Received 26 June 2024, accepted 10 July 2024, date of publication 15 July 2024, date of current version 25 July 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3427674

RESEARCH ARTICLE

DDoS-UNet: Incorporating Temporal Information Using Dynamic Dual-Channel UNet for Enhancing Super-Resolution of Dynamic MRI

SOUMICK CHATTERJEE[1,2,3], (Member, IEEE), CHOMPUNUCH SARASAEN[2,4,5], GEORG ROSE[4,5], ANDREAS NÜRNBERGER[1,6], (Member, IEEE), AND OLIVER SPECK[2,4,6,7]

[1]Data and Knowledge Engineering Group, Faculty of Computer Science, Otto von Guericke University Magdeburg, 39106 Magdeburg, Germany
[2]Biomedical Magnetic Resonance, Otto von Guericke University Magdeburg, 39106 Magdeburg, Germany
[3]Genomics Research Centre, Human Technopole, 20157 Milan, Italy
[4]Research Campus STIMULATE, Otto von Guericke University Magdeburg, 39106 Magdeburg, Germany
[5]Institute for Medical Engineering, Otto von Guericke University Magdeburg, 39106 Magdeburg, Germany
[6]Centre for Behavioural Brain Sciences, 39106 Magdeburg, Germany
[7]German Centre for Neurodegenerative Diseases, 39120 Magdeburg, Germany

Corresponding author: Soumick Chatterjee (soumick@ieee.org)

This work was supported by the International Graduate School MEMoRIAL at Otto von Guericke University (OVGU) Magdeburg, Germany, through European Structural and Investment Funds (ESF) under the Program ‘‘Sachsen-Anhalt WISSENSCHAFT Internationalisierung’’ under Project ZS/2016/08/80646.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Local Ethics Committee of Otto von Guericke University Magdeburg, Saxony-Anhalt, Germany.

ABSTRACT Magnetic resonance imaging (MRI) provides high spatial resolution and excellent soft-tissue contrast without using harmful ionising radiation. Dynamic MRI is an essential tool for interventions to visualise movements or changes of the target organ. However, such MRI acquisitions with high temporal resolution suffer from limited spatial resolution - also known as the spatio-temporal trade-off of dynamic MRI. Several approaches, including deep learning based super-resolution approaches by treating each timepoint as individual volumes. This research addresses this issue by creating a deep learning model which attempts to learn both spatial and temporal relationships. A modified 3D UNet model, DDoS-UNet, is proposed - which takes the low-resolution volume of the current timepoint along with a prior image volume. Initially, the network is supplied with a static high-resolution planning scan as the prior image along with the low-resolution input to super-resolve the first timepoint. Then it continues step-wise by using the super-resolved timepoints as the prior image while super-resolving the subsequent timepoints. The model performance was tested with 3D dynamic data that was undersampled to different in-plane levels and achieved an average SSIM value of 0.951±0.017 while reconstructing only 4% of the k-space - which could result in a theoretical acceleration factor of 25. The proposed approach can be used to reduce the required scan-time while achieving high spatial resolution - consequently alleviating the spatio-temporal trade-off of dynamic MRI, by incorporating prior knowledge of spatio-temporal information from the available high-resolution planning scan and the existing temporal redundancy of time-series images into the network model.

INDEX TERMS MRI reconstruction, undersampled MRI, dynamic MRI, super-resolution, dual-channel training, deep learning.

I. INTRODUCTION

Magnetic resonance imaging (MRI) does not rely on ionising radiation and can provide high spatial resolution with superior

The associate editor coordinating the review of this manuscript and approving it for publication was Chulhong Kim.

visualisation of soft-tissue contrast. MR images can also offer better differentiation between fat, water and muscle than other imaging modalities. Therefore, image guidance based on MRI is a favourable tool for identifying and characterising tumours in interventions [1], [2]. Interventional applications in real-time or near real-time, such as MR-guided liver biopsy, show excellent contrast between the target organ or structure and adjacent soft tissue while visualising the changes of internal organs during an examination. In such applications, dynamic MRI is used, which is obtained by acquiring the k-space data (in frequency domain) continuously and reconstructing a sequence of images over time [3]. However, while achieving high temporal resolution, these acquisitions suffer from the restricted spatial resolution because only a limited part of the data can be measured (undersampling). Consequently, the resultant image might have reconstruction artefacts due to the violation of the Nyquist criterion [4], and also leads to image resolution loss. This is known as the spatio-temporal trade-off of dynamic MRI and has been demonstrated as one of the main research problems [5], [6], [7], [8]. Although common approaches such as compressed sensing [6] can utilise the spatial and temporal correlation of the data to accelerate the data acquisition, the iterative processes could hinder real-time applications such as intervention MRI.

Super-resolution (SR) is a process of estimating a high-resolution image from a low-resolution counterpart. Several deep learning based super-resolution algorithms have been proposed [9], [10], [11], [12], [13]. The existing SR techniques can be categorised into two major groups: single image super-resolution (SISR) and video super-resolution (VSR). In contrast to SISR, VSR exploits the temporal information in a sequence of images to enhance the spatial resolution and frame rate [14], [15], [16]. Additionally, some literature investigated the use of temporal information incorporation and reported its potential for improving the image quality of dynamic MRI reconstruction [17], [18], [19].

To further improve the super-resolved image quality, additional prior information had been integrated into the super-resolution process [20], [21]. The prior information can be incorporated in multi-channel training to enhance the results [22]. A multi-channel network allows better feature extractions when learning with multiple types of channels [23]. Multi-channel training has been used across numerous applications including image recognition [24], speech recognition [25], [26], audio classification [27], natural language processing [28], etc. This paper extends the previous work into the temporal domain [29] by exploiting dual-channel inputs (prior image and low-resolution image) in the deep learning model - to learn the temporal relationship between timepoints, while also learning the spatial relationship between low- and high-resolution images, to perform SISR, using the proposed DDoS (**D**ynamic **D**ual-channel **o**f **S**uper-resolution) approach.

A. RELATED WORK

The UNet architecture [30], including its 3D version [31], is a versatile neural network consisting of two paths: contraction and expansion. Originally proposed for image segmentation, different flavours of UNet have been developed and deployed in plenty of applications such as image segmentation [32], [33], [34], [35], audio source separation [36], [37], [38] and image reconstruction [39], [40]. 3D UNet and its variants have been used for MR super-resolution as well [29], [41], [42]. Furthermore, UNet has been extended to multi-channel and dual-branch to incorporate prior information [22].

Previous work attempting to super-resolve 3D dynamic MRIs treats each timepoint as a single 3D volume and then super-resolves them individually [29]. But in this way, the inherent relationship between the different timepoints of the dynamic MRI is not utilised, which might be possible to exploit to improve the super-resolution performance. Dynamic MRIs can be considered as 3D videos. For super-resolving 2D videos, recurrent networks are commonly employed, which utilise the aforementioned relationship [43], [44], [45]. However, these types of networks are typically more computationally expensive during training - making them difficult to employ for a 3D volumetric scenario. It is worth mentioning that super-resolution of dynamic medical imaging is not limited to MRI, it is also applicable to modalities like endoscopy [46], [47].

Since medical images are mainly used for diagnosis, evaluation using perception-based metrics are more suitable than pixel-wise metrics. Perceptual loss [48] has demonstrated the ability to improve image quality perceptually, yielding superior results and reducing blurriness than classical pixel-based metrics such as L1 or L2 [49], [50]. A recent study from [51] presented that deep feature extractions, which were obtained from the trained network, could be utilised to deal with excessive blurry images and showed that perceptual similarity is an important property that has been shared among deep visual representations. Previous work [29] has also demonstrated the potential of applying a perceptual loss network to improve the results of image super-resolution.

B. CONTRIBUTIONS

This paper extends the research of Single-Image Super-Resolution (SISR) of dynamic MRIs treating each timepoint as individual 3D volumes, by incorporating the temporal information into the network model using the proposed **DDoS-UNet** framework. The proposed method super-resolves the low-resolution dynamic MRI with the help of a static prior scan, and by exploiting the temporal relationship between the different timepoints. The method has been evaluated using Cartesian undersampling by taking different amounts of the centre k-space data, up to a theoretical acceleration factor of 25.

II. METHODOLOGY

In this work, the dynamic training data was initially generated from the benchmark dataset due to the lack of dynamic abdominal data. After that, it was undersampled, and a modified UNet model was trained on that. The dual-channel input consists of the low-resolution image of the current timepoint and the super-resolved image of the previous timepoint. The network was trained and tested with different levels of undersampling.

A. SUPER-RESOLUTION RECONSTRUCTION

The reconstruction of the high-resolution image from the corresponding low-resolution image can be modelled as:

$$\hat{I}_{HR} = \mathcal{F}(I_{LR}; \theta) \tag{1}$$

where I_{LR} is the low-resolution image, $\hat{I}_{HR}$ is the super-resolved image, $\mathcal{F}$ is the mapping function which models the spatial super-resolution relationship between the corresponding low- and high-resolution images using a given set of parameters θ [52]. The SR image reconstruction is an ill-posed problem to approximate the super-resolved image from a given low-resolution counterpart. The SR reconstruction using neural networks can be expressed as an objective function:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \sum_{n=1}^{N} \mathcal{L}(\mathcal{F}(\mathbf{x}_n; \boldsymbol{\theta}), \mathbf{y}_n) + \mathcal{R}(\boldsymbol{\theta}) \tag{2}$$

where $\mathbf{x}_n$ is a low-resolution input sample I_{LR} provided to the network from a training set consisting of N items, $\mathbf{y}_n$ is the corresponding high-resolution ground-truth I_{HR}, $\mathcal{F}$ is the neural network as the mapping function (Eq. 1), the operator $\mathcal{L}$ defines the loss function between the predicted super-resolved image $\hat{I}_{HR}$ and the corresponding ground-truth I_{HR}, and $R(\theta)$ is a regularisation term.

B. NETWORK ARCHITECTURE

The 3D UNet architecture from the previous work [29] was extended using multi-channel for supplying prior information [22] to create the proposed **D**ynamic **D**ual-channel **o**f **S**uper-resolution UNet architecture (DDoS-UNet, or simply DDoS), as shown in Fig. 1. The basic architecture of the UNet is similar to the previous work [29] - except for two differences, having contracting (encoding) and expanding (decoding) paths. The contracting path is made of three blocks, each of the blocks comprises two pairs of 3D convolutional layers (kernel size:3, stride:1, padding:1) and ReLU activation functions, followed by average pool layers (kernel size: 2) - making the output size of the block half the size of the input received by that block. The expanding path also consists of three blocks, each consisting of a pair of trilinear upsampling layer (scale factor:2) and 3D convolutional layer (kernel size:1, stride:1, padding:0), unlike the original work which used 3D convolutional transpose layers (first difference with the earlier model); followed by a convolutional block similar to the contracting path, except for the pooling layers. It is noteworthy that initial experiments were performed using 3D convolutional transpose layers similar to the earlier model, but for volumetric super-resolution this model resulted in checkerboard artefacts [53]. This can be attributed to the fact that overlapped portions of the patches are averaged in the patch-based super-resolution - mitigating the checkerboard problem, but in volumetric super-resolution, there is no averaging operation that could mitigate this effect. Each block of the expanding path increases the size of its input by a factor of two. Inside these expanding path blocks, after upsampling the input using trilinear-convolution pair, the output is concatenated with the input coming from a similar depth of the contraction path - known as skip connections. The initial layer of the network provides an output of 64 feature maps. Then, each block of the contraction path increases the number of feature maps by two, whereas each of the expanding path blocks decreases it by two. Finally, a 3D convolutional layer (kernel size: 1, stride: 1, padding: 0) is applied to merge all the feature maps to generate the final output. The other difference between the earlier UNet [29] and this DDoS-UNet is the fact that the initial layer of the network receives two input channels rather than one.

