simultaneous motion modeling and resolution enhancement for 4D-MRI. The proposed CoSF-Net integrates a GAN-based SR model into the coarse-to-fine registration model to present a coarse-SR-fine framework. We also upgraded the model by considering prior knowledge and limited 4D-MRI datasets into account. Our experimental results using real-patient datasets showed that CoSF-Net can handle motion estimation and image resolution enhancement in a unified model. Moreover, CoSF-Net has been shown to successfully recover 4D-MR images with a better spatiotemporal resolution than the state-of-the-art networks and algorithms.

References

[1] L. Xing, B. Thorndyke, E. Schreibmann, Y. Yang, T.-F. Li, G.-Y. Kim, G. Luxton, and A. Koong, "Overview of image-guided radiation therapy," *Medical Dosimetry*, vol. 31, no. 2, pp. 91–112, Jun. 2006.

[2] M. J. Menten, A. Wetscherek, and M. F. Fast, "MRI-guided lung SBRT: Present and future developments," *Physica Medica*, vol. 44, pp. 139–149, Dec. 2017.

[3] R. Otazo, P. Lambin, J.-P. Pignol, M. E. Ladd, H.-P. Schlemmer, M. Baumann, and H. Hricak, "MRI-guided radiation therapy: An emerging paradigm in adaptive radiation oncology," *Radiology*, vol. 298, no. 2, pp. 248–260, Feb. 2021.

[4] J. Cai, G. W. Miller, T. A. Altes, P. W. Read, S. H. Benedict, E. E. de Lange, G. D. Cates, J. R. Brookeman, J. P. Mugler, and K. Sheng, "Direct measurement of lung motion using hyperpolarized helium-3 MR tagging," *International Journal of Radiation Oncology, Biology, Physics*, vol. 68, no. 3, pp. 650–653, Jul. 2007.

[5] J. Cai, P. W. Read, J. M. Larner, D. R. Jones, S. H. Benedict, and K. Sheng, "Reproducibility of interfraction lung motion probability distribution function using dynamic MRI: Statistical analysis," *International Journal of Radiation Oncology, Biology, Physics*, vol. 72, no. 4, pp. 1228–1235, Nov. 2008.

[6] H. Ge, J. Cai, C. R. Kelsey, and F.-F. Yin, "Quantification and minimization of uncertainties of internal target volume for stereotactic body radiation therapy of lung cancer," *International Journal of Radiation Oncology, Biology, Physics*, vol. 85, no. 2, pp. 438–443, Feb. 2013.

[7] I. Vergalasova and J. Cai, "A modern review of the uncertainties in volumetric imaging of respiratory-induced target motion in lung radiotherapy," *Medical Physics*, vol. 47, no. 10, pp. e988–e1008, 2020.

[8] M. Fast, A. van de Schoot, T. van de Lindt, C. Carbaat, U. van der Heide, and J.-J. Sonke, "Tumor trailing for liver SBRT on the MR-Linac," *International Journal of Radiation Oncology, Biology, Physics*, vol. 103, no. 2, pp. 468–478, Feb. 2019.

[9] J. Cai, Z. Chang, Z. Wang, W. Paul Segars, and F.-F. Yin, "Four-dimensional magnetic resonance imaging (4D-MRI) using image-based respiratory surrogate: A feasibility study," *Medical Physics*, vol. 38, no. 12, pp. 6384–6394, Dec. 2011.

[10] Y. Liu, F.-F. Yin, B. G. Czito, M. R. Bashir, and J. Cai, "T2-weighted four dimensional magnetic resonance imaging with result-driven phase sorting," *Medical Physics*, vol. 42, no. 8, pp. 4460–4471, Aug. 2015.

[11] Y. Liu, X. Zhong, B. G. Czito, M. Palta, M. R. Bashir, B. M. Dale, F.-F. Yin, and J. Cai, "Four-dimensional diffusion-weighted MR imaging (4D-DWI): A feasibility study," *Medical Physics*, vol. 44, no. 2, pp. 397–406, Feb. 2017.

[12] T. Li, D. Cui, E. S. Hui, and J. Cai, "Time-resolved magnetic resonance fingerprinting for radiotherapy motion management," *Medical Physics*, vol. 47, no. 12, pp. 6286–6293, Dec. 2020.

[13] W. Harris, F.-F. Yin, J. Cai, and L. Ren, "Volumetric cine magnetic resonance imaging (VC-MRI) using motion modeling, free-form deformation and multi-slice undersampled 2D cine MRI reconstructed with spatio-temporal low-rank decomposition," *Quantitative Imaging in Medicine and Surgery*, vol. 10, no. 2, pp. 432–450, Feb. 2020.

