Projection-Based Cascaded U-Net Model for MR Image Reconstruction

Amir Aghabiglou[a], Ender M. Eksioglu[b,*]

[a]*Graduate School of Science, Engineering and Technology, Istanbul Technical University, Istanbul, Turkey*

[b]*Electronics and Communication Engineering Department, Istanbul Technical University, Istanbul, Turkey*

Abstract

Background and Objective: Recent studies in deep learning reveal that the U-Net stands out among the diverse set of deep models as an effective network structure, especially for imaging inverse problems. Initially, the U-Net model was developed to solve segmentation problems for biomedical images while using an annotated dataset. In this paper, we will study a novel application of the U-Net structure for the important inverse problem of MRI reconstruction. Deep networks are particularly efficient for the speed-up of the MR image reconstruction process by decreasing the data acquisition time, and they can significantly reduce the aliasing artifacts caused by the undersampling in the k-space. Our aim is to develop a novel and efficient cascaded U-Net framework for reconstructing MR images from undersampled k-space data. The new framework should have improved reconstruction performance when compared to competing methodologies.

*Corresponding author

Email addresses: `aghabaiglou17@itu.edu.tr` (Amir Aghabiglou), `eksioglue@itu.edu.tr` (Ender M. Eksioglu)

Methods: In this paper, a novel cascaded framework utilizing the U-Net as a sub-block is being proposed. The introduced U-Net cascade structure is applied to the magnetic resonance image reconstruction problem. The connection between the cascaded U-Nets is realized in the form of a recently developed projection-based updated data consistency layer. The novel structure is implemented in the PyTorch environment, which is one of the standards for deep learning implementations. The recently created fastMRI dataset which forms an important benchmark for MRI reconstruction is used for training and testing purposes.

Results: We present simulation results comparing the novel method with a variety of competitive deep networks. The new cascaded U-Net structure's PSNR performance stands on average 1.28 dB higher than the baseline U-Net. The improvement, when compared to the standard CNN, is on average 3.32 dB.

Conclusions: The proposed cascaded U-Net configuration results in an improved reconstruction performance when compared to the CNN, the cascaded CNN, and also the singular U-Net structures, where the singular U-Net forms the baseline reconstruction method from the fastMRI package. The use of the projection-based updated data consistency layer also leads to improved quantitative (including SSIM, PSNR, and NMSE results) and qualitative results when compared to the use of the conventional data consistency layer.

Keywords: Magnetic resonance imaging, Image reconstruction, Deep learning, Cascaded networks, U-Net, Updated data consistency

1. Introduction

Magnetic resonance imaging (MRI) is an important modality for clinical diagnosis because it provides high resolution for soft tissue analysis [1]. However, the data acquisition process for MR images is relatively long when compared to other imaging methods, because the data acquisition occurs in the k-space (Fourier) domain in a point-wise and successive manner. Moreover, the speed of this sequential data acquisition process is hampered by physiological and hardware constraints [2]. Therefore, the MR image acquisition becomes susceptible to motion artifacts. MR image reconstruction constitutes a trade-off among the acquired amount of k-space data, the quality of the reconstructed images, and reconstruction speed. Deep networks have facilitated a novel family of reconstruction algorithms from undersampled k-space data. Considering all these issues, a variety of deep learning (DL) and MRI acceleration methods like compressed sensing (CS)-MRI [2, 3] have been proposed, with the major research goal being to decrease the acquisition time without reducing image quality. Accordingly, effective CS-MRI models which utilize denoising as a sub-step of iterative reconstruction have been developed [4, 5]. In this setting denoisers such as block matching 3D (BM3D) are employed to enforce the sparse prior onto the reconstructed MR images. Using these iterative methods, an opportunity has been provided to improve MRI reconstruction performance by applying a varying regularization parameter. However, the sparsity regularized CS-MRI based reconstruction methods are computationally expensive and rather slow due to the iterative nature of optimization solutions [6].

2. Related Works

Newly developed deep learning based frameworks for MR reconstruction enable the researchers to tackle the difficulties related to modeling-based iterative methods. One of these deep frameworks which gained attention in solving the inverse problem of MRI reconstruction was the study performed by Schlemper *et al.* [7, 8]. The proposed algorithm incorporates a "Deep Cascade" of Convolutional Neural Networks with interleaved data consistency (DC) stages. The Deep Cascade-CNN (DC-CNN) performs the back-propagation process through the DC block by deriving the Jacobian of this layer. Motivated by the DC-CNN of [7], a cascaded CNN model that utilizes a novel, projection-based updated data consistency layer has been introduced in [9]. Deep learning based methods significantly accelerate the MRI reconstruction process (e.g. for Deep Cascade-MRI each image reconstruction takes around 23ms). Recently, with inspiration from the Deep Cascade model, a compound cascade algorithm has been proposed by Qiao *et al.* [10]. They have applied both data-based and model-based algorithms to achieve better performance compared to other state-of-the-art MRI reconstruction models. The authors have used the iterative Approximate Message Passing (AMP) algorithm alongside the convolutional neural network to de-alias and reconstruct MR images. These two models were connected with data consistency layers to each other. Another study that leveraged DC-layer was conducted by YanWu *et al.* [11]. Their proposed SAT-Net (self-attention convolutional neural network) was developed by using a self-attention CNN and was applied on cartilage images. This deep residual CNN uses long-range dependencies to reconstruct MR images. One more compound network has been put forward by combining the convolutional neural network and parallel imaging. By using this framework, the authors have reconstructed real-time MR images [12]. Using a similar approach, recently a wavelet-based deep cascaded CNN has been put forward by modifying the U-Net model architecture [13]. In this study, they have replaced pooling and unpooling layers in the U-Net model with discrete wavelet transform and inverse wavelet transform, respectively.

MR image reconstruction actually proceeds towards the minimization of a loss function calculated by using the fully-sampled k-space data as the target. The input to the model is the under-sampled k-space data (or the initial rough image estimate calculated using zero-filling) [14]. Based on the same idea, to accelerate the MR imaging S. Wang *et al.* [15] suggested an off-line CNN to identify the connection between MR images that have been acquired from the under-sampled k-space data and the target ground truth data. Their network showed that it can reinstate the details and reconstruct structures that are lost in the masking step.

Besides CNN, U-Net models have been proposed in the literature for automatic segmentation purposes in MR images [16, 17]. Attention gated (AG) networks were introduced in [18] for medical image analysis. In this work, the AG model is integrated into CNN and U-net architectures for MR image classification and segmentation problems. Squeeze-and-Excitation block (SE block) was proposed for the first time in [19]. The SE block updates the features in a channel-wise manner without increasing the computational cost. The USE-Net which employs SE blocks inside a U-Net structure was offered in [20] for zonal MR image segmentation. In this particular study, a combination of three datasets was used for training and testing the model [20].

Despite the fact that the U-Net was originally built for image segmentation problems in bio-imaging, the U-Net has been used as a strong baseline for MR image reconstruction. The U-Net has a strong ability of preserving high-resolution features through concatenation in the up-sampling process. In a recent study [21], a deep learning algorithm based on k-space interpolation using the U-Net has been proposed, where the input and output are in the k-domain and are complex-valued. Hence in this work, the undersampled k-space measurements have been utilized without performing the Inverse Discrete Fourier transform (IDFT). In another recent work, a framework based on the U-Net for speeding up MR imaging with sub-Nyquist sampling strategies has been developed [22]. Here the zero-filling image is used as the input to the U-Net model, and the reconstruction is improved by employing the data consistency layer, which provides enforced consistency with the measured k-space data [22]. In [23] on the other hand, CNN models based on the Least Absolute Deviations (LAD-L1) and Least Square Errors (LSE-L2) loss functions have been compared in their capability to achieve fast cardiac MRI reconstruction. The authors showed that their network is almost 150x faster than the compared CS reconstruction method, with even better image quality and for larger acceleration factors [23].