Since the UNet-like architectures requires the image dimensions of the input to be the same as the output (ground-truth), the low-resolution input volumes were interpolated using trilinear interpolation with the interpolation factor equivalent to the acceleration factor before providing them as input to the DDoS-UNet model.

1) DDOS: WORKING MECHANISM AND THEORY

The DDoS-UNet works with dynamic MRIs while using the static planning scan as a prior image. Initially, the network is supplied with a patient-specific fully sampled high-resolution (HR) static prior scan on the first channel and the first timepoint (TP0) of the undersampled low-resolution (LR) dynamic MRI on the second channel. It is to be noted that the static planning scan is acquired with the same protocol as the dynamic scan, but they are not co-registered. Given this pair of HR-LR images, DDoS-UNet super-resolves the LR to obtain the TP0 of the super-resolved (SR) HR dynamic MRI. This initial phase is termed here as the "Antipasto" phase as it precedes the main reconstruction phase. The reconstruction phase starts by supplying this SR-TP0 on the first channel, while the LR-TP1 is supplied on the second channel of the network to generate SR-TP1. This process is continued recursively for all the subsequent timepoints. This can be formulated by modifying Eq. 1 as:

$$\hat{\mathbf{y}}_t = \mathcal{F}(\mathbf{x}_t, \hat{\mathbf{y}}_{t-1}; \boldsymbol{\theta}) \tag{3}$$

where $\hat{\mathbf{y}}_{n,t}$ is the super-resolved timepoint, $\mathbf{x}_{n,t}$ is the low-resolution timepoint, $\hat{\mathbf{y}}_{n,t-1}$ is the super-resolved previous timepoint, $\mathcal{F}$ is the super-resolution model that maps those three images, and θ is the set of parameters of $\mathcal{F}$. The network training process of the DDoS-UNet can be expressed by

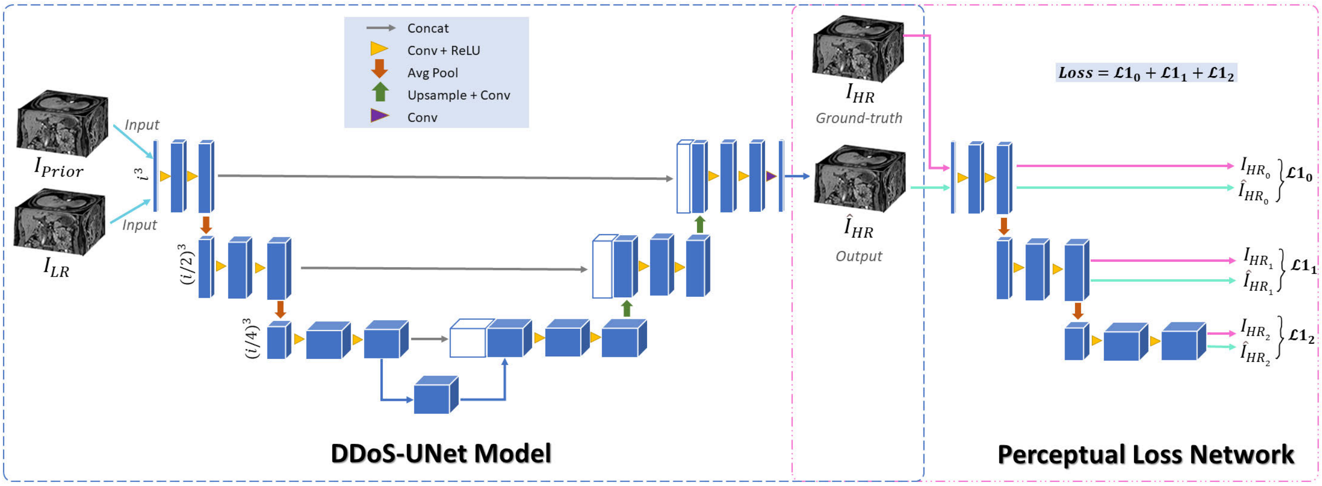


FIGURE 1. DDoS-UNet: Network architecture and training phase. I_{prior} is the high-resolution prior image (super-resolved or high-resolution previous timepoint), I_{LR} is the low-resolution current timepoint, I_{HR} is the high-resolution ground-truth of the current timepoint (only during the training phase), and $\hat{I}_{HR}$ is the super-resolved current timepoint output from the model. The loss between I_{HR} and $\hat{I}_{HR}$ is calculated using a perceptual loss network, which is then optimised to train the model.

modifying Eq. 2 as:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \left[\left(\sum_{s=1}^{S} \left(\mathcal{L}(\mathcal{F}(\mathbf{x}_{s,1}, \mathbf{y}_{s,0}; \boldsymbol{\theta}), \mathbf{y}_{s,1}) \right) \right. \right. \\ \left. + \left(\sum_{t=2}^{T_s} \mathcal{L}(\mathcal{F}(\mathbf{x}_{s,t}, \hat{\mathbf{y}}_{s,t-1}; \boldsymbol{\theta}), \mathbf{y}_{s,t}) \right) + \mathcal{R}(\boldsymbol{\theta}) \right] \tag{4}$$

where $\mathbf{x}_{s,t}$ is the low-resolution input volume at timepoint t of subject s, $\mathbf{y}_{s,t}$ is the corresponding high-resolution ground-truth, $\hat{\mathbf{y}}_{s,t-1}$ is the previous super-resolved timepoint, $\mathcal{F}$ is the neural network as the mapping function (Eq. 3), the operator $\mathcal{L}$ defines the loss function between the predicted super-resolved image and the corresponding ground-truth, $R(\theta)$ is a regularisation term, when T_s is the number of timepoint subject s has and S is the number of subjects present in the training dataset. Here, $\mathcal{F}(\mathbf{x}_{s,1}, \mathbf{y}_{s,0}; \boldsymbol{\theta})$ is the Antipasto phase where $\mathbf{x}_{s,1}$ is the low-resolution volume at the first dynamic timepoint and $\mathbf{y}_{s,0}$ is the high-resolution static volume, while the rest is the actual reconstruction phase.

The authors hypothesise that the network learns two different representations: the temporal relationship between $\hat{\mathbf{y}}_t$ and $\hat{\mathbf{y}}_{t-1}$ and the super-resolution relationship between $\mathbf{x}_t$ and $\hat{\mathbf{y}}_t$. If Ψ is the DDoS relationship and $\hat{\boldsymbol{\theta}}$ is the set of parameters of the DDoS network learnt using the Eq. 4, this hypothesis can be formulated as:

$$\Psi(\hat{\boldsymbol{\theta}}) \ni \{\mathcal{R}_1(\mathbf{x}_t, \hat{\mathbf{y}}_t), \mathcal{R}_2(\hat{\mathbf{y}}_{t-1}, \hat{\mathbf{y}}_t)\} \tag{5}$$

where $\mathcal{R}_1$ is the super-resolution relationship and $\mathcal{R}_2$ is the temporal relationship.

It is worth mentioning that the patch-based super-resolution idea from the previous work [29] was dropped in this current research due to the working theory of DDoS-UNet. Due to physiological movements, the organs can move in and out of the 24^3 patches (as used in the previous work). Consequently, the supplied $\mathbf{x}_t$ and $\hat{\mathbf{y}}_{t-1}$ patches might not contain similar organs - making the hypothesis of the temporal relationship operator $\mathcal{R}_2$ of Eq. 5 invalid. Hence, this work performs volumetric super-resolution (using complete 3D volumes) instead of 3D patch-based super-resolution.

C. DATA

The proposed method was trained using the publicly available abdominal benchmark dataset: the CHAOS dataset (T1-dual images, in- and opposed phase) [54], comprising 80 volumes (40 subjects, in-phase and opposed-phase for each subject). Dynamic training data was generated artificially by applying random elastic deformation, explained in detail in Sec. II-C1. The dataset was divided into training and validation sets with a ratio of 70:30.

For testing the approach, high-resolution 3D static (breath-hold) and 3D ‘‘pseudo’’-dynamic (free-breathing) scans for 25 timepoints of five healthy subjects were acquired using a 3T MRI (Siemens Magnetom Skyra). Prior to imaging, informed consent was obtained from each subject. Each subject's static (acquired with breath-hold) and dynamic scans were acquired in different sessions using the same sequence, parameters, and volume coverage. The static scan can be considered as another timepoint of the dynamic series which is acquired with a long gap in time. During the acquisition of the dynamic scans, the subjects were asked to breath slowly - a likely scenario during real interventions due to sedation. All the datasets (except the high-resolution static scans) were artificially undersampled to simulate the low-resolution datasets. The acquisition parameters of the datasets are listed in Table 1.

TABLE 1. MRI acquisition parameters for the CHAOS dataset and subject-wise 3D dynamic scans. Static scans were performed using the same subject-wise sequence parameters as the dynamic scans for one timepoint (TP), acquired in a different session.

	CHAOS (40 Subjects)	Protocol 1 (2 Subjects)	Protocol 2 (1 Subject)	Protocol 3 (1 Subject)	Protocol 4 (1 Subject)
Sequence	T1 Dual In-Phase & Opposed-Phase	T1w Flash 3D	T1w Flash 3D	T1w Flash 3D	T1w Flash 3D
Resolution	1.44 x 1.44 x 5 - 2.03 x 2.03 x 8 mm^3	0.90 x 0.90 x 4 mm^3	0.90 x 0.90 x 4 mm^3	0.90 x 0.90 x 4 mm^3	1.00 x 1.00 x 4 mm^3
FOV x, y, z	315 x 315 x 240 - 520 x 520 x 280 mm^3	300 x 225 x 176 mm^3	350 x 262 x 176 mm^3	350 x 262 x 192 mm^3	350 x 262 x 176 mm^3
Encoding matrix	256 x 256 x 26 - 400 x 400 x 50	320 x 240 x 44	384 x 288 x 44	384 x 288 x 48	352 x 264 x 44
Phase/Slice oversampling	-	10/0 %	10/0 %	10/0 %	10/0 %
TR	110.17 - 255.54 ms	2.37 ms	2.40 ms	2.40 ms	2.31 ms
TE	4.60 - 4.64 ms (In-Phase) 2.30 ms (Opposed-Phase)	1.00 ms	1.02 ms	1.02 ms	0.97 ms
Flip angle	80°	8°	8°	8°	8°
Bandwidth	-	920 Hz/Px	930 Hz/Px	930 Hz/Px	950 Hz/Px
GRAPPA factor	None	None	None	None	None
Phase/Slice partial Fourier	-	Off/Off	Off/Off	Off/Off	Off/Off
Phase/Slice resolution	-	50/64 %	50/64 %	50/64 %	50/64 %
Fat saturation	-	On	On	On	On
Time per TP	-	10.52 sec	12.80 sec	13.96 sec	11.36 sec

TABLE 2. Effective resolutions and estimated acquisition times (per TP) of the dynamic and static datasets after performing different levels of artificial undersampling.