[14] J. Yuan, O. L. Wong, Y. Zhou, K. Y. Chueng, and S. K. Yu, "A fast volumetric 4D-MRI with sub-second frame rate for abdominal motion monitoring and characterization in MRI-guided radiotherapy," *Quantitative Imaging in Medicine and Surgery*, vol. 9, no. 7, pp. 1303–1314, Jul. 2019.

[15] M. Frueh, A. Schilling, S. Gatidis, and T. Kuestner, "Real time landmark detection for within- and cross subject tracking with minimal human supervision," *IEEE Access*, vol. 10, pp. 81 192–81 202, 2022.

[16] C. Wang and F.-F. Yin, "4D-MRI in radiotherapy," in *Magnetic Resonance Imaging*. IntechOpen, Jul. 2019.

[17] J. N. Freedman, O. J. Gurney-Champion, S. Nill, A.-M. Shiarli, H. E. Bainbridge, H. C. Mandeville, D.-M. Koh, F. McDonald, M. Kachelrieß, U. Oelfke, and A. Wetscherek, "Rapid 4d-MRI reconstruction using a deep radial convolutional neural network: Dracula," *Radiotherapy and Oncology*, vol. 159, pp. 209–217, 2021.

[18] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," *arXiv:1707.02921 [cs]*, Jul. 2017.

[19] S. Park, H. M. Gach, S. Kim, S. J. Lee, and Y. Motai, "Autoencoder-inspired convolutional network-based super-resolution method in MRI," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 9, pp. 1–13, 2021.

[20] C. Zhao, B. E. Dewey, D. L. Pham, P. A. Calabresi, D. S. Reich, and J. L. Prince, "SMORE: A self-supervised anti-aliasing and super-resolution algorithm for MRI using deep learning," *IEEE Transactions on Medical Imaging*, vol. 40, no. 3, pp. 805–817, 2021.

[21] E. Ferdian, A. Suinesiaputra, D. J. Dubowitz, D. Zhao, A. Wang, B. Cowan, and A. A. Young, "4dflownet: Super-resolution 4d flow MRI using deep learning and computational fluid dynamics," *Frontiers in Physics*, vol. 8, 2020.

[22] A. Oar, G. Liney, R. Rai, S. Deshpande, L. Pan, M. Johnston, M. Jameson, S. Kumar, and M. Lee, "Comparison of four dimensional computed tomography and magnetic resonance imaging in abdominal radiotherapy planning," *Physics and Imaging in Radiation Oncology*, vol. 7, pp. 70–75, 2018.

[23] W. Huang, H. Yang, X. Liu, C. Li, I. Zhang, R. Wang, H. Zheng, and S. Wang, "A coarse-to-fine deformable transformation framework for unsupervised multi-contrast MR image registration with dual consistency constraint," *IEEE Transactions on Medical Imaging*, vol. 40, no. 10, pp. 2589–2599, 2021.

[24] M. A. Morales, D. Izquierdo-Garcia, I. Aganj, J. Kalpathy-Cramer, B. R. Rosen, and C. Catana, "Implementation and validation of a three-dimensional cardiac motion estimation network," *Radiology: Artificial Intelligence*, vol. 1, no. 4, p. e180080, 2019.

[25] E. Martín-González, T. Sevilla, A. Revilla-Orodea, P. Casaseca-de-la Higuera, and C. Alberola-López, "Groupwise non-rigid registration with deep learning: An affordable solution applied to 2d cardiac cine MRI reconstruction," *Entropy (Basel, Switzerland)*, vol. 22, no. 6, p. 687, 2020.

[26] J. Pan, D. Rueckert, T. Küstner, and K. Hammernik, "Efficient image registration network for non-rigid cardiac motion estimation," in *Machine Learning for Medical Image Reconstruction*, ser. Lecture Notes in Computer Science. Springer International Publishing, 2021, pp. 14–24.

[27] N. R. F. Huttinga, C. A. T. van den Berg, P. R. Luijten, and A. Sbrizzi, "MR-MOTUS: model-based non-rigid motion estimation for MR-guided radiotherapy using a reference image and minimal k-space data," *Physics in Medicine and Biology*, vol. 65, no. 1, p. 015004, 2020.

[28] T. Küstner, J. Pan, H. Qi, G. Cruz, C. Gilliam, T. Blu, B. Yang, S. Gatidis, R. Botnar, and C. Prieto, "LAPNet: Non-rigid registration derived in k-space for magnetic resonance imaging," *IEEE Transactions on Medical Imaging*, vol. 40, no. 12, pp. 3686–3697, 2021.

[29] H. Xiao, R. Ni, S. Zhi, W. Li, C. Liu, G. Ren, X. Teng, W. Liu, W. Wang, Y. Zhang, H. Wu, H.-F. V. Lee, L.-Y. A. Cheung, H.-C. C. Chang, T. Li, and J. Cai, "A dual-supervised deformation estimation model (DDEM) for constructing ultra-quality 4D-MRI based on a commercial low-quality 4D-MRI for liver cancer radiation therapy," *Medical Physics*, vol. 49, no. 5, pp. 3159–3170, Feb. 2022.

