Received: 9 August 2020 | Revised: 18 March 2021 | Accepted: 14 April 2021

DOI: 10.1002/mrm.28827

FULL PAPER

Magnetic Resonance in Medicine

Systematic evaluation of iterative deep neural networks for fast parallel MRI reconstruction with sensitivity-weighted coil combination

Kerstin Hammernik[1,2] | **Jo Schlemper**[3] | **Chen Qin**[1,4] | **Jinming Duan**[1,5] | **Ronald M. Summers**[6] | **Daniel Rueckert**[1,2]

[1]Department of Computing, Imperial College London, London, United Kingdom

[2]Chair for AI in Healthcare and Medicine, Klinikum Rechts der Isar, Technical University of Munich, Munich, Germany

[3]Hyperfine Research Inc., Guilford, CT, USA

[4]Institute for Digital Communications, School of Engineering, University of Edinburgh, Edinburgh, United Kingdom

[5]School of Computer Science, University of Birmingham, Birmingham, United Kingdom

[6]NIH Clinical Center, Bethesda, MD, USA

Correspondence
Kerstin Hammernik, Department of Computing, Imperial College London, South Kensington Campus, London SW7 2AZ, UK.
Email: k.hammernik@imperial.ac.uk

Funding information
EPSRC Programme Grant, Grant Number: EP/P001009/1; Intramural Research Programs of the National Institutes of Health Clinical Center, Project Number: 1Z01 CL040004

Purpose: To systematically investigate the influence of various data consistency layers and regularization networks with respect to variations in the training and test data domain, for sensitivity-encoded accelerated parallel MR image reconstruction.

Theory and Methods: Magnetic resonance (MR) image reconstruction is formulated as a learned unrolled optimization scheme with a down-up network as regularization and varying data consistency layers. The proposed networks are compared to other state-of-the-art approaches on the publicly available fastMRI knee and neuro dataset and tested for stability across different training configurations regarding anatomy and number of training samples.

Results: Data consistency layers and expressive regularization networks, such as the proposed down-up networks, form the cornerstone for robust MR image reconstruction. Physics-based reconstruction networks outperform post-processing methods substantially for $R = 4$ in all cases and for $R = 8$ when the training and test data are aligned. At $R = 8$, aligning training and test data is more important than architectural choices.

Conclusion: In this work, we study how dataset sizes affect single-anatomy and cross-anatomy training of neural networks for MRI reconstruction. The study provides insights into the robustness, properties, and acceleration limits of state-of-the-art networks, and our proposed down-up networks. These key insights provide essential aspects to successfully translate learning-based MRI reconstruction to

Preliminary data for this manuscript were submitted to the fastMRI challenge (team *holykspace*) and were presented as oral presentation at ISMRM 2020.

clinical practice, where we are confronted with limited datasets and various imaged anatomies.

KEYWORDS

data consistency, deep learning, domain shift, down-up networks, fastMRI, iterative image reconstruction, parallel imaging

1 | INTRODUCTION

Parallel imaging (PI)[1-3] forms the foundation of accelerated data acquisition in magnetic resonance imaging (MRI), which is tremendously time-consuming. In the last decade, PI combined with compressed sensing (CS) techniques have resulted in substantial improvements in acquisition speed and image quality.[4-9] Although PI-CS can achieve state-of-the-art performance, designing effective regularization schemes and tuning of hyper-parameters are not trivial. Starting in 2016, deep learning algorithms have become extremely popular and effective tools in data-driven learning of inverse problems and have enabled progress beyond the limitations of CS.

Deep learning for image reconstruction is an enormously fast-growing field, which makes it challenging to keep an overview over the different approaches. For details on the developments of deep learning for MRI reconstruction, we refer the interested reader to survey papers.[10-13] In this work, we only focus on reviewing relevant approaches for 2D MRI reconstruction. Table 1 gives a compact overview of already

TABLE 1 Overview of related work for 2D MRI reconstruction

Related work	Acquisition	Coil comb	Application	Data consistency	Network	Adversarial
Aggarwal et al[14]	MC	SENSE	Neuro	Proximal mapping	CNN	No
Akçakaya et al[15]	MC	RSS	Neuro, cardiac	ACL	CNN	No
Duan et al[16]	MC	SENSE	MSK	Variable splitting	CNN	No
Eo et al[17]	SC	–	Neuro	Cross-domain	CNN	No
Hammernik et al[18]	MC	SENSE	MSK	Gradient descent	VN	No
Hyun et al[19]	SC	–	Neuro	k-space correction	UNET	No
Johnson et al[20]	MC	SENSE	MSK	Gradient descent	UNET	No
Knoll et al[21]	SC/MC	RSS	MSK, neuro	No	UNET	No
Lee et al[22]	SC/MC	RSS	Neuro	No	UNET	No
Mardani et al[23]	SC	–	MSK, abdominal	k-space projection	CNN	Yes
Pezotti et al[24]	MC	RSS	MSK	Gradient descent	CS-NET	No
Putzky et al[25]	SC	–	MSK	Gradient descent	RIM	No
Qin et al[26]	SC	–	Cardiac	Proximal mapping	RNN	No
Quan et al[27]	SC	–	MSK, neuro	Loss penalty	WNET	Yes
Ran et al[28]	SC	–	MSK, neuro	Cross-domain	CNN	No
Schlemper et al[29]	SC	–	Cardiac	Proximal mapping	CNN	No
Seitzer et al[30]	SC	–	Cardiac	Proximal mapping	CNN	Yes
Sriram et al[31]	MC	RSS	MSK	k-space	UNET	No
Wang et al[32]	MC	RSS	MSK	Proximal mapping	RNN	No
Yang et al[33]	SC	–	Neuro	ADMM	CTA	No
Yang et al[34]	SC	–	Neuro	Loss penalty	UNET	Yes
Zhang et al[35]	MC	RSS	Neuro	Loss penalty	CNN	Yes
Zhu et al[36]	SC/MC	SENSE	Neuro	No	FC + CNN	No

Acquisition: MC, multi-coil; SC, single-coil. Coil comb: RSS, root-sum-of-squares; SENSE, sensitivity-weighted coil combination[2]. Data consistency: ACL, auto-calibration lines, no training database required; ADMM, Alternating Direction Method of Multipliers. Network: CNN, convolutional neural network; CTA, combination of convolutions and trainable activations; FC, fully connected layer; RNN, recurrent neural Network; VN, variational network.

published approaches.[14-36] The different approaches can be distinguished based on (1) the acquisition type, that is, single-coil or multi-coil reconstruction, (2) the type of coil combination used in multi-coil approaches, (3) the type of application, (4) realization of consistency to measured *k*-space data, (5) network architecture, and (6) the use of adversarial training strategies in addition to commonly used similarity measures, for example, mean squared error (MSE). Several things can be noted: A majority of approaches work on single-coil reconstruction; however, the standard approach for MR acquisition is PI. Within the multi-coil approaches, different types of coil combinations, that is, root-sum-of-squares (RSS) and sensitivity-weighted combination occur, similar to a preference towards sensitivity encoding (