	Protocol 1		Protocol 2		Protocol 3		Protocol 4	
	Resolution (mm^3)	Acq. Time (sec)	Resolution (mm^3)	Acq. Time (sec)	Resolution (mm^3)	Acq. Time (sec)	Resolution (mm^3)	Acq. Time (sec)
high-resolution (Ground-truth)	0.90 x 0.90 x 4	8.76	0.90 x 0.90 x 4	10.68	0.90 x 0.90 x 4	11.76	1.00 x 1.00 x 4	9.38
10% of k-space	2.70 x 2.70 x 4	0.88	2.70 x 2.70 x 4	1.07	2.70 x 2.70 x 4	1.18	3.00 x 3.00 x 4	0.94
6.25% of k-space	3.60 x 3.60 x 4	0.55	3.60 x 3.60 x 4	0.67	3.60 x 3.60 x 4	0.74	4.00 x 4.00 x 4	0.59
4% of k-space	4.50 x 4.50 x 4	0.35	4.47 x 4.47 x 4	0.43	4.47 x 4.47 x 4	0.47	4.99 x 4.99 x 4	0.38

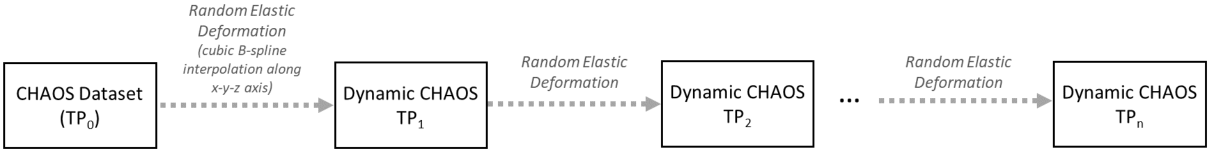

CHAOS Dataset
(TP0)
Random Elastic Deformation
(cubic B-spline interpolation along x-y-z axis)
Dynamic CHAOS
TP1
Random Elastic Deformation
Dynamic CHAOS
TP2
...
Random Elastic Deformation
Dynamic CHAOS
TPn

FIGURE 2. Flowchart of dynamic data generation. In this work, n = 24 and the total number of TP is 25 timepoints.

1) DYNAMIC DATA GENERATION

Since large dynamic MRI datasets that would be required for training are not available publicly, an artificial dynamic dataset was created. This was achieved by applying random elastic deformation of TorchIO [55] on the volumes from the CHAOS dataset. Figure 2 illustrates the dynamic data generation mechanism with the help of a flowchart.

Random displacement fields were generated using TorchIO's random elastic deformation with five control points, 5-20-20 mm of maximum displacements along x-y-z dimensions, respectively, and two locked borders. The displacement fields were then applied to the volumes of the CHAOS dataset using cubic B-spline interpolation, considering them as TP0, to generate artificial TP1. Then, a new set of random displacement fields with the same parameters were generated and applied on TP1 to generate TP2. In this manner, 24 artificial timepoints (TP1 - TP24) were generated for each of the volumes present in the original dataset. The displacement field tries to imitate the movement induced by breathing during a dynamic acquisition. The displacement field was set to expand and/or contract more in the anterosuperior (front-back) and the superoinferior (up-down) but less in the lateral (left-right) direction - to keep the deformation as realistic as possible. However, this manner of generating artificial breathing motion is not equivalent to physiological motion. It is to be noted that the goal of using this kind of artificial motion was to create a dataset from which a network can learn the pseudo-temporal relationship between two subsequent timepoints. This process results in an artificially created dynamic dataset - CHAOS dynamic, comprising 25 timepoints in total for each volume.

2) UNDERSAMPLING

The training data - the original CHAOS dataset and the artificially created CHAOS dynamic dataset, as well as the testing data (3D dynamic scans) were artificially undersampled in-plane using MRUnder [56], [57], available on Github: https://github.com/soumickmj/MRUnder, by taking only 10%, 6.25%, and 4% of the centre k-space, as shown in Fig. 3. By taking the centre of the k-space, the undersampling happened in both in-plane directions. If only the undersampling of the phase-encoding (PE) direction is considered, this results in MR acceleration factors (i.e. how many times the acquisitions will be faster if these undersampling is performed) of 3, 4, and 5, respectively. Considering the actual amount of data used (i.e. in both directions while taking the centre of the k-space) during SR reconstruction, this results in theoretical acceleration factors of 10, 16, and 25, respectively.

The effective resolutions and estimated acquisition times for each of the dynamic test datasets are calculated using Eq. 6 and shown in Table 2

$$T_{acq} = PE_n \times TR \times S_m \tag{6}$$

where T_{acq} is the estimated acquisition time, given the number of phase-encoding lines PE_n, the repetition time TR, and the number of slices acquired S_m [29]. During the calculation of Table 2, phase/slice resolution and phase/slice oversampling (Table 1) were also taken into consideration while calculating PE_n and S_m.

D. IMPLEMENTATION, TRAINING, AND INFERENCE

The proposed model was trained on 3D volumes from the artificially created dynamic version of a publicly available benchmark dataset, as summarised in Fig. 4. Fig. 5 shows an overview of the inference steps. The inference process starts (following Eq. 3) with the Antipasto phase - by supplying the high-resolution patient-specific static scan as a prior image on the first channel of the network (as $\hat{\mathbf{y}}_{t-1}$ is not yet available), and by supplying $\mathbf{x}_t$ (in this case, $\mathbf{x}_1$) on the second channel of the network.

It is to be noted that the static scan has the same resolution, contrast and volume coverage as the high-resolution ground-truth dynamic scan. However, to keep the testing environment similar to a real-life scenario and keep a fast speed of inference, the static and dynamic datasets were not co-registered, as registration is typically time-consuming. After this, the network super-resolves $\mathbf{x}_1$ to $\hat{\mathbf{y}}_1$. Now for the next timepoint, $\hat{\mathbf{y}}_1$ and $\mathbf{x}_2$ are supplied as input to the network and the network provides $\hat{\mathbf{y}}_2$ as output.

The implementation was done using PyTorch [58], and the training and the inference were performed using Nvidia Tesla V100 GPUs. Following the hypothesis of using batch size one to be able to learn an exact mapping function between the specific pair of low- and high-resolution images [57], batch size during training and inference in this research was also set to one. The loss during training was calculated using perceptual loss [48], with the help of a perceptual loss network [35], and was minimised using the Adam optimiser with a learning rate of 10^{-4} for 100 epochs. The code of the implementation is available on GitHub: https://github.com/soumickmj/DDoS.

1) PERCEPTUAL LOSS

Similar to the previous work [29], perceptual loss [48] was employed to compute the loss during training. For the same, the initial three blocks of the frozen pre-trained (on 7T MRA scans, for the task of vessel segmentation) UNet MSS model was used as the perceptual loss network (PLN) [35]. The job of this PLN is to extract "deep features" of different abstraction levels at the different levels of the PLN, from the super-resolved volumes and their corresponding ground-truths. The features extracted from the super-resolved output and the ground-truth were compared against each other using mean absolute error (L1 loss). Finally, all these L1 losses were added together and backpropagated. This perceptual loss L between a ground-truth image $\mathbf{y}_{s,t}$ and the corresponding predicted image $\hat{\mathbf{y}}_{s,t}$ can be formulated as:

$$L_{\mathbf{y}_{s,t},\hat{\mathbf{y}}_{s,t}} = \sum_{b=1}^{B}\sum_{f=1}^{F_b} |f_{\mathbf{y}_{s,t}} - f_{\hat{\mathbf{y}}_{s,t}}| \tag{7}$$

where B is the number of blocks of the PLN to be used for loss calculation, F_b is the number of features block b can generate (depending upon the network architecture), $f_{\mathbf{y}_t}$ is a particular feature generated feature from the ground-truth $\mathbf{y}_t$, and $f_{\hat{\mathbf{y}}_t}$ is the corresponding feature generated from the prediction $\hat{\mathbf{y}}_t$.

E. EVALUATION CRITERIA

The quality of super-resolution was evaluated quantitatively with the help of the structural similarity index (SSIM) [59], the peak signal-to-noise ratio (PSNR), and the normalised root mean squared error (NRMSE). The perceptual quality of the output was evaluated with the help of SSIM, which compares luminance, contract, and structure terms between two given images x and y, which for this research represent the output and ground-truth, respectively, using the following formula:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{8}$$

where $\mu_x, \mu_y, \sigma_x, \sigma_y$ and σ_{xy} are the local means, standard deviations, and cross-covariance for images x and y, respectively. $c_1 = (k_1L)^2$ and $c_2 = (k_2L)^2$, where L is the dynamic range of the pixel-values, $k_1 = 0.01$ and $k_2 = 0.03$. Moreover, the quality of the super-resolution was

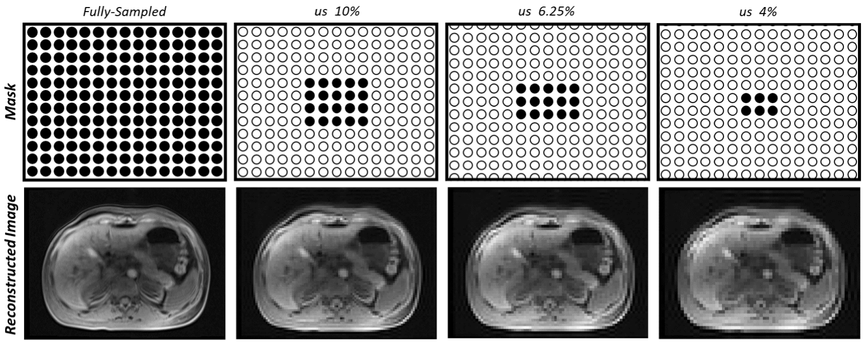


FIGURE 3. Graphical representation of masks and the corresponding reconstructed images. The undersampled masks were generated by taking only 10%, 6.25%, and 4% of the centre k-space. The data points in black denote the sampled points, and the white dots denote the undersampled data.

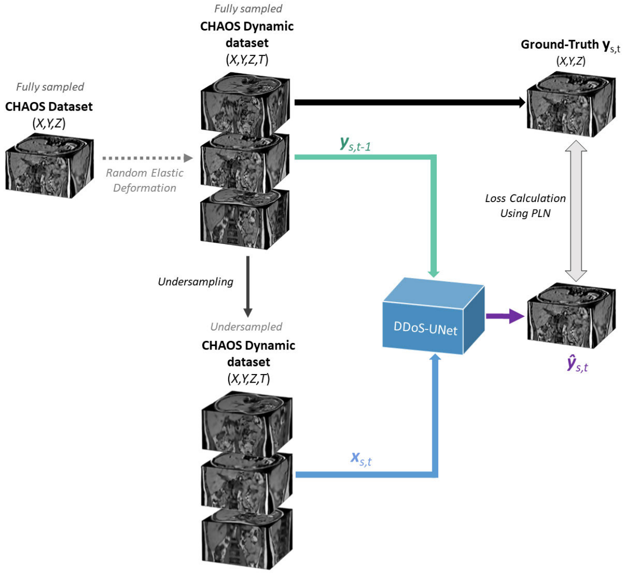


FIGURE 4. Method Overview: Training. Initially, random elastic deformation (TorchIO [55]) is applied to the CHAOS dataset (fully sampled) to generate the artificial CHAOS dynamic dataset. The CHAOS dynamic dataset is then undersampled to generate the final training dataset. The model is subsequently trained by providing low-resolution (undersampled) current timepoint ($\mathbf{x}_{s,t}$) along with the high-resolution (fully sampled) previous timepoint ($\mathbf{y}_{s,t-1}$) as input, and the output is compared against the ground-truth high-resolution current timepoint ($\mathbf{y}_{s,t}$).

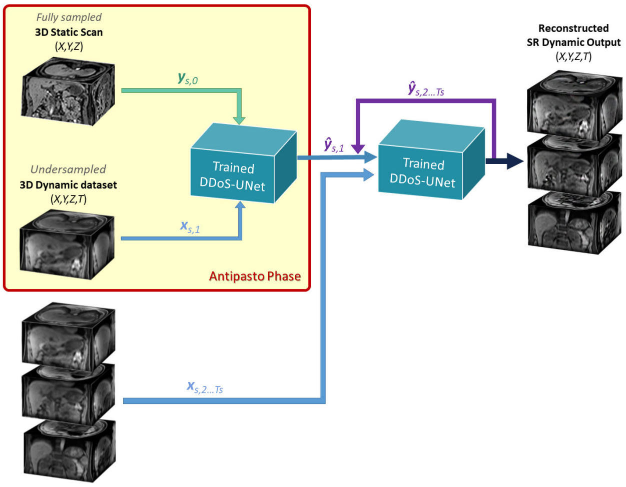


FIGURE 5. Method Overview: Inference. A 3D static subject-specific planning scan (fully sampled) is supplied as the high-resolution prior image ($\hat{\mathbf{y}}_{s,0}$), along with the first low-resolution (undersampled) timepoint ($\mathbf{x}_{s,1}$) of the 3D dynamic dataset, are supplied as input to the trained DDoS-UNet model, and the model super-resolves $\mathbf{x}_{s,1}$ to obtain $\hat{\mathbf{y}}_{s,1}$. This initial phase is called the "Antipasto" phase. $\hat{\mathbf{y}}_{s,1}$ is subsequently supplied as input, together with the next low-resolution timepoint $\mathbf{x}_{s,2}$ to the same trained DDoS-UNet model to obtain $\hat{\mathbf{y}}_{s,2}$. This process is continued recursively until all the timepoints of the low-resolution (undersampled) 3D dynamic dataset are super-resolved, by supplying pairs of $\hat{\mathbf{y}}_{s,T_S-1}$ and $\mathbf{x}_{s,T_S}$ to obtain each of the $\hat{\mathbf{y}}_{s,T_S}$.

measured statistically with the help of PSNR and NRMSE, both of which are calculated using the mean-square error (m) between x and y as:

$$PSNR(x, y) = 10 \log_{10} \left(\frac{R^2}{m} \right) \tag{9}$$

where R is the maximum fluctuation in the input image, and

$$NRMSE(x, y) = \frac{\sqrt{m}\sqrt{N}}{||y||} \tag{10}$$

where $|| \cdot ||$ denotes the Frobenius norm, N is the number of elements in the data, and y is the ground-truth.

The statistical significance of the differences in the quantitative metrics for the proposed method against the other baselines was computed using the Mann-Whitney U test. Apart from quantitative evaluations, the results were also compared qualitatively.

III. RESULTS

The performance of the DDoS-UNet was compared for three different levels of undersampling: 10%, 6.25%, and 4% of the centre k-space, against the low-resolution input, traditional trilinear interpolation, Fourier-interpolated input (zero-padded k-space), and finally against two different baseline deep learning models: two UNet models identical to the DDoS-UNet except for the initial layer (unlike DDoS-UNet, these UNets received one input) - one of them trained on the original CHAOS dataset (T1-dual images, in- and opposed phase) [54], and the other one was trained using artificial

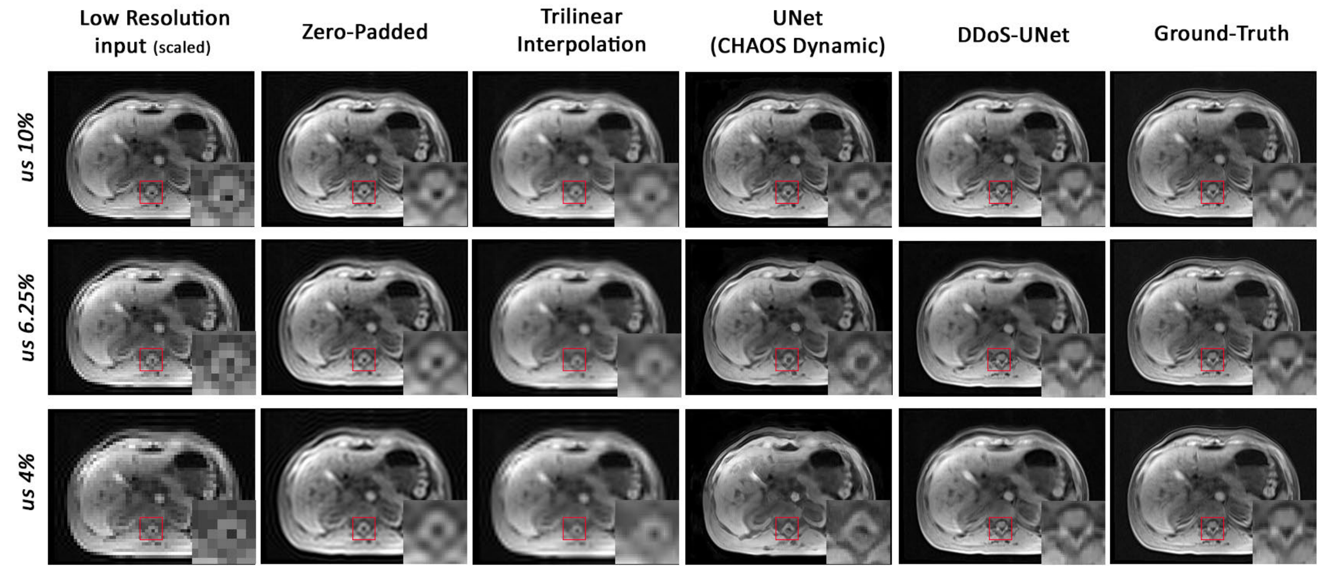


FIGURE 6. Comparative results of low-resolution (10%, 6.25%, and 4% of k-space) 3D dynamic data of the same slice. From left to right: low-resolution images (scaled-up, nearest-neighbour interpolation), interpolated input (Trilinear), <ero-padded reconstruction, output of UNet trained on CHAOS dynamic dataset, output of DDoS-UNet and ground-truth images.

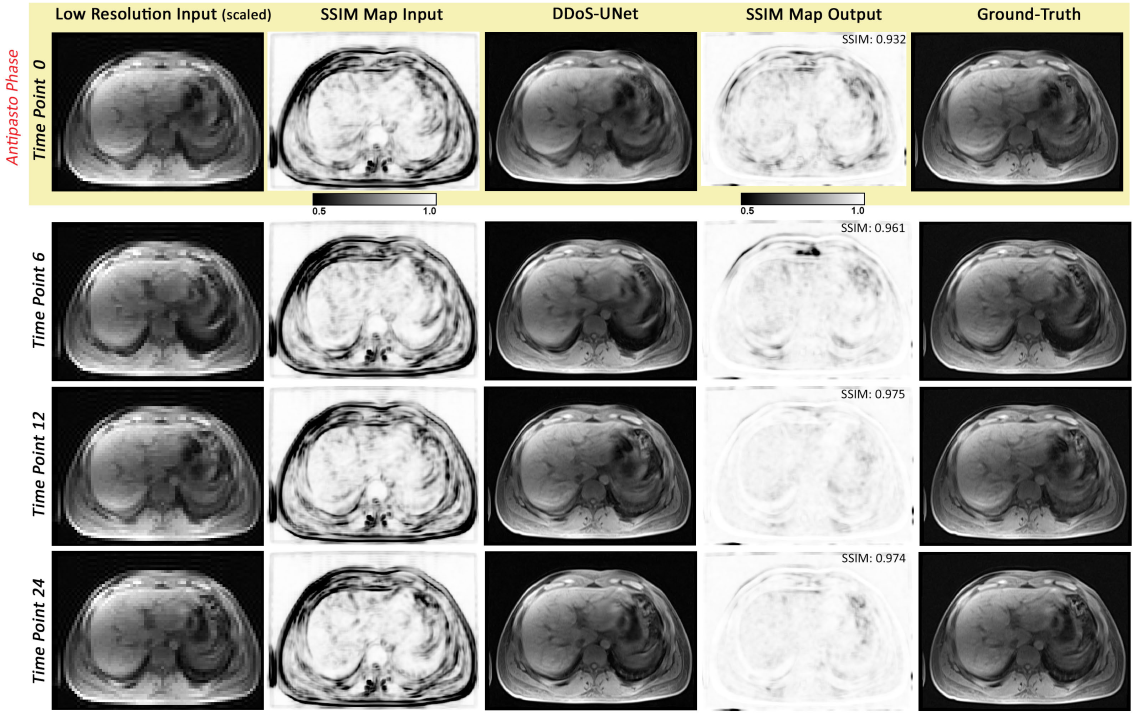


FIGURE 7. An example comparison of the low-resolution input of 4% of k-space with the super-resolution (SR) result of the DDoS-UNet over four different time points, compared against the high-resolution ground-truth using SSIM maps (generated with a local window size of seven).

dynamic CHAOS (see Sec. II-C1). The training dataset of the second UNet was identical to the training dataset of the DDoS-UNet. In case of DDoS-UNet, the model utilises information from the previous timepoint in terms of the prior image, while both the UNets attempt to super-resolve the provided input timepoint without any additional help. The models were evaluated on real dynamic datasets of five subjects, each consisting of 25 timepoints (details in Sec. II-C). The inference process for the DDoS-UNet was started with the patient-specific prior high-resolution

TABLE 3. The average and the standard deviation of SSIM, PSNR, and NRMSE. The table shows the results for different resolutions. For all comparisons of the DDoS-UNet against the baselines, the p-values were always less than 0.0001.

	10% of k-space			6.25% of k-space			4% of k-space		
	SSIM	PSNR	NRMSE	SSIM	PSNR	NRMSE	SSIM	PSNR	NRMSE
Trilinear Interpolation	0.872±0.014	28.631±1.364	0.192±0.023	0.821±0.017	26.770±1.226	0.238±0.024	0.765±0.022	25.248±1.298	0.283±0.025
Zero-padded	0.949±0.013	36.138±1.753	0.082±0.016	0.910±0.018	29.761±1.640	0.124±0.019	0.863±0.021	32.520±1.508	0.170±0.025
UNet (CHAOS)	0.967±0.006	38.359±1.580	0.021±0.004	0.944±0.010	35.623±1.552	0.029±0.005	0.916±0.015	32.658±1.598	0.041±0.007
UNet (CHAOS Dynamic)	0.959±0.012	37.376±1.275	0.024±0.003	0.941±0.012	35.113±1.566	0.031±0.006	0.914±0.012	33.620±1.035	0.036±0.004
DDoS-UNet	**0.980±0.006**	**41.824±2.070**	**0.014±0.003**	**0.967±0.011**	**39.494±2.121**	**0.019±0.005**	**0.951±0.017**	**37.557±2.179**	**0.024±0.006**

static scan and first low-resolution timepoint as input and then continued by supplying the previous super-resolved timepoint with the current low-resolution timepoint to super-resolve the current timepoint (as explained in Sec. II-B1).

Fig. 6 shows a qualitative comparison of the results obtained by the different methods for different levels of undersampling. It can be observed that the proposed DDoS-UNet managed to restore finer details better than the other methods. Moreover, both the baseline UNet models show better anatomical structures than the zero-padded reconstructions. Furthermore, the comparison with the help of SSIM maps between the input (low-resolution images) and output (super-resolved images) of the DDoS-UNet are shown in Fig. 7. It reveals that the reconstruction quality of the initial timepoint is not very good, but the network manages to recover from the initial struggle during the Antipasto phase, and manages to reconstruct the subsequent timepoints much better and consistently over all the timepoints. This can be attributed to the fact that the static and dynamic scans are acquired in two different sessions, and they are not co-registered. Finally, Fig. 8 shows the qualitative comparisons of the different methods for two regions of interest (ROI). It shows the proposed DDoS-UNet framework results in better reconstruction performance than the baseline UNet models. Between the baseline UNets, UNet trained on CHAOS dynamic dataset managed to recover finer anatomical details better than the UNet model trained on CHAOS dataset.

Table 3 presents the quantitative results for all the methods. It can be observed that both the baseline UNet models (trained on the original CHAOS dataset and on the CHAOS dynamic dataset) outperformed the non-DL baselines: trilinear interpolation and zero-padded reconstruction (sinc interpolation), and the proposed DDoS-UNet method outperformed all the baselines for all three undersampling factors in all three metrics with statistical significance (p-values always less than 0.001). It can be further observed that the UNet trained on the original CHAOS dataset outperformed the UNet trained on the CHAOS dynamic dataset. Fig. 9 shows the resultant SSIM and PSNR values over all subjects and timepoints by means of box plots. It can be observed that the improvements obtained by the proposed method increase with the increase in the undersampling factor. Fig. 10 portrays the SSIM values over the different timepoints averaged for all five subjects. It can be seen that after the initial timepoint TP0 (Antipasto phase), the proposed DDoS-UNet achieved consistently better SSIM values compared to all the other methods. Finally, Fig. 11 shows the average SSIM values over the different timepoints (excluding the Antipasto phase) for each subject. The median values over TP1 to TP24 for each subject resulted in SSIM values in the range 0.988 to 0.975, 0.980 to 0.960, and 0.970 to 0.945, for 10%, 6.25%, and 4% of k-space, respectively. Fig. 10 and 11 show that the proposed DDoS-UNet is able to reconstruct different protocols and subjects efficiently while being stable over different timepoints.

A. COMPARISON AGAINST PREVIOUS WORK

The proposed method was also compared against the previously-proposed fine-tuning based super-resolution of dynamic MRI [29], referred to here as ‘‘Fine-tuned SR’’. Fig. 12 shows a qualitative comparison of these method with the proposed DDoS-UNet while super-resolving 4% of the centre k-space. It can be seen that although the Fine-tuned SR approach could restore the information from highly undersampled input, the result is very smooth and fails to recover the details of anatomy and fine structures compared to the DDoS-UNet. However, Fine-tuned SR works with patches - making it suitable for GPUs with lesser memories (e.g. 12GB Nvidia GeForce RTX 2080 TI) than the ones required for DDoS-UNet (e.g. 32GB Nvidia Tesla V100). But, working with patches also increases the processing time. During inference, DDoS-UNet took (on average of all subjects) 0.36 seconds for each timepoint (9 seconds for 25 TPs). On the other hand, Fine-tuned SR took 2 minutes per timepoint (50 minutes for 25 TPs - average over all five subjects) while working with a batch size of 1900 on the same GPU. If the same inference is performed using an inferior GPU (Nvidia GeForce RTX 2080 TI) with a batch size of 96 (similar to training and fine-tuning stages), the inference time increases for each timepoint. This makes DDoS-UNet better suitable for real-time or near real-time applications than the Fine-tuned SR approach. Moreover, the time required for inferring one timepoint using Fine-tuned SR depends on the matrix size of the volumes, as a larger matrix would result in more patches if the same patch size and strides are used. If Fine-tuned SR also works with the whole volume, this difference in inference time can be resolved. But in that case, there will be only one forward pass using one static volume for fine-tuning (currently it uses all the possible patches) - making it unsuitable. Moreover, the Fine-tuned SR approach requires an additional step after training the model

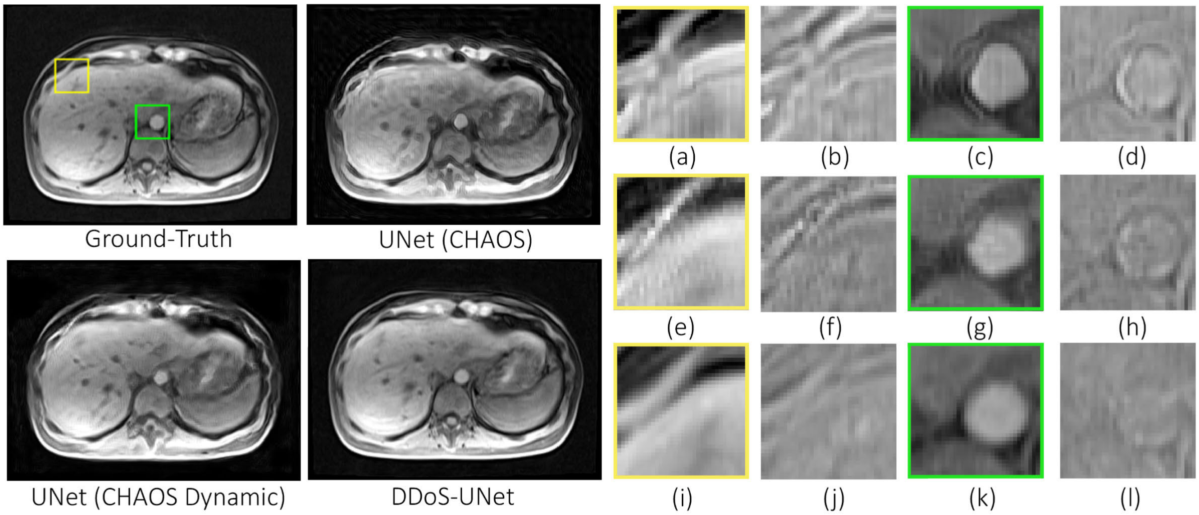


FIGURE 8. An example of reconstructed results from UNet baselines and DDoS-UNet, compared against its ground-truth (GT) for low-resolution images from 4% of k-space. From left to right, top to bottom: ground-truth, SR result of the UNet baseline (UNet CHAOS), SR result of the UNet baseline trained on CHAOS dynamic (UNet CHAOS Dynamic) and SR result of the DDoS-UNet. For the yellow ROI, (a-b): UNet CHAOS and the difference image from GT, (e-f): SR result of UNet CHAOS Dynamic, and (i-j): SR result of DDoS-UNet and the difference image from GT. The images on the right are identical examples for the green ROI. It can be observed that the difference images of DDoS-UNet have considerably fewer structures than the other two - indicating that there is the least amount of difference between the DDoS-UNet and the ground-truth compared to the other models.

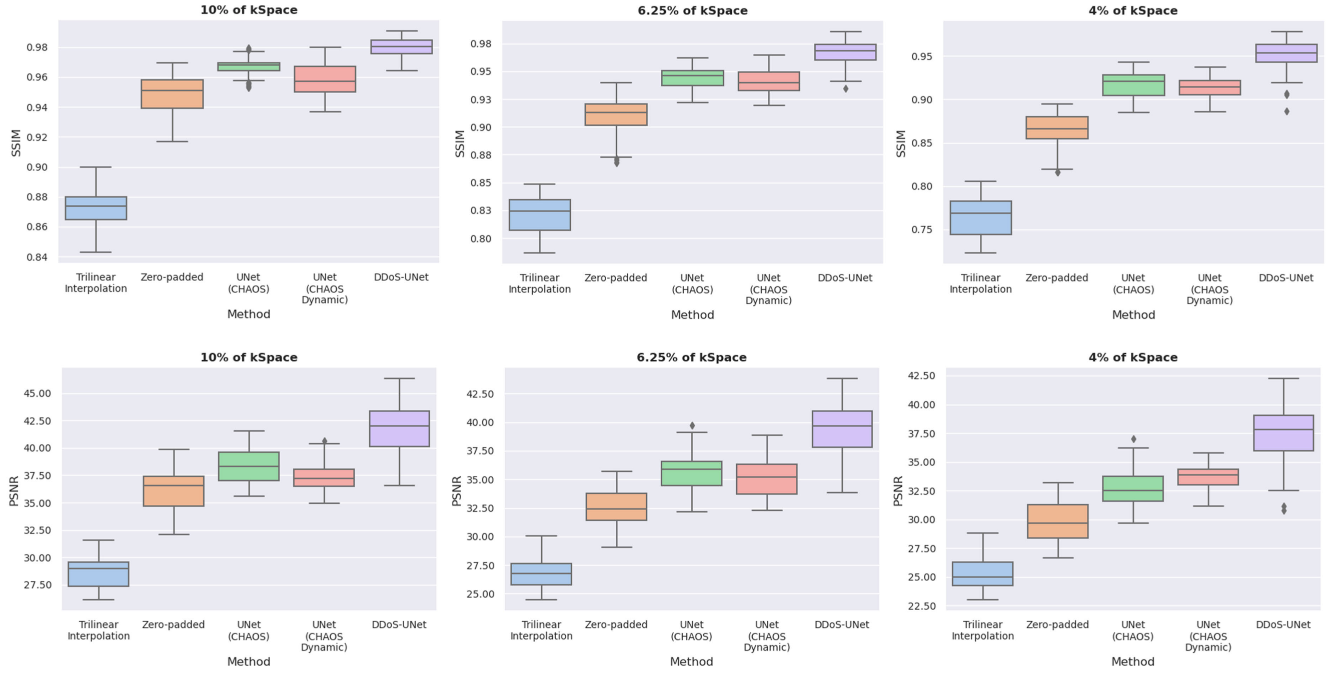


FIGURE 9. Quantitative comparison of different methods using SSIM and PSNR - for all subjects and timepoints combined, for different levels of undersampling. For all comparisons of the DDoS-UNet against the baselines, the p-values were always less than 0.0001.

- the step of fine-tuning using subject-specific static scans. Depending upon the available resources, the fine-tuning can take 8-10 hours - which can be avoided using DDoS-UNet as it does not require this step.

IV. DISCUSSION

This paper presents the **D**ynamic **D**ual-channel **o**f **S**uper-resolution using **UNet** (DDoS-UNet) framework and shows its applicability for reconstructing low-resolution (undersampled) dynamic MRIs up to a theoretical acceleration factor of 25. The quantitative and qualitative results demonstrate the superiority of the proposed method.

The UNet model trained on the original CHAOS dataset performed better quantitatively than the UNet model trained on the CHAOS dynamic dataset, even though the latter had 25 times more volumes (24 artificially created timepoints on top of the original one). This can be attributed to the quality of the CHAOS dynamic dataset. Due to the repeated

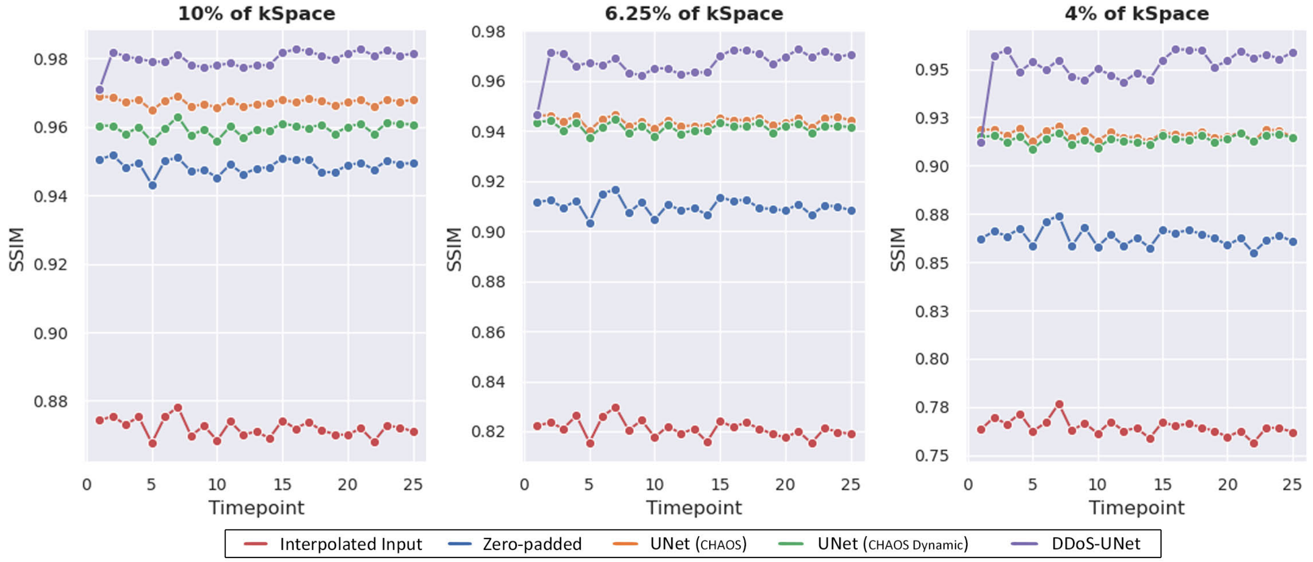


FIGURE 10. Line plot showing the average SSIM values for each subject across all timepoints, for different levels of undersampling. An initial drop can be observed for the first timepoint for DDoS-UNet, which is referred to here as the Antipasto phase; thereafter, the network performs with stability for the remaining timepoints.

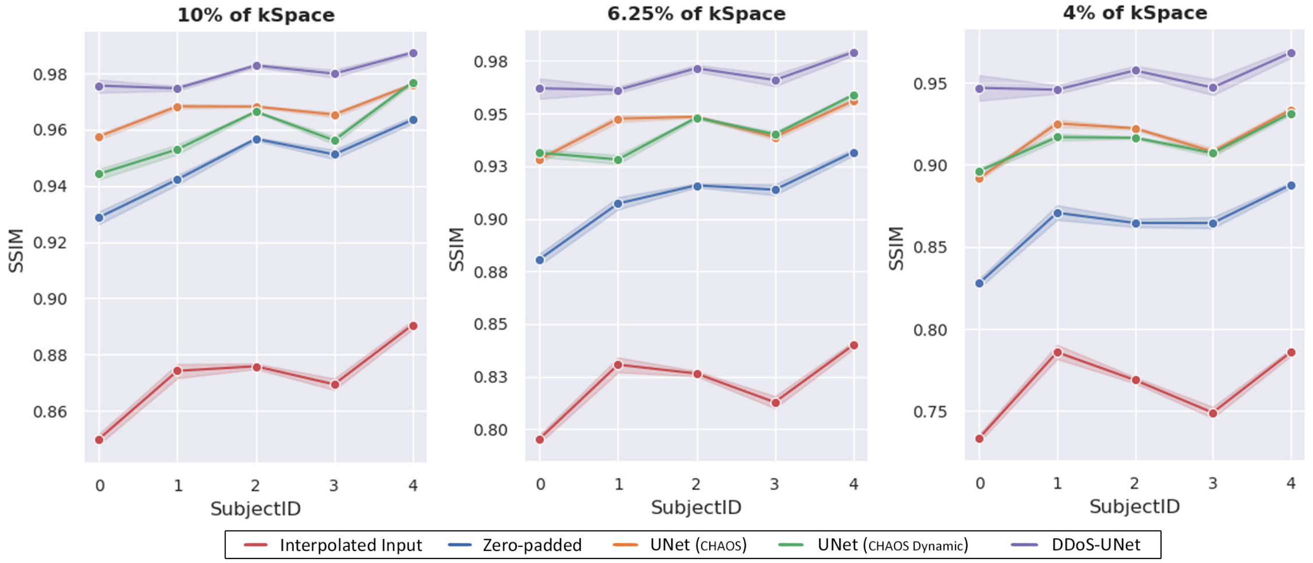


FIGURE 11. Line plot showing the mean and 95% confidence interval of the resultant SSIM values over the different timepoints (excluding the initial one, the Antipasto phase) for each subject. The red, blue, orange, green, and violet lines represent the reconstruction results of trilinear interpolation, zero-padding (sinc interpolation), UNet trained on CHAOS dataset, UNet trained on CHAOS Dynamic dataset, and DDoS-UNet, respectively. It is to be noted that the dots (of each subject) are connected only to help in comparing the results and do not represent any additional relationship (e.g. ordering) among the subjects.

applications of the random elastic deformation on the original dataset, which includes interpolation, the sharpness of the later timepoints decreased due to the accumulated interpolation errors. This might have also negatively impacted the results of the DDoS-UNet. Improving the quality of the artificial dynamic dataset might improve the performance of both of these models. Ultimately, if the DDoS-UNet is trained using real dynamic MRIs, the reconstruction performance might improve further. It is worth mentioning, however, that for the highest undersampling factor (4% of the k-space), UNet trained on CHAOS dynamic dataset resulted in better PSNR than UNet trained on CHAOS dataset, and also visual comparison (Fig. 8) revealed that the UNet trained on CHAOS dynamic managed to restore finer anatomical details better.

The fundamental difference between the proposed method and the baselines is the temporal prior. The baseline UNet models only receive the low-resolution image as input.

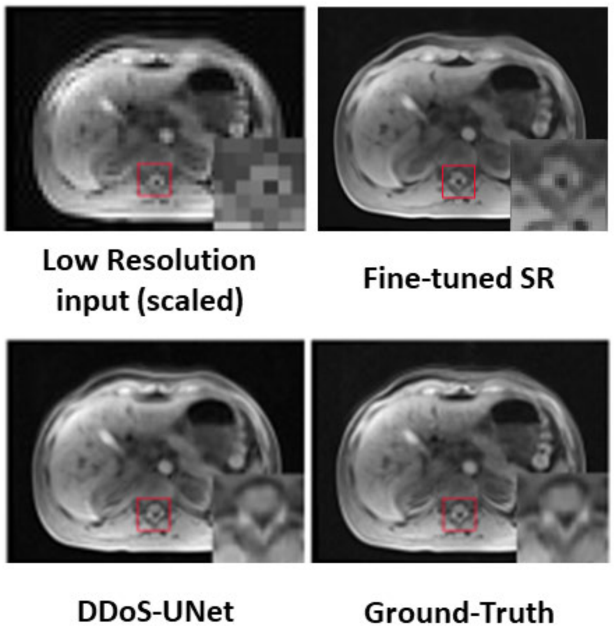


FIGURE 12. A qualitative comparison of the previously proposed method [29] (referred to as Fine-tuned SR) against the proposed DDoS-UNet while super-resolving 4% of the centre k-space. The red ROI shows the spinal cord area. As can be seen in the image, although the Fine-tuned SR approach could restore the information from the low-resolution input, the result is very smooth and fails to recover the details of anatomy and fine structures compared to the DDoS-UNet.

But the DDoS-UNet receives the super-resolved previous timepoint (as the temporal prior) along with the current low-resolution image as input. Hence, the authors attribute the improvements observed with DDoS-UNet to the addition of the prior image. The authors hypothesise that the network managed to take help from this prior image to super-resolve the current timepoint as they are temporally related. This confirms the hypothesis of the authors formulated in Eq. 5.

A final observation can be made regarding the results of the DDoS-UNet for the different timepoints. The result of the initial timepoint was considerably worse compared to the rest of the other timepoints (similar or better than the UNets, and always better than the non-DL baselines), as can be seen in Figures 10 and 7. This initial timepoint was reconstructed by supplying the high-resolution subject-specific static scan as the prior image, referred here as the Antipasto phase, whereas the remaining timepoints were reconstructed by supplying the super-resolved previous timepoint as the prior image. The static scan has a big temporal difference from the first timepoint of the dynamic scan as they were acquired in different sessions, while the subsequent timepoints of the dynamic scan were closer in time. The network faces difficulties reconstructing the initial timepoint, but then recovers from it after super-resolving the first one and then maintaining its performance steadily for all subsequent timepoints. This also supports the hypothesis that the DDoS-UNet learnt both spatial and temporal relationships, as shown in Eqs. 3 and 5 in Sec. II-B1.

The reconstruction (inference) time using the proposed DDoS-UNet was approximately 0.36 seconds for each timepoint (9 seconds for 25 TPs) while reconstructing using an Nvidia Tesla V100 GPU. Fast reconstruction time, coupled with the high speed of acquisition (shown in Table 2), this method shows the potential to acquire and reconstruct each timepoint of a 3D dynamic acquisition within 0.71 seconds (for 4% of k-space with Protocol 1) - making it a potential candidate for near real-time MR acquisitions. The acquisition time can be further reduced using techniques such as parallel imaging, as shown in the earlier work [29]. The focus of this paper is on abdominal imaging; however, this method might also be used for other types of dynamic imaging, e.g. cardiac imaging. Moreover, this approach might be adapted to super-resolve other dynamic imaging modalities, e.g. endoscopy.

V. CONCLUSION AND FUTURE WORK

This research proposes the DDoS-UNet model to perform 3D volumetric super-resolution of low-resolution dynamic MRIs by using a subject-specific high-resolution prior planning scan and exploiting the spatio-temporal relationship present in the dynamic MRI. The proposed network was trained using an artificially created dynamic dataset from the CHAOS abdominal benchmark dataset and then was tested using dynamic MRIs comprising of 25 timepoints. It was observed that even though the network was trained using a dataset with MRI acquisition parameters very different from the test set, the network was able to super-resolve the given input images with high accuracy - even for high undersampling factors. The proposed method resulted in 0.951±0.017 SSIM while super-resolving the highest undersampling experimented in this research (i.e. 4% centre k-space), whereas the baseline UNet (model without supplying the super-resolved previous timepoint as prior information) resulted in 0.916±0.015. The results show that the proposed network managed to mitigate the spatio-temporal problem of dynamic MRI by performing spatial super-resolution with the help of the temporal relationship present in the data without compromising the acquisition speed. Given the reconstruction speed of the proposed approach, this can be a candidate for near real-time dynamic acquisition scenarios, such as interventional MRI.

The proposed approach employs a multi-channel approach to supply the prior image (initially, the high-resolution static scan, then the super-resolved volumes). However, other approaches such as dual-branch have also been proposed [22], which might also be used to supply such prior images to the network. Such an architecture can deal with the prior image and the low-resolution image differently (i.e. different weights applied on each), whereas the current initial layer of the network treats them equally and merges them as an internal representation in the initial layer. Moreover, DDoS-UNet is interesting in interventional setup. During interventions, devices such as catheters are used, which were not present in the training set. The authors plan to extend the current research by evaluating the proposed model's reconstruction performance for such devices.

ACKNOWLEDGMENT

(Soumick Chatterjee and Chompunuch Sarasaen contributed equally to this work.)

AUTHOR CONTRIBUTIONS STATEMENT

Soumick Chatterjee and Oliver Speck designed the study, Soumick Chatterjee and Andreas Nürnberger developed the method, Soumick Chatterjee performed the experiments, Chompunuch Sarasaen analysed the results and created the plots, Soumick Chatterjee and Chompunuch Sarasaen wrote the manuscript, Andreas Nürnberger and Oliver Speck supervised the study, Georg Rose, Andreas Nürnberger, and Oliver Speck acquired the funding, all authors reviewed and approved the manuscript.

DATA AVAILABILITY

Data cannot be made used in this research is available from the first author upon request.

ADDITIONAL INFORMATION

All methods were carried out in accordance with relevant guidelines and regulations. The code of this project is publicly available on GitHub: https://github.com/soumickmj/DDoS.

REFERENCES

[1] J. Barkhausen, T. Kahn, G. A. Krombach, C. K. Kuhl, J. Lotz, D. Maintz, J. Ricke, S. O. Schoenberg, T. J. Vogl, and F. K. Wacker, "White paper: Interventional MRI: Current status and potential for development considering economic perspectives—Part 1: General application," in *RöFo-Fortschritte Auf dem Gebiet der Röntgenstrahlen und der Bildgebenden Verfahren*, vol. 189. New York, NY, USA: Georg Thieme Verlag KG, 2017, pp. 611–623.

[2] A. H. Mahnken, J. Ricke, and K. E. Wilhelm, *CT-and MR-Guided Interventions in Radiology*, vol. 22. Cham, Switzerland: Springer, 2009.

[3] M. A. Bernstein, K. F. King, and X. J. Zhou, *Handbook MRI Pulse Sequences*. Amsterdam, The Netherlands: Elsevier, 2004.

[4] C. E. Shannon, "Communication in the presence of noise," *Proc. IEEE*, vol. 72, no. 9, pp. 1192–1201, Jan. 1984.

[5] M. Lustig, J. M. Santos, D. L. Donoho, and J. M. Pauly, "kt SPARSE: High frame rate dynamic mri exploiting spatio-temporal sparsity," in *Proc. 13th Annu. Meeting (ISMRM)*, vol. 2420, 2006, pp. 1.

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

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

[8] S. Zhang, K. T. Block, and J. Frahm, "Magnetic resonance imaging in real time: Advances using radial FLASH," *J. Magn. Reson. Imag.*, vol. 31, no. 1, pp. 101–109, Jan. 2010.

[9] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2016.

[10] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 1132–1140.

[11] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 105–114.

[12] K. Zeng, H. Zheng, C. Cai, Y. Yang, K. Zhang, and Z. Chen, "Simultaneous single- and multi-contrast super-resolution for brain MRI images based on a convolutional neural network," *Comput. Biol. Med.*, vol. 99, pp. 133–141, Aug. 2018.

[13] X. He, Y. Lei, Y. Fu, H. Mao, W. J. Curran, T. Liu, and X. Yang, "Super-resolution magnetic resonance imaging reconstruction using deep attention networks," *Proc. SPIE*, vol. 11313, Sep. 2020, Art. no. 113132J.

[14] T. Yamaguchi, H. Fukuda, R. Furukawa, H. Kawasaki, and P. Sturm, "Video deblurring and super-resolution technique for multiple moving objects," in *Proc. Asian Conf. Comput. Vis.* Cham, Switzerland: Springer, 2010, pp. 127–140.

[15] J. Caballero, C. Ledig, A. Aitken, A. Acosta, J. Totz, Z. Wang, and W. Shi, "Real-time video super-resolution with spatio-temporal networks and motion compensation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2848–2857.

[16] A. Lucas, S. López-Tapia, R. Molina, and A. K. Katsaggelos, "Generative adversarial networks and perceptual losses for video super-resolution," *IEEE Trans. Image Process.*, vol. 28, no. 7, pp. 3312–3327, Jul. 2019.

[17] J. Rasch, V. Kolehmainen, R. Nivajärvi, M. Kettunen, O. Gröhn, M. Burger, and E.-M. Brinkmann, "Dynamic MRI reconstruction from undersampled data with an anatomical prescan," *Inverse Problems*, vol. 34, no. 7, Jul. 2018, Art. no. 074001.

[18] A. Kofler, M. Dewey, T. Schaeffter, C. Wald, and C. Kolbitsch, "Spatio-temporal deep learning-based undersampling artefact reduction for 2D radial cine MRI with limited training data," *IEEE Trans. Med. Imag.*, vol. 39, no. 3, pp. 703–717, Mar. 2020.

[19] 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," *Sci. Rep.*, vol. 10, no. 1, pp. 1–13, Aug. 2020.

[20] C. A. Segall, A. K. Katsaggelos, R. Molina, and J. Mateos, "Bayesian resolution enhancement of compressed video," *IEEE Trans. Image Process.*, vol. 13, no. 7, pp. 898–911, Jul. 2004.

[21] S. P. Belekos, N. P. Galatsanos, and A. K. Katsaggelos, "Maximum a posteriori video super-resolution using a new multichannel image prior," *IEEE Trans. Imag. Process.*, vol. 19, pp. 1451–1464, 2010.

[22] S. Chatterjee, A. Sciarra, M. Dünnwald, S. Oeltze-Jafra, A. Nürnberger, and O. Speck, "Retrospective motion correction of MR images using prior-assisted deep learning," in *Proc. Med. Imag. Meets (NeurIPS)*, Dec. 2020, pp. 1–5.

[23] S. Araki, T. Hayashi, M. Delcroix, M. Fujimoto, K. Takeda, and T. Nakatani, "Exploring multi-channel features for denoising-autoencoder-based speech enhancement," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2015, pp. 116–120.

[24] P. Barros, S. Magg, C. Weber, and S. Wermter, "A multichannel convolutional neural network for hand posture recognition," in *Proc. Int. Conf. Artif. Neural Netw.*, 2014, pp. 403–410.

[25] Z.-Q. Wang, J. Le Roux, and J. R. Hershey, "Multi-channel deep clustering: Discriminative spectral and spatial embeddings for speaker-independent speech separation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2018, pp. 1–5.

[26] A. A. Nugraha, A. Liutkus, and E. Vincent, "Multichannel audio source separation with deep neural networks," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 24, no. 9, pp. 1652–1664, Sep. 2016.

[27] J. Casebeer, Z. Wang, and P. Smaragdis, "Multi-view networks for multi-channel audio classification," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 940–944.

[28] C. Xu, W. Huang, H. Wang, G. Wang, and T.-Y. Liu, "Modeling local dependence in natural language with multi-channel recurrent neural networks," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, 2019, pp. 5525–5532.

[29] C. Sarasaen, S. Chatterjee, M. Breitkopf, G. Rose, A. Nürnberger, and O. Speck, "Fine-tuning deep learning model parameters for improved super-resolution of dynamic MRI with prior-knowledge," *Artif. Intell. Med.*, vol. 121, Nov. 2021, Art. no. 102196.

[30] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervent.*, 2015, pp. 234–241.

[31] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning dense volumetric segmentation from sparse annotation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Interv.*, 2016, pp. 424–432.

[32] F. Milletari, N. Navab, and S.-A. Ahmadi, ''V-Net: Fully convolutional neural networks for volumetric medical image segmentation,'' in *Proc. 4th Int. Conf. 3D Vis. (3DV)*, Oct. 2016, pp. 565–571.

[33] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, ''UNet++: A nested U-Net architecture for medical image segmentation,'' in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Cham, Switzerland: Springer, 2018, pp. 3–11.

[34] O. Oktay, J. Schlemper, L. Le Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y Hammerla, B. Kainz, B. Glocker, and D. Rueckert, ''Attention U-Net: Learning where to look for the pancreas,'' 2018, *arXiv:1804.03999*.

[35] S. Chatterjee, K. Prabhu, M. Pattadkal, G. Bortsova, C. Sarasaen, F. Dubost, H. Mattern, M. de Bruijne, O. Speck, and A. Nürnberger, ''DS6, deformation-aware semi-supervised learning: Application to small vessel segmentation with noisy training data,'' 2020, *arXiv:2006.10802*.

[36] A. Jansson, E. Humphrey, N. Montecchio, R. Bittner, A. Kumar, and T. Weyde, ''Singing voice separation with deep u-net convolutional networks,'' in *Proc. 18th Int. Soc. Music Inf. Retr. Conf.*, 2017, pp. 23–27.

[37] D. Stoller, S. Ewert, and S. Dixon, ''Wave-U-Net: A multi-scale neural network for end-to-end audio source separation,'' 2018, *arXiv:1806.03185*.

[38] H.-S. Choi, J.-H. Kim, J. Huh, A. Kim, J.-W. Ha, and K. Lee, ''Phase-aware speech enhancement with deep complex U-Net,'' 2019, *arXiv:1903.03107*.

[39] C. M. Hyun, H. P. Kim, S. M. Lee, S. Lee, and J. K. Seo, ''Deep learning for undersampled MRI reconstruction,'' *Phys. Med. Biol.*, vol. 63, no. 13, Jun. 2018, Art. no. 135007.

[40] Z. Iqbal, D. Nguyen, G. Hangel, S. Motyka, W. Bogner, and S. Jiang, ''Super-resolution ^{1}H magnetic resonance spectroscopic imaging utilizing deep learning,'' *Frontiers Oncol.*, vol. 9, p. 1010, Oct. 2019.

[41] C.-H. Pham, C. Tor-Díez, H. Meunier, N. Bednarek, R. Fablet, N. Passat, and F. Rousseau, ''Multiscale brain MRI super-resolution using deep 3D convolutional networks,'' *Computerized Med. Imag. Graph.*, vol. 77, Oct. 2019, Art. no. 101647.

[42] S. Chatterjee, A. Sciarra, M. Dünnwald, R. V. Mushunuri, R. Podishetti, R. N. Rao, G. D. Gopinath, S. Oeltze-Jafra, O. Speck, and A. Nürnberger, ''ShuffleUNet: Super resolution of diffusion-weighted MRIs using deep learning,'' in *Proc. 29th Eur. Signal Process. Conf. (EUSIPCO)*, Aug. 2021, pp. 940–944.

[43] Y. Huang, W. Wang, and L. Wang, ''Video super-resolution via bidirectional recurrent convolutional networks,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 1015–1028, Apr. 2018.

[44] W. Yang, J. Feng, G. Xie, J. Liu, Z. Guo, and S. Yan, ''Video super-resolution based on spatial–temporal recurrent residual networks,'' *Comput. Vis. Image Understand.*, vol. 168, pp. 79–92, Mar. 2018.

[45] B. N. Chiche, A. Woiselle, J. Frontera-Pons, and J.-L. Starck, ''Stable long-term recurrent video super-resolution,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 827–836.

[46] M. Hayat, S. Aramvith, and T. Achakulvisut, ''Combined channel and spatial attention-based stereo endoscopic image super-resolution,'' in *Proc. TENCON IEEE Region 10 Conf. (TENCON)*, Oct. 2023, pp. 920–925.

[47] M. Hayat and S. Aramvith, ''E-SEVSR—Edge guided stereo endoscopic video super-resolution,'' *IEEE Access*, vol. 12, pp. 30893–30906, 2024.

[48] J. Johnson, A. Alahi, and L. Fei-Fei, ''Perceptual losses for real-time style transfer and super-resolution,'' in *Proc. Eur. Conf. Comput. Vis.*, Cham, Switzerland: Springer, 2016, pp. 694–711.

[49] L. A. Gatys, A. S. Ecker, and M. Bethge, ''Image style transfer using convolutional neural networks,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2414–2423.

[50] V. Ghodrati, J. Shao, M. Bydder, Z. Zhou, W. Yin, K.-L. Nguyen, Y. Yang, and P. Hu, ''MR image reconstruction using deep learning: Evaluation of network structure and loss functions,'' *Quant. Imag. Med. Surg.*, vol. 9, no. 9, pp. 1516–1527, Sep. 2019.

[51] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, ''The unreasonable effectiveness of deep features as a perceptual metric,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 586–595.

[52] Z. Wang, J. Chen, and S. C. H. Hoi, ''Deep learning for image super-resolution: A survey,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 10, pp. 3365–3387, Oct. 2021.

[53] A. Odena, V. Dumoulin, and C. Olah, ''Deconvolution and checkerboard artifacts,'' *Distill*, vol. 1, no. 10, p. e3, Oct. 2016.

[54] A. E. Kavur, ''CHAOS challenge–combined (CT-MR) healthy abdominal organ segmentation,'' *Med. Image Anal.*, vol. 69, Apr. 2021, Art. no. 101950.

[55] F. Pérez-García, R. Sparks, and S. Ourselin, ''TorchIO: A Python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning,'' *Comput. Methods Programs Biomed.*, vol. 208, Sep. 2021, Art. no. 106236.

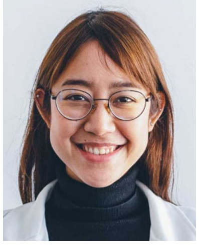
[56] S. Chatterjee, ''Soumickmj/Mrunder: Initial release,'' Eur. Org. Nucl. Res., OpenAIRE, Zenodo, Version v0.1, CERN, Geneva, Switzerland, Jun. 2020.

[57] S. Chatterjee, M. Breitkopf, C. Sarasaen, H. Yassin, G. Rose, A. Nürnberger, and O. Speck, ''ReconResNet: Regularised residual learning for MR image reconstruction of undersampled Cartesian and radial data,'' 2021, *arXiv:2103.09203*.

[58] A. Paszke, ''PyTorch: An imperative style, high-performance deep learning library,'' in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, 2019, pp. 8024–8035.

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

SOUMICK CHATTERJEE (Member, IEEE) received the B.C.A. degree from Punjab Technical University, India, in 2013, the M.Sc. degree in computer science from the St. Xavier's College, Kolkata, India, in 2017, with a master's project in text classification using machine learning, and the Ph.D. degree (summa cum laude) in computer science, specializing in artificial intelligence applied to medical physics from Otto von Guericke University Magdeburg, Germany, in 2022, with this thesis titled ''Reducing Artefacts in MRI using Deep Learning: Enhancing Automatic Image Processing Pipelines.'' He began his career with tech entrepreneurship, co-founding Supernova Techlink, Kolkata, and serving as the Chief Software Architect, from 2011 to 2017. Subsequently, he worked as a Research Scholar with the Data and Knowledge Engineering Group (Faculty of Computer Science) and the Department of Biomedical Magnetic Resonance (Faculty of Natural Sciences), Otto von Guericke University Magdeburg, from 2018 to 2022. He is currently a Computer Scientist. He is a Postdoctoral Researcher with the Glastonbury Group, Genomics Research Centre, Human Technopole, Milan, Italy. His research interests include machine learning, image processing, MRI, applied physics, and statistical genetics.



CHOMPUNUCH SARASAEN received the B.Sc. degree in radiological technology and the M.Eng. degree in biomedical engineering in Thailand, and the Ph.D. degree in engineering from Otto von Guericke University Magdeburg, Germany, in 2024. She is a Biomedical Engineer and Data Scientist. Her research interests include image processing, image reconstruction, and image registration based on synchrotron micro- and nano-CT.

GEORG ROSE received the Ph.D. degree in statistical physics from the University of Düsseldorf, Germany, with a focus on field theory and mean-field approximation. As a Postdoctoral Researcher with the Department of Neurology, he researched brain modeling and stroke management strategies. In 1995, he joined Philips Research Laboratories, Aachen, Germany, working on tomographic and functional imaging, while continuing his stroke management research. Since 2006, he has been a Full Professor and the Chair for healthcare telematics and medical engineering with Otto von Guericke University Magdeburg, Germany, with a focus on medical imaging, medical electronics, and brain–machine interfaces. He is currently a Full Professor with Otto von Guericke University Magdeburg.

ANDREAS NÜRNBERGER (Member, IEEE) received the Diploma degree in computer science from the Technical University of Braunschweig, Germany, in 1996, and the Ph.D. degree in computer science from Otto von Guericke University Magdeburg, Germany, in 2001. Following the Ph.D. degree, he joined the University of California at Berkeley as a Postdoctoral Fellow, where he worked on adaptive soft computing and visualization techniques for information retrieval systems in collaboration with BTexact Technologies, U.K. Since 2003, he has held the position of a Junior Professor for information retrieval. Since 2007, he as a tenured Professor of data and knowledge engineering with Otto von Guericke University Magdeburg. He is currently a Computer Scientist and an Emmy Noether Fellow with German Science Foundation, a Senior Professor of computer science, and the Head of the Research Group "Data and Knowledge Engineering," Faculty of Computer Science, Otto von Guericke University Magdeburg, Germany. Additionally, he served as the Vice Dean (Research), from 2012 to 2014, and the Dean, from 2014 to 2020, of the Faculty of Computer Science at his university. His research interests include machine learning and data mining, human–computer interaction, information structuring and organization, cross-lingual and multilingual information retrieval, image and video retrieval, and image processing. His leadership extends beyond his research group, having served as the Vice President for the conferences and meetings and the Vice President of IEEE Transactions on Human-Machine Systems and the IEEE Systems, Man, and Cybernetics (SMC) Society. His editorial contributions include roles with IEEE Transactions on Cybernetics and *International Journal of Knowledge-Based and Intelligent Engineering Systems*.

OLIVER SPECK is a physicist, senior professor of Physics, and head of the Department of Biomedical Magnetic Resonance in the Faculty of Natural Sciences at Otto von Guericke University Magdeburg in Germany. He completed his diploma thesis in Physics at the University Hospital Freiburg and the University of Freiburg's Department of Physics in 1994, and pursued his PhD at the same institutions from 1994 to 1997. Subsequently, he worked as a research associate at the University Hospital Freiburg's MR Centre before moving to the UCLA Research and Education Institute in Torrance, California, USA, where he worked until 1999. He returned to the University Hospital Freiburg as a staff research associate and deputy head of the section of MR-Physics, leading the Emmy-Noether-Group from 1999 to 2005. He achieved habilitation in Medical Physics at the Department of Diagnostic Radiology, University of Freiburg in 2005. Since 2006, he has been the Professor for Biophysics and the Director of the Department of Biomedical Magnetic Resonance at Otto von Guericke University Magdeburg. His research interests include motion correction, undersampled reconstruction, and ultra-highfield MRI.