[30] A. Ranjan and M. J. Black, "Optical flow estimation using a spatial pyramid network," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 2720–2729, ISSN: 1063-6919.

[31] N. Gunnarsson, J. Sjölund, and T. B. Schön, "Learning a deformable registration pyramid," in *Segmentation, Classification, and Registration*

of Multi-modality Medical Imaging Data, ser. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2021, pp. 80–86.

[32] J. Lv, Z. Wang, H. Shi, H. Zhang, S. Wang, Y. Wang, and Q. Li, "Joint progressive and coarse-to-fine registration of brain MRI via deformation field integration and non-rigid feature fusion," *IEEE Transactions on Medical Imaging*, vol. 41, no. 10, pp. 2788–2802, 2022.

[33] T. N. van de Lindt, M. F. Fast, W. van den Wollenberg, J. Kaas, A. Betgen, M. E. Nowee, E. P. Jansen, C. Schneider, U. A. van der Heide, and J.-J. Sonke, "Validation of a 4d-MRI guided liver stereotactic body radiation therapy strategy for implementation on the MR-linac," *Physics in Medicine and Biology*, vol. 66, no. 10, 2021.

[34] G. Li, J. Wei, M. Kadbi, J. Moody, A. Sun, S. Zhang, S. Markova, K. Zakian, M. Hunt, and J. O. Deasy, "Novel super-resolution approach to time-resolved volumetric 4-Dimensional magnetic resonance imaging with high spatiotemporal resolution for multi-breathing cycle motion assessment," *International Journal of Radiation Oncology, Biology, Physics*, vol. 98, no. 2, pp. 454–462, Jun. 2017.

[35] F. Odille, A. Bustin, B. Chen, P.-A. Vuissoz, and J. Felblinger, "Motion-corrected, super-resolution reconstruction for high-resolution 3d cardiac cine MRI," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, ser. Lecture Notes in Computer Science. Springer International Publishing, 2015, pp. 435–442.

[36] S. Wang, C. Qin, N. Savioli, C. Chen, D. P. O'Regan, S. Cook, Y. Guo, D. Rueckert, and W. Bai, "Joint motion correction and super resolution for cardiac segmentation via latent optimisation," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part III*. Springer-Verlag, 2021, pp. 14–24.

[37] Q. Beirinckx, B. Jeurissen, M. Nicastro, D. H. J. Poot, M. Verhoye, A. J. d. Dekker, and J. Sijbers, "Model-based super-resolution reconstruction with joint motion estimation for improved quantitative MRI parameter mapping," *Computerized Medical Imaging and Graphics*, vol. 100, p. 102071, 2022.

[38] D. B. Pier, A. Gholipour, O. Afacan, C. Velasco-Annis, S. Clancy, K. Kapur, J. A. Estroff, and S. K. Warfield, "3d super-resolution motion-corrected MRI: Validation of fetal posterior fossa measurements," *Journal of neuroimaging : official journal of the American Society of Neuroimaging*, vol. 26, no. 5, pp. 539–544, 2016.

[39] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca, "VoxelMorph: A learning framework for deformable medical image registration," *IEEE Transactions on Medical Imaging*, vol. 38, no. 8, pp. 1788–1800, Aug. 2019.

[40] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in Neural Information Processing Systems*, vol. 27. Curran Associates, Inc., 2014.

[41] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, ser. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2015, pp. 234–241.

[42] M. Jaderberg, K. Simonyan, A. Zisserman, and K. Kavukcuoglu, "Spatial transformer networks," in *Advances in Neural Information Processing Systems*, vol. 28. Curran Associates, Inc., 2015.

[43] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," *arXiv:1611.07004 [cs]*, 2018. [Online]. Available: http://arxiv.org/abs/1611.07004

[44] E. Schonfeld, B. Schiele, and A. Khoreva, "A u-net based discriminator for generative adversarial networks," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2020, pp. 8204–8213.

[45] Z. Huang, J. Zhang, Y. Zhang, and H. Shan, "DU-GAN: Generative adversarial networks with dual-domain u-net based discriminators for low-dose CT denoising," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–12, 2022.

[46] P. Kancharla and S. S. Channappayya, "Improving the visual quality of generative adversarial network (GAN)-generated images using the multi-scale structural similarity index," in *25th IEEE International Conference on Image Processing (ICIP)*, 2018, pp. 3908–3912.

[47] Z. Wang, E. Simoncelli, and A. Bovik, "Multiscale structural similarity for image quality assessment," in *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers*, vol. 2, 2003, pp. 1398–1402 Vol.2.