In a comparative study, two deep learning models including residual network and U-Net for MR image reconstruction have been evaluated based on diverse loss functions by using a cardiac MRI dataset. Results reveal that both of the architectures result in similar outputs, and additionally, they found that the utilization of a perceptual loss function outperformed the Dssim, L1, and L2 loss functions [14]. In another study, a variational network that learns the effective priors has been offered, with different acceleration factors, for the purpose of shortening the acquisition and reconstruction time of MR knee images. Additionally, their proposed variational network preserved the original MR image appearance plus pathologies that were not included in the training data set [24].

When adopting deep learning for MR image reconstruction, one should carefully evaluate various features like the effects of loss function choices and network architecture. In [25], the authors put forward a cross-domain CNN called KIKI-net by reconstructing the final output image by forward propagating the under-sampled k-space data through the entire network. They revealed that in terms of restoring tissue structures and removing aliasing artifacts, the combination of K-net and I-net performs much better than the single-domain convolutional neural networks, with an average PSNR improvement of 2.29 dB [25]. Furthermore, another dual-domain architecture has been offered in [26]. In this paper, the authors designed a hybrid framework termed W-net, which operates on both k-space and image domains. Their model in the k-space domain contains a residual U-Net that works with complex values. The structure includes another U-Net in the image domain working with real values. However, in terms of quantitative results, the deep-cascade framework outperforms this model. From a qualitative point of view, the deep-cascade model reconstructs images with better visual quality.

Another state-of-the-art algorithm is the Primal-Dual-Net (PD-net) which was applied to MR images in [27]. This paper implements unrolling of the cross-domain model for MR image reconstruction in the case of aggressively under-sampled k-space data. [28] offers an enhanced recursive residual network that has superior reconstruction abilities with clearly restored structural features and high image quality. Lately, a hybrid CNN has been put forward for MR image reconstruction based on a single-image super-resolution technique. By leveraging the HybridNet, high-quality images have been achieved for three different datasets. The reconstruction PSNR is higher when compared to other deep learning based approaches [29].

Regarding all aforementioned issues, DL-based model realizations from the literature tend to outperform conventional, modeling based reconstruction methods in both reconstruction quality and speed [30]. In this work on the other hand, we introduce a cascaded network structure employing U-Net networks in a cascaded, unfolded framework. The U-Nets are linked together through the novel updated data consistency (UDC) layer. The introduced new model is trained and tested on the recently introduced fastMRI benchmark dataset [31]. The introduced model's performance is compared with baseline U-Net (Fig. 1) [31, 32, 33], cascaded CNN [7, 8] and projection-based cascaded CNN [9]. The results indicate that the developed cascaded U-Net structure with intermediary UDC layers outperforms all of the mentioned deep learning based MRI reconstruction network variants.

This paper main contributions can be summarized as follow:

- Cascade structures were recently introduced into the MRI reconstruction literature using conventional CNN networks. In this study, we have proposed a novel cascaded U-Net framework that also utilizes recently introduced projection-based structures in the encoding-decoding

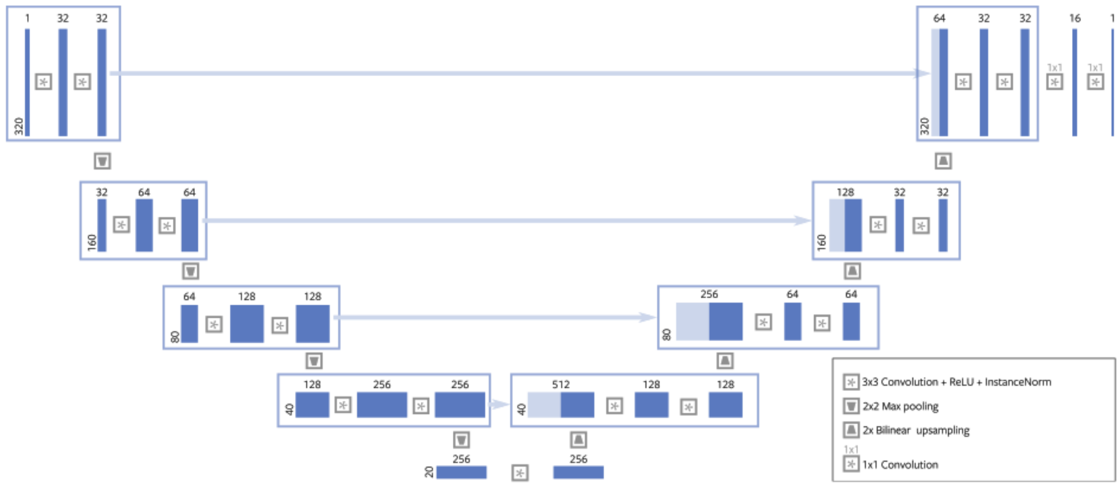

Figure 1: Single-coil baseline U-Net architecture [31]. A multi-channel feature map is denoted by the blue rectangular boxes. The slide size is presented at the lower-left edge of the boxes. The number at the top of each box shows the number of channels.

pathway.

- We replaced the data fidelity block of the MRI reconstruction pipeline with a recently introduced updated data consistency layer.

- In literature, cascade frameworks for MRI reconstruction were trained and tested with rather small datasets, some including up to only 10 fully sampled cardiac MR images. We benchmarked cascade and projection-based cascade structures using the recently introduced and rather large fastMRI dataset. This dataset was firstly developed by a joint group of New York University (NYU) researchers together with a Facebook AI Research (FAIR) team for the fastMRI image reconstruction challenge. This dataset has gained quite attraction in the MRI reconstruction community and has been employed as a benchmark in quite a few recent studies.

In Section 2, the contribution and novelty of this work in comparison to previous deep networks are discussed. Section 3 introduces the details of the proposed novel network structures. Section 4 provides detailed information about the dataset, experimental setup, and evaluation parameters. The quantitative and visual image reconstruction results are presented in the subsequent section. Finally, the conclusions are provided in the final section.

3. Method

3.1. A New Model for Deep Learning MRI Reconstruction: Cascaded U-Nets

In previous works from the literature, MR image reconstruction methodologies based on Convolutional Neural Networks and U-Nets have been offered as successful models. In the MRI reconstruction setting, data is acquired through the following equation:

$$y = \mathscr{F}_{\Omega} x_{orig} \tag{1}$$

In this equation, y indicates the observed data in the k-domain, where $\mathscr{F}_{\Omega}$ is the undersampled Fourier transform modeling the data acquisition. x_{orig} on the other hand, specifies the vectorized form for the ground truth image. One of the efficient deep models for solving the inverse problem of reconstructing x_{orig} utilizes cascaded CNNs (DC-CNN), and its details are provided in [7, 8]. This DC-CNN model places a data fidelity (or data consistency) layer (DC layer) after each network. Each single CNN reconstructs an intermediary output which goes through the succeeding DC layer as an input. Then, the DC layer enforces the observation data in the k-space domain onto the reconstructed output, with the procedure given in the below equation:

$$x_{out} = \mathscr{F}^{-1}\{\overline{M} \circ (\mathscr{F} x_{in}) + \widehat{y}\} \tag{2}$$

Here, $\mathscr{F}$ and $\mathscr{F}^{-1}$ indicate 2D Discrete Fourier Transform (2D-DFT) and its inverse operator (2D-IDFT), respectively. $\overline{M}$ is the inverse of mask function which has been used for undersampling the ground truth data in the k-space domain. Moreover, "$\circ$" and x_{in} are the point-wise multiplication operator and the input image (intermediary reconstructed image), respectively. Here, $\widehat{y}$ is the ground truth observation in k-space domain which is defined in (1). A novel, projection-based updated DC layer on the other hand was introduced in [9]. The UDC layer incorporates a secondary output, r_{out}, which corresponds to the estimate of the unobserved part of Fourier information. The UDC layer creates both a rectified output image x_{out} in accordance with the regular DC layer, and also another residual image r_{out} which is defined as follow.

$$r_{out} = \mathscr{F}^{-1}\{\overline{M} \circ (\mathscr{F} x_{in})\} \tag{3}$$

Fig. 2 gives a pictorial description of this updated data consistency layer. These secondary outputs from the UDC layers, alongside the final intermediary reconstructed image, get collected in the concatenation layer near the end of the structure via skip connections.

In this work, we introduce a cascaded deep learning framework that incorporates the U-Net as the building block. The cascaded U-Net model with standard DC layers is given in Fig. 3(a). We also develop a cascaded model

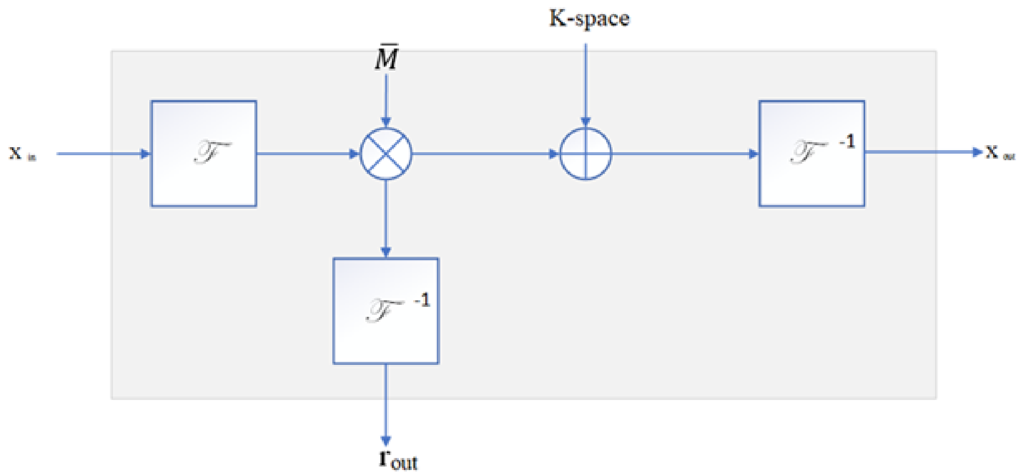

Figure 2: Updated data consistency layer (UDC) block diagram [9].

with UDC layers acting as the intermediaries. This novel cascaded U-Net model utilizing UDC layers on the other hand is shown in Fig. 3(b). At the final stage of this novel structure, the last network reconstructs the final image estimate by using the input feature maps coming from the concatenation layer, which brings together the secondary outputs from the UDC layers and the image estimate formed before this final stage.

In this paper, to provide a fair overview and comparison, multiple models including standard CNN, cascaded CNN, projection-based cascaded CNN, and U-Nets were trained and validated using the fastMRI dataset. Two different types of masking functions have been utilized with multiple undersampling ratios. The novel models proposed in this work include the cascaded U-Net structure and the projection-based cascaded U-Net structure. These novel structures are depicted in Fig. 3. Their performance results are compared with the previously mentioned models from the literature. In this

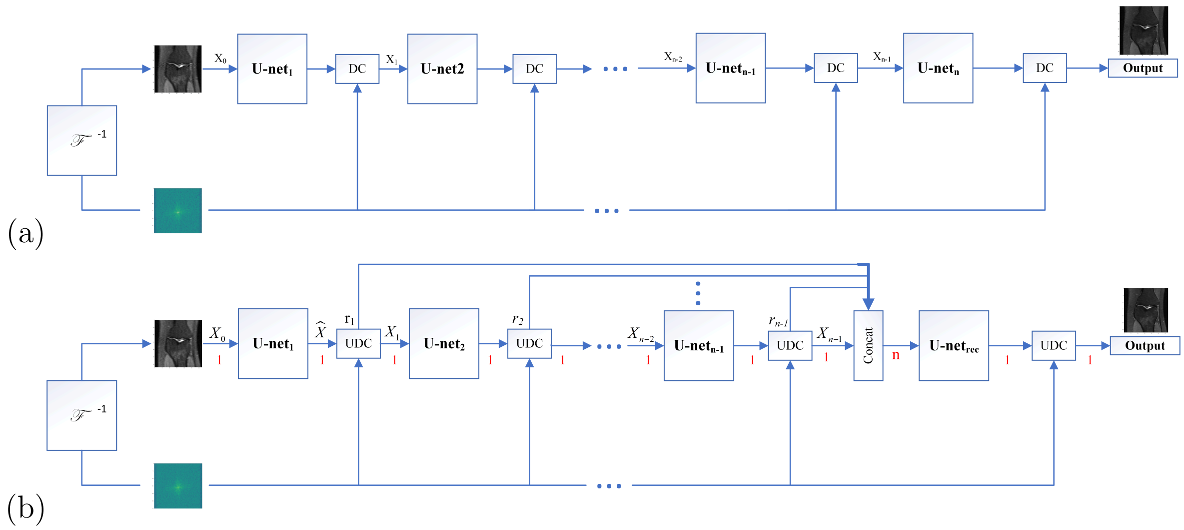

Figure 3: The block diagram of (a) cascaded U-Net with standard DC layer, and (b) projection-based cascaded U-Net with updated DC layers. The input and output channel sizes are shown with red values.

simulation setting, single-channel ground truth images with real values are undersampled using random and equispaced Cartesian mask functions. Subsequently, zero-filled (ZF) images are given as input to the cascaded blocks of the U-Net with residual network blocks and intermediary DC or UDC layers. However, in the projection-based cascade U-Net, the fifth block is a reconstruction module with multiple feature maps at its input. The secondary outputs (r_{out}) from the initial UDC layers concatenated with the primary output of the final UDC layer forms the input to this last module. Finally, the loss is calculated between the target image and the final reconstructed image estimate.

In Algorithm 1, we give an algorithmic representation for the cascaded network as introduced in this paper. Here $\mathcal{N}$ denotes the deep network as

for $i = 1 : n-1$ **do**

$\widehat{x}_i = \mathcal{N}(x_{i-1})$

$(x_i, r_i) = \text{UDC}(\widehat{x}_i, y)$

end for

$\mathbf{R} = [r_1, r_2, ..., r_{n-1}]$

$\widehat{x}_n = \mathcal{N}(x_{n-1}, \mathbf{R})$

$x_n = \text{UDC}(\widehat{x}_n, y)$

Algorithm 1: The steps for the projection-based cascaded network using UDC layers.

utilized in the recursive, cascaded structure. In our implementation, we take $\mathcal{N}$ to be a U-Net as defined in Fig. 1. Here, UDC is the updated data consistency layer. $\widehat{x}_i$ are the intermediary outputs of the cascaded blocks. Furthermore, r_i are the secondary residual outputs of the UDC layers. x_i is the rectified image output of the UDC layer. To the best of our knowledge, the proposed scheme gives the first use of U-Nets in a cascaded structure. The UDC is also utilized together with the U-Nets for the first time. This novel cascaded structure which employs the U-Net together with the UDC results in very competitive MRI reconstruction performance.

3.2. Architecture

In this study, all deep networks based on U-Net and CNN have been trained using single-channel MR images with real values. As illustrated in Fig. 3, initially inverse Fast Fourier Transform will be applied to the undersampled k-space data. Then, the obtained zero-filling (ZF) images will be normalized as a pre-processing step. These ZF input images will get processed by the deep networks to generate high-quality reconstructed images. Meanwhile, all image slices are center cropped to the size of 320x320 to make all slices from the dataset have the same dimensions.

3.2.1. CNN based networks

The utilized CNN network architecture is the same as the one mentioned in the literature [7, 8]. The model is detailed in Fig. 4, and it includes five convolution layers. In the first convolutional layer, the input and output channel sizes are set to one and 64 respectively, with the filter kernel size equaling three. Subsequent three layers have both input and output channel sizes equal to 64, with the filter kernel size being 3. The final layer, namely C_{rec} will project the 64 input maps into a single image output. Throughout this structure, all convolution layers are followed by ReLU units (rectified linear units) acting as the activation function. Moreover, the input and output of the network are to get a residual network which can improve performance. For the training of all the CNN-based networks, Adam optimizer was utilized with a learning rate of 10^{-4}, weight decay of 10^{-7}, and beta values equal to (0.9, 0.999).

3.2.2. U-Net based networks

In our realizations, we have benefited from the U-Net models for MRI reconstruction as developed and shared by the Facebook AI Research (FAIR) team [31]. The employed U-Net pipeline includes two different deep convolutional networks in the down-sampling and up-sampling paths. The down-sampling path has 2 blocks of 3x3 convolutions. It employs ReLU as the activation function. These blocks perform down-sampling by using max-

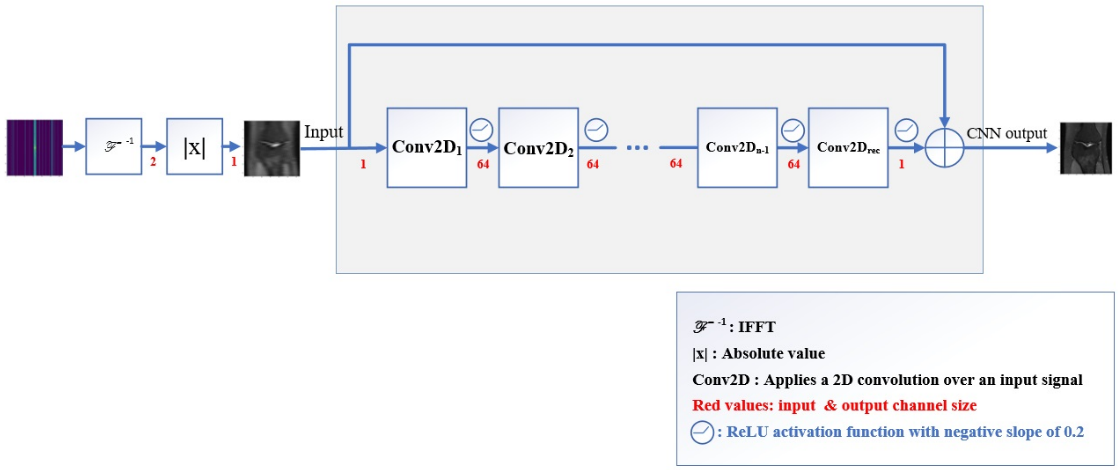

Figure 4: The reconstruction simulation pipeline using the standard Convolutional Neural Network (CNN) architecture.

pooling layers with stride 2 to halve the resolution. On the other side, the Up-sampling layer has the same structure as the down-sampling, although it uses bilinear layers to double each spatial dimension. At the output layer, it has a 1x1 convolution layer which makes the output channels to be one. Unlike conventional CNN, our U-Net model has 2 outputs. In this model, the second output stores residual image which will be carried via skip connection to the input layer of the final U-Net. This final network will create the ultimate reconstruction based on the residual images and the intermediary rectified image. Furthermore, as a depth of the network, the cascaded and projection-based cascade U-Net network consists of five U-Net (n_c =5) models that each one is connected to a residual layer. As shown in Fig. 3(b) the output of these layers goes inside a UDC layer in order to save the data fidelity.

At the training stage, the deep models are trained by using 973 image volumes. The gradient is calculated as a running average of its recent magnitude (RMSProp) and the error term used is the mean element-wise absolute value difference (L1 loss function) between the reconstructed image and the desired output (ground truth image). Furthermore, during each batch processing step of the training, a random or equispaced (uniform) Cartesian sampling mask is selected to realize the k-domain undersampling. The number of U-Net pooling layers is chosen as four, and the dropout probability is set to zero by default. Additionally, the learning rate is set equal to 0.001. The period for learning rate and the multiplicative factor of learning rate decay are set to 40 and 0.1, respectively. Moreover, the strength of weight decay regularization is chosen as equal to zero.

3.3. Simulation Setting

3.3.1. Dataset

Deep convolutional networks have been proven their performance versus competing state-of-the-art methods in various applications [34]. Hence deep learning based methodologies have found widespread usage, however, their applicability is challenged by the availability of training datasets for the particular application under consideration [35, 36]. In the MRI reconstruction setting, some studies tried to solve this problem by eliminating the need for pretraining. As an example, [37] combines the DL-based model with compressed sensing ideas to reconstruct MR images without the need for pretraining, to alleviate the dependency on training datasets. In this reference-driven model, they improved the accuracy and reconstructed MR image quality by leveraging k-space data fidelity. However, the more common effective approach has been the creation and employment of MR image datasets for training deep learning based reconstruction structures aimed at accelerated MR image reconstruction. The fastMRI dataset presents a very recent and rather complete example for an MR image dataset aimed at MRI reconstruction, complete with the training and testing code for a baseline network [31]. This dataset includes the data for different types of MR images in various formats. Both single and multi-coil images are included. The image data are stored in the form of fully sampled k-space information, reconstructed ground-truth images from the fully sampled data, and DICOM format images.

In this paper, we have employed the single-coil image dataset. There are a total of 1372 single-coil MRI image volumes in the fastMRI dataset. These volumes are divided into four groups of training, validation, test, and challenge subsets. Table 1 gives a rundown of the number of volumes and the total number of image slices in these four groups.

Table 1: Number of image volumes and image slices in the fastMRI single-coil dataset [31]

Subset name	Volumes	Slices
Training	973	34742
Validation	199	7135
Test	108	3903
Challenge	92	3305

Each volume includes both the fully sampled k-space data and the corresponding fully sampled single-coil image reconstruction. We start our simulation from the fully sampled and real-valued ground truth images, by subsampling them in the k-space to model the MR data acquisition process. The ZF reconstructions from the subsampled data are used as the input to the networks. Undersampling is performed with three different acceleration strategies and two different masking schemes. Undersampling scenarios include the 4-fold subsampling, 8-fold subsampling, and the aggregate subsampling. In aggregate subsampling, the sampling ratio is randomly set to either four or eight with equal probability for each. Setting the acceleration factor (subsampling ratio) equal to four or eight will result in keeping only 25% or 12.5% of the k-space information, respectively. The ground truth image slices are employed as the target image of the networks, where the network backpropagation error is calculated as the L1 pixel-wise difference between the reconstructed image and target image.

3.3.2. Experimental setup

For the reconstruction simulations of this paper, we utilized Python 3.6 together with Pytorch version 1.4.0. We took advantage of two GPUs in the form of GeForce RTX 2080 Ti with 11GB of memory for each. The batch size was selected as 16 for baseline U-Net and standard CNN realizations. However, the batch size was reduced to four in the case of the cascaded networks. Meanwhile, the number of training epochs was set to 20, which was observed to be adequate for all of the different network realizations.

3.3.3. Evaluation Methodology

We tabulate the results evaluating the performance of the different models in the form of three metrics, which include normalized mean squared error (NMSE), peak signal to noise ratio (PSNR), and structural similarity index measure (SSIM). NMSE metric is calculated by measuring the pixel-wise difference between the ground truth and network output images. PSNR is evaluated as the ratio between the maximum possible power of the signal (image intensity across a volume) and the power of the distorting noise that affects the fidelity of its representation. SSIM on the other hand is a perceptual index that measures similarity between images by using mutual dependencies between neighboring pixels. NMSE between the network output image R_i and the desired output G_t is given by the following expression.

$$\mathrm{NMSE}(R_i, G_t) = \frac{\|R_i - G_t\|_2^2}{\|G_t\|_2^2} \tag{4}$$

Here, $\|\cdot\|_2^2$ designates the squared Euclidean norm. The PSNR between R_i and G_t is defined as follows.

$$\mathrm{PSNR}(R_i, G_t) = 10\ \log_{10}\frac{max(G_t)^2}{\mathrm{MSE}(R_i, G_t)} \tag{5}$$

Here, $\mathrm{MSE}(R_\mathrm{i},G_\mathrm{t})$ is the mean square error between R_i and G_t, which is given by:

$$\mathrm{MSE}(R_i, G_t) = \frac{1}{n}\|R_i - G_t\|_2^2 \tag{6}$$

Here, n indicates the number of entries in the ground truth volume G_t. It should be mentioned that lower values for PSNR indicate inferior reconstruction. On the contrary, a lower value for NMSE shows better reconstruction. The SSIM between two image patches P_1 and P_2 is given by:

$$\mathrm{SSIM}(P_1, P_2) = \frac{(2\mu_{P_1}\mu_{P_2} + C_1)(2\sigma_{P_1P_2} + C_2)}{(\mu_{P_1}^2 + \mu_{P_2}^2 + C_1)(\sigma_{P_1}^2 + \sigma_{P_2}^2 + C_2)} \tag{7}$$

Here, μ_{P_1} and μ_{P_2} are the average pixel values of P_1 and P_2, respectively. ${\sigma^2}_{\mathrm{P}_1}$ and ${\sigma^2}_{\mathrm{P}_2}$ on the other hand, are the corresponding pixel variances. Additionally, $\sigma_{\ \mathrm{P}_1\ \mathrm{P}_2}$ indicate the covariance value between these two patches. C_1 and C_2 are defined as below to stabilize the division.

$$C_1 = (0.01L)^2 \tag{8}$$

$$C_2 = (0.03L)^2 \tag{9}$$

L is defined as:

$$L = max(G_t) \tag{10}$$

4. Results

4.1. Quantitative results

In this work, we have implemented the standard CNN and its cascaded variant as used in [7, 8]. Another important deep reconstruction paradigm from the literature, which we have realized is the baseline U-Net as implemented in the fastMRI framework [31]. In all of these models, we have employed the regular DC layer for the data fidelity block, as is done in all the original implementations. In addition to these models, we have also implemented the cascaded CNN with UDC layers acting as the linking blocks. The novel models advanced in this paper are the cascaded U-Net structures, firstly with the regular DC layer and secondly with the updated DC layer as the linking intermediary blocks. The evaluated quantitative results include PSNR, SSIM, and NMSE. The results are tabulated in Tables 2 and

3, for different mask types and subsampling ratios. From these results, it becomes apparent that projection-based cascaded U-Net with UDC layers performs the best among all the evaluated variants, with the cascaded U-Net using DC layers coming as second. The cascaded U-Net with UDCs has improved performance in all of the calculated metrics. This proposed network increases averagely the PSNR performance by almost 1.28 dB and 3.32 dB when compared to the baseline U-Net and the standard CNN, respectively. Furthermore, when we compare Tables 2 and 3, we notice that almost all of the models get better results with the equispaced, uniform masking function. We can deduce that these two tables verify each other's results when we compare the difference among models.

Table 2: Simulation results for various models with random subsampling mask

Acceleration	4-fold				8-fold				Aggregate				Runtime (s)
Network	Loss	NMSE	SSIM	PSNR	Loss	NMSE	SSIM	PSNR	Loss	NMSE	SSIM	PSNR	
Zero-filling	-	0.0416	0.711	29.876	-	0.077	0.603	26.921	-	0.061	0.651	28.217	0
CNN[7]	0.308	0.034	0.755	30.880	0.451	0.069	0.637	27.462	0.386	0.054	0.688	28.867	0.34
CNN+DC	0.307	0.033	0.759	31.012	0.450	0.067	0.638	27.609	0.385	0.052	0.689	29.041	0.344
Cascade CNN [7, 9]	0.280	0.0265	0.790	32.412	0.417	0.0548	0.654	28.639	0.355	0.042	0.715	30.152	1.75
Projection-based cascaded CNN [9]	**0.272**	**0.0247**	**0.801**	**32.872**	**0.400**	**0.049**	**0.678**	**29.213**	**0.342**	**0.038**	**0.731**	**30.704**	**1.83**
Baseline U-Net [31]	0.281	0.0268	0.785	32.419	0.380	0.043	0.692	29.952	0.335	0.036	0.732	30.934	0.53
Baseline U-Net + DC layer (new)	0.270	0.024	0.801	32.968	0.373	0.040	0.701	30.307	0.326	0.033	0.744	31.358	0.55
Cascade U-Net (new)	0.261	0.0224	0.813	33.585	0.356	0.035	0.715	31.185	0.314	0.030	0.756	32.038	2.76
Projection-based cascaded U-Net (new)	**0.260**	**0.0221**	**0.816**	**33.685**	**0.3551**	**0.0353**	**0.719**	**31.251**	**0.310**	**0.029**	**0.761**	**32.231**	**2.77**

Note that in these tables, the metrics have been measured using the ground truth (fully sampled) image and the reconstructed image. The runtime has been evaluated for the joint reconstruction of a whole image volume with 32 slices. Additionally, it can be observed that the results obtained in Tables 2 and 3 verify each other. Furthermore, we have performed statistical

Table 3: Simulation results for various models with equispaced (uniform) Cartesian subsampling mask

Acceleration	4-fold				8-fold				Aggregate				Runtime (s)
Network	Loss	NMSE	SSIM	PSNR	Loss	NMSE	SSIM	PSNR	Loss	NMSE	SSIM	PSNR	
Zero-filling	-	0.041	0.709	29.833	-	0.078	0.602	26.898	-	0.0611	0.650	28.208	0
CNN	0.306	0.033	0.760	30.934	0.451	0.068	0.640	27.476	0.385	0.053	0.692	28.918	0.32
CNN+DC	0.303	0.032	0.769	31.028	0.450	0.067	0.641	27.624	0.384	0.051	0.694	29.079	0.33
Cascaded CNN	0.280	0.026	0.78	32.379	0.419	0.055	0.654	28.579	0.356	0.042	0.714	30.139	1.72
Projection-based cascaded CNN	**0.271**	**0.024**	**0.802**	**32.894**	**0.399**	**0.048**	**0.678**	**29.217**	**0.342**	**0.038**	**0.731**	**30.713**	**1.85**
Baseline U-Net	0.274	0.025	0.795	32.763	0.376	0.041	0.698	30.210	0.330	0.034	0.736	31.136	0.49
Baseline U-Net + DC layer	0.269	0.024	0.802	33.037	0.370	0.039	0.703	30.494	0.324	0.032	0.746	31.501	0.53
Cascaded U-Net	0.259	0.022	0.816	33.680	0.35	0.035	0.718	31.267	0.312	0.029	0.757	32.168	2.9
Projection-based cascaded U-Net	**0.258**	**0.0219**	**0.818**	**33.760**	**0.353**	**0.034**	**0.722**	**31.383**	**0.309**	**0.029**	**0.762**	**32.326**	**3.1**

assessment tests for the provided quantitative results. The performed statistical tests include one-way analysis of variance (ANOVA) tests in addition to paired t-tests to confirm the statistical significance of the performance differences throughout all assessment metrics and setups. We set the threshold p-value at α=0.05. The ANOVA test results indicated that the p-values for all simulation settings, including different masks, acceleration factors, and assessment metrics are less than a threshold of 0.01, indicating 99% confidence. Moreover, the introduced models' performances were also evaluated pair-wise against the competing CNN, cascade CNN and U-net results using paired t-tests. The paired t-tests culminated overwhelmingly in p-values less than 0.05, suggesting 95% confidence in the simulation results of our proposed framework.

4.2. Qualitative results

Fig. 5 presents a particular fully sampled image slice and its associated fully sampled k-space data. The fully sampled or ground truth images are used as the target images. On the other hand, Fig. 6 includes the random and equispaced Cartesian undersampling masks with 4-fold, 8-fold, and aggregate acceleration. The corresponding zero-filled image is also shown in this figure, and these ZF images are input to the networks.

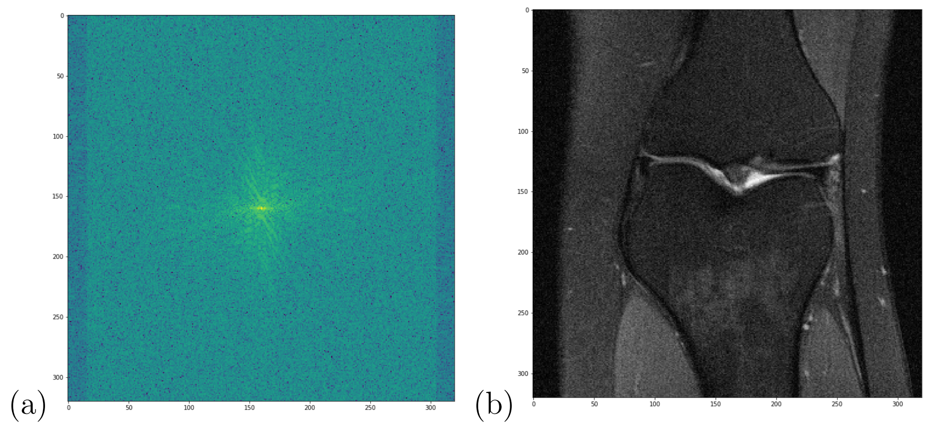

Figure 5: Example for (a) k-space data, (b) Ground truth image.

Figs. 7,8 and 9 display the reconstructed images by the aforementioned deep networks, for the cases of 8-fold, aggregate and 4-fold acceleration factors, respectively. All these images qualitatively verify the quantitative results as summarized in Table 2. We can see that the projection-based cascade U-Net output images are of a higher perceptual quality than the reconstructed images by the other networks and details are recovered better.

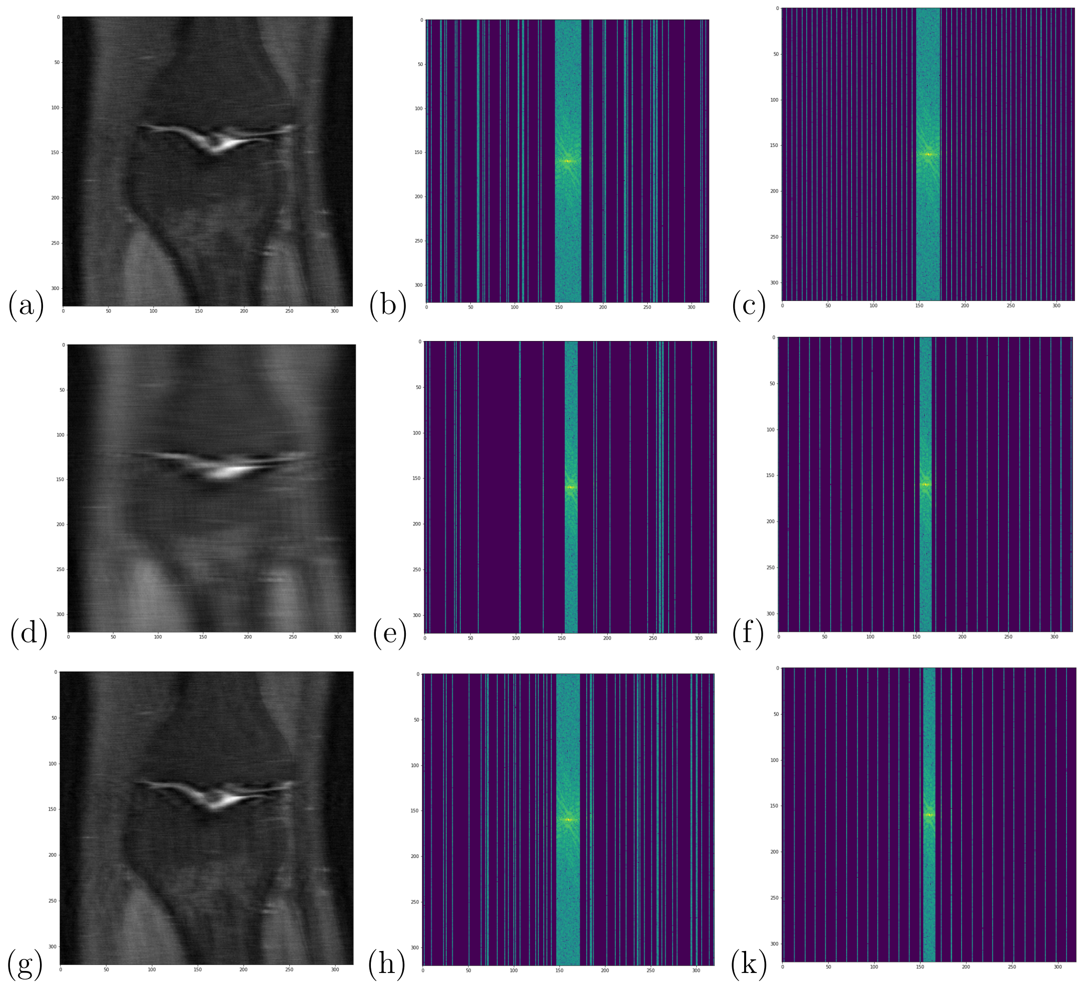

Figure 6: (a) 4-fold zero-filling image, (b) 4-fold undersampled k-space with random mask function, (c) 4-fold undersampled k-space with equispaced mask function, (d) 8-fold zero-filling image, (e) 8-fold undersampled k-space with random mask function, (f) 8-fold undersampled k-space with equispaced mask function, (g) aggregate zero-filling image, (h) aggregate undersampled k-space with random mask function, (k) aggregate undersampled k-space with equispaced mask function.

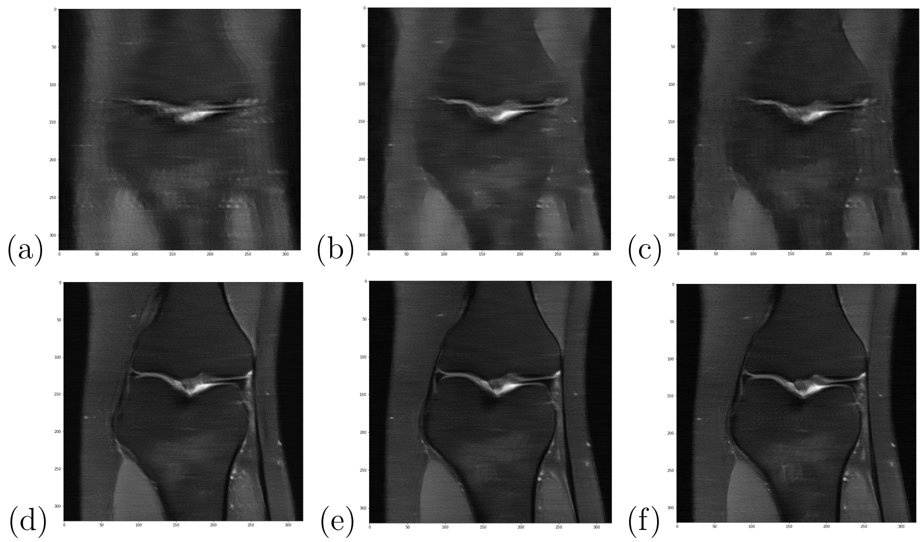

Figure 7: Reconstructed images for 8-fold subsampling with random mask function in the case of (a) Standard CNN, (b) Cascaded CNN, (c) projection-based cascaded CNN, (d) U-Net baseline, (e) Novel cascaded U-Net, (f) Novel projection-based cascaded U-Net.

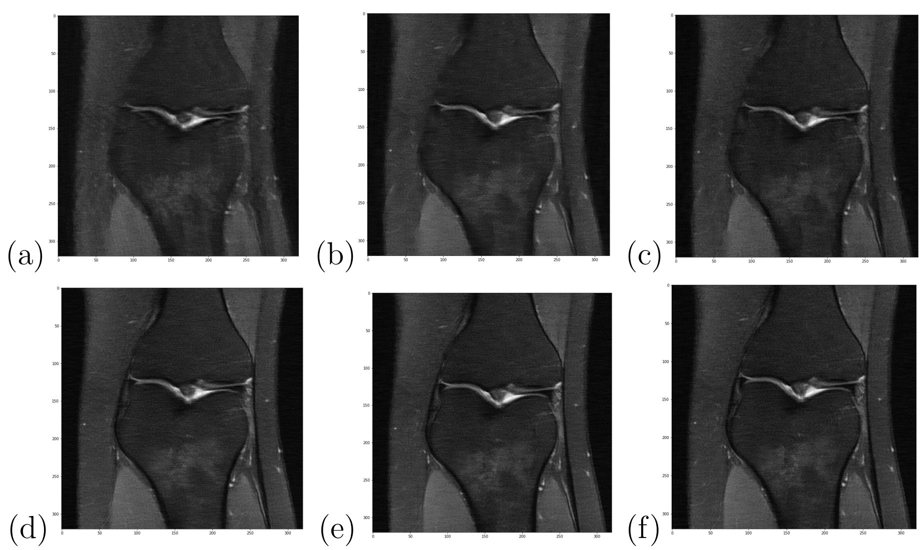

Figure 8: Reconstructed images with random mask function for aggregate subsampling for (a) Standard CNN, (b) Cascaded CNN, (c) projection-based cascaded CNN, (d) U-Net baseline, (e) Novel cascaded U-Net, (f) Novel projection-based cascaded U-Net.

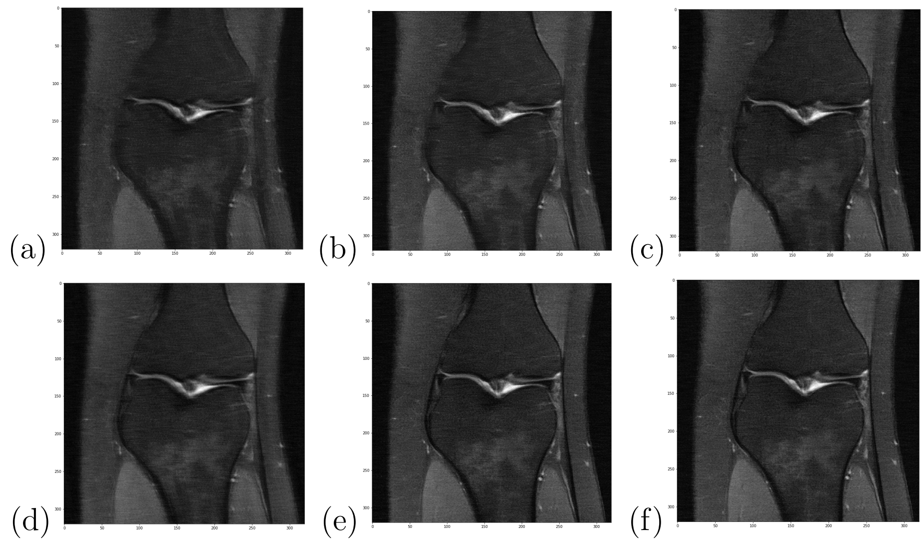

Figure 9: Reconstructed images for 4-fold subsampling with random mask function in the case of (a) Standard CNN, (b) Cascaded CNN, (c) projection-based cascaded CNN, (d) U-Net baseline, (e) Novel cascaded U-Net, (f) Novel projection-based cascaded U-Net.

5. Conclusion

U-Net has been one of the most commonly used deep learning structures in biomedical imaging applications. In this paper, we proposed and evaluated a novel projection-based cascaded U-Net structure for the important problem of MRI reconstruction from highly undersampled k-space observation data. We evaluated the introduced model's applicability and performance by using the rather recent fastMRI dataset as a benchmark. Random and uniform Cartesian undersampling masks of differing ratios have been employed to model the MRI data acquisition in the k-space. The zero-filling image estimate forms the input to the realized networks. The U-Net is employed as the building block of the novel cascaded U-Net structure. The intermediary updated data consistency layers firstly rectify the reconstructed images at their input and forward these rectified images to the successive network stage. The UDC also creates a secondary output, which gets collected at a later stage of the structure for further processing. Applying this novel method in a residual setting, we observed that the projection-based cascade U-Net structure with UDCs outperforms a variety of previously introduced deep networks that are aimed at MRI reconstruction. The introduced novel model can produce images with improved SNR results. The results for the new model can maintain texture details better when compared to the prior networks.

Conflict of interest statement

None of the authors have any conflicts to declare, financial or otherwise.

Acknowledgement

This work was supported by TUBITAK (The Scientific and Technological Research Council of Turkey) under project no. 119E248.

References

[1] K. G. Hollingsworth, Reducing acquisition time in clinical MRI by data undersampling and compressed sensing reconstruction, Physics in Medicine and Biology 60 (21) (2015) R297–R322. doi:10.1088/0031-9155/60/21/r297.
URL https://doi.org/10.1088%2F0031-9155%2F60%2F21%2Fr297

[2] M. Lustig, D. L. Donoho, J. M. Santos, J. M. Pauly, Compressed Sensing MRI, IEEE Signal Processing Magazine 25 (2) (2008) 72–82. doi:10.1109/MSP.2007.914728.

[3] A. K. Tanc, E. M. Eksioglu, MRI reconstruction with joint global regularization and transform learning, Computerized Medical Imaging and Graphics 53 (2016) 1–8. doi:https://doi.org/10.1016/j.compmedimag.2016.06.004.
URL http://www.sciencedirect.com/science/article/pii/S0895611116300520

[4] E. M. Eksioglu, Decoupled algorithm for MRI reconstruction using nonlocal block matching model: BM3D-MRI, Journal of Mathematical Imaging and Vision 56 (3) (2016) 430–440.

[5] E. M. Eksioglu, A. K. Tanc, Denoising AMP for MRI Reconstruction: BM3D-AMP-MRI, SIAM Journal on Imaging Sciences 11 (3) (2018) 2090–2109. arXiv:https://doi.org/10.1137/18M1169655, doi:10.1137/18M1169655.
URL https://doi.org/10.1137/18M1169655

[6] D. Liang, J. Cheng, Z. Ke, L. Ying, Deep Magnetic Resonance Image Reconstruction: Inverse Problems Meet Neural Networks, IEEE Signal Processing Magazine 37 (1) (2020) 141–151. doi:10.1109/MSP.2019.2950557.

[7] J. Schlemper, J. Caballero, J. V. Hajnal, A. N. Price, D. Rueckert, A Deep Cascade of Convolutional Neural Networks for Dynamic MR Image

Reconstruction, IEEE Transactions on Medical Imaging 37 (2) (2017) 491–503. doi:10.1109/TMI.2017.2760978.

[8] J. Schlemper, J. Caballero, J. V. Hajnal, A. Price, D. Rueckert, A deep cascade of convolutional neural networks for MR image reconstruction, in: International Conference on Information Processing in Medical Imaging, Springer, 2017, pp. 647–658.

[9] D. Kocanaogullari and E. M. Eksioglu, Deep Learning for MRI Reconstruction Using a Novel Projection Based Cascaded Network, in: 2019 IEEE 29th International Workshop on Machine Learning for Signal Processing (MLSP), 2019. doi:10.1109/MLSP.2019.8918715.

[10] X. Qiao, J. Du, L. Wang, Z. He, Y. Jia, A Model-Based Deep Network for MRI Reconstruction Using Approximate Message Passing Algorithm, in: ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2020, pp. 1105–1109.

[11] Y. Wu, Y. Ma, J. Liu, J. Du, L. Xing, Self-attention convolutional neural network for improved MR image reconstruction, Information Sciences 490 (2019) 317–328. doi:https://doi.org/10.1016/j.ins.2019.03.080.
URL http://www.sciencedirect.com/science/article/pii/S0020025519302919

[12] Z. Zhou, F. Han, V. Ghodrati, Y. Gao, W. Yin, Y. Yang, P. Hu, Parallel imaging and convolutional neural network combined fast MR image re-

construction: Applications in low-latency accelerated real-time imaging, Medical Physics 46 (8) (2019) 3399–3413.

[13] S. Ramanarayanan, B. Murugesan, K. Ram, M. Sivaprakasam, DC-WCNN: A deep cascade of wavelet based convolutional neural networks for MR image reconstruction, in: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), IEEE, 2020, pp. 1069–1073.

[14] V. Ghodrati, J. Shao, M. Bydder, Z. Zhou, W. Yin, K.-L. Nguyen, Y. Yang, P. Hu, MR Image Reconstruction Using Deep Learning: Evaluation of Network Structure and Loss Functions, Quantitative Imaging in Medicine and Surgery 9 (9) (2019) 1516.

[15] S. Wang, Z. Su, L. Ying, X. Peng, S. Zhu, F. Liang, D. Feng, D. Liang, Accelerating magnetic resonance imaging via deep learning, in: 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI), IEEE, 2016, pp. 514–517.

[16] S. Li, Y. Chen, S. Yang, W. Luo, Cascade dense-unet for prostate segmentation in MR images, in: International Conference on Intelligent Computing, Springer, 2019, pp. 481–490.

[17] D. Lachinov, E. Vasiliev, V. Turlapov, Glioma segmentation with cascaded U-net, in: International MICCAI Brainlesion Workshop, Springer, 2018, pp. 189–198.

[18] J. Schlemper, O. Oktay, M. Schaap, M. Heinrich, B. Kainz, B. Glocker, D. Rueckert, Attention gated networks: Learning to leverage salient regions in medical images, Medical Image Analysis 53 (2019) 197–207.

doi:https://doi.org/10.1016/j.media.2019.01.012.
URL https://www.sciencedirect.com/science/article/pii/S1361841518306133

[19] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.

[20] L. Rundo, C. Han, Y. Nagano, J. Zhang, R. Hataya, C. Militello, A. Tangherloni, M. S. Nobile, C. Ferretti, D. Besozzi, M. C. Gilardi, S. Vitabile, G. Mauri, H. Nakayama, P. Cazzaniga, Use-net: Incorporating squeeze-and-excitation blocks into u-net for prostate zonal segmentation of multi-institutional mri datasets, Neurocomputing 365 (2019) 31–43. doi:https://doi.org/10.1016/j.neucom.2019.07.006.
URL https://www.sciencedirect.com/science/article/pii/S0925231219309245

[21] Y. Han, L. Sunwoo, J. C. Ye, k -Space Deep Learning for Accelerated MRI, IEEE Transactions on Medical Imaging 39 (2) (2019) 377–386. doi:10.1109/TMI.2019.2927101.

[22] C. M. Hyun, H. P. Kim, S. M. Lee, S. Lee, J. K. Seo, Deep learning for undersampled MRI reconstruction, Physics in Medicine & Biology 63 (13) (2018) 135007.

[23] C. M. Sandino, N. Dixit, J. Y. Cheng, S. S. Vasanawala, Deep convolutional neural networks for accelerated dynamic magnetic resonance imaging, preprint (2017).

[24] K. Hammernik, T. Klatzer, E. Kobler, M. P. Recht, D. K. Sodickson, T. Pock, F. Knoll, Learning a variational network for reconstruction of accelerated MRI data, Magnetic Resonance in Medicine 79 (6) (2018) 3055–3071. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/mrm.26977, doi:https://doi.org/10.1002/mrm.26977.
URL https://onlinelibrary.wiley.com/doi/abs/10.1002/mrm.26977

[25] T. Eo, Y. Jun, T. Kim, J. Jang, H.-J. Lee, D. Hwang, KIKI-net: cross-domain convolutional neural networks for reconstructing undersampled magnetic resonance images, Magnetic Resonance in Medicine 80 (5) (2018) 2188–2201.

[26] R. Souza, R. Frayne, A hybrid frequency-domain/image-domain deep network for magnetic resonance image reconstruction, in: 2019 32nd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), IEEE, 2019, pp. 257–264.

[27] J. Cheng, H. Wang, L. Ying, D. Liang, Model learning: Primal dual networks for fast MR imaging, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2019, pp. 21–29.

[28] L. Bao, F. Ye, C. Cai, J. Wu, K. Zeng, P. C. van Zijl, Z. Chen, Undersampled MR image reconstruction using an enhanced recursive residual network, Journal of Magnetic Resonance 305 (2019) 232–246.

[29] Y. Zheng, B. Zhen, A. Chen, F. Qi, X. Hao, B. Qiu, A hybrid convolutional neural network for super-resolution reconstruction of MR images, Medical Physics 47 (7) (2020) 3013–3022. doi:https://doi.org/10.1002/mp.14152.
URL https://aapm.onlinelibrary.wiley.com/doi/abs/10.1002/mp.14152

[30] D. J. Lin, P. M. Johnson, F. Knoll, Y. W. Lui, Artificial Intelligence for MR Image Reconstruction: An Overview for Clinicians, Journal of Magnetic Resonance Imaging (2020).

[31] J. Zbontar, F. Knoll, A. Sriram, M. J. Muckley, M. Bruno, A. Defazio, M. Parente, K. J. Geras, J. Katsnelson, H. Chandarana, et al., fastMRI: An open dataset and benchmarks for accelerated MRI, arXiv preprint arXiv:1811.08839 (2018).

[32] Z. Ramzi, P. Ciuciu, J.-L. Starck, Benchmarking MRI Reconstruction Neural Networks on Large Public Datasets, Applied Sciences 10 (5) (2020) 1816.

[33] Z. Ramzi, P. Ciuciu, J.-L. Starck, Benchmarking deep nets MRI reconstruction models on the fastMRI publicly available dataset, in: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), IEEE, 2020, pp. 1441–1445.

[34] A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with deep convolutional neural networks, in: Advances in neural information processing systems, 2012, pp. 1097–1105.

[35] O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional Networks for Biomedical Image Segmentation, in: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, Springer, Springer International Publishing, 2015, pp. 234–241.

[36] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, L. D. Jackel, Backpropagation applied to handwritten zip code recognition, Neural Computation 1 (4) (1989) 541–551.

[37] D. Zhao, F. Zhao, Y. Gan, Reference-Driven Compressed Sensing MR Image Reconstruction Using Deep Convolutional Neural Networks without Pre-Training, Sensors 20 (1) (2020) 308.