[48] Y. Liu, F.-F. Yin, Z. Chang, B. G. Czito, M. Palta, M. R. Bashir, Y. Qin, and J. Cai, "Investigation of sagittal image acquisition for 4D-MRI with body area as respiratory surrogate," *Medical Physics*, vol. 41, no. 10, p. 101902, 2014.

[49] S. Klein, M. Staring, K. Murphy, M. A. Viergever, and J. P. W. Pluim, "Elastix: A toolbox for intensity-based medical image registration," *IEEE Transactions on Medical Imaging*, vol. 29, no. 1, pp. 196–205, Jan. 2010.

[50] C. Zhang, Y. Li, and G.-H. Chen, "Accurate and robust sparse-view angle CT image reconstruction using deep learning and prior image constrained compressed sensing (DL-PICCS)," *Medical Physics*, vol. 48, no. 10, pp. 5765–5781, 2021.

[51] V. Vishnevskiy, T. Gass, G. Szekely, C. Tanner, and O. Goksel, "Isotropic total variation regularization of displacements in parametric image registration," *IEEE Transactions on Medical Imaging*, vol. 36, no. 2, pp. 385–395, Feb. 2017.

[52] J. P. Thirion, "Image matching as a diffusion process: An analogy with Maxwell's demons," *Medical Image Analysis*, vol. 2, no. 3, pp. 243–260, Sep. 1998.

[53] A. Hore and D. Ziou, "Image quality metrics: PSNR vs. SSIM," in *20th International Conference on Pattern Recognition*. IEEE, Oct. 2010, pp. 2366–2369.

[54] S. Zhi, M. Kachelrieß, and X. Mou, "High-quality initial image-guided 4D CBCT reconstruction," *Medical Physics*, vol. 47, no. 5, pp. 2099–2115, 2020.

[55] B. Huang, H. Xiao, W. Liu, Y. Zhang, H. Wu, W. Wang, Y. Yang, Y. Yang, G. W. Miller, T. Li, and J. Cai, "MRI super-resolution via realistic downsampling with adversarial learning," *Physics in Medicine Biology*, vol. 66, no. 20, p. 205004, 2021.

[56] S. Zhi, M. Kachelrieß, F. Pan, and X. Mou, "CycN-Net: A convolutional neural network specialized for 4D CBCT images refinement," *IEEE Transactions on Medical Imaging*, vol. 40, no. 11, pp. 3054–3064, Nov. 2021.

[57] S. Zhi, M. Kachelrieß, and X. Mou, "Spatiotemporal structure-aware dictionary learning-based 4D CBCT reconstruction," *Medical Physics*, vol. 48, no. 10, pp. 6421–6436, 2021.

[58] A. Ammar, O. Bouattane, and M. Youssfi, "Automatic spatio-temporal deep learning-based approach for cardiac cine MRI segmentation," in *Networking, Intelligent Systems and Security*, vol. 237. Singapore: Springer Singapore, 2022, pp. 59–73.

[59] S. Wang, H. Cheng, L. Ying, T. Xiao, Z. Ke, H. Zheng, and D. Liang, "DeepcomplexMRI: Exploiting deep residual network for fast parallel MR imaging with complex convolution," *Magnetic Resonance Imaging*, vol. 68, pp. 136–147, 2020.

[60] J. Zou, C. Li, S. Jia, R. Wu, T. Pei, H. Zheng, and S. Wang, "SelfCoLearn: Self-Supervised Collaborative Learning for Accelerating Dynamic MR Imaging," vol. 9, no. 11, p. 650.

[61] S. Wang, Z. Ke, H. Cheng, S. Jia, L. Ying, H. Zheng, and D. Liang, "DIMENSION: Dynamic MR imaging with both k-space and spatial prior knowledge obtained via multi-supervised network training," vol. 35, no. 4, p. e4131.

[62] T. Küstner, N. Fuin, K. Hammernik, A. Bustin, H. Qi, R. Hajhosseiny, P. G. Masci, R. Neji, D. Rueckert, R. M. Botnar, and C. Prieto, "CINENet: deep learning-based 3d cardiac CINE MRI reconstruction with multi-coil complex-valued 4d spatio-temporal convolutions," *Scientific Reports*, vol. 10, no. 1, p. 13710, 2020.

[63] J. Yang, T. Küstner, P. Hu, P. Liò, and H. Qi, "End-to-end deep learning of non-rigid groupwise registration and reconstruction of dynamic MRI," *Frontiers in Cardiovascular Medicine*, vol. 9, p. 880186, 2022.

[64] T. Küstner, J. Pan, C. Gilliam, H. Qi, G. Cruz, K. Hammernik, T. Blu, D. Rueckert, R. Botnar, C. Prieto, and S. Gatidis, "Self-supervised motion-corrected image reconstruction network for 4d magnetic resonance imaging of the body trunk," *APSIPA Transactions on Signal and Information Processing*, vol. 11, no. 1, 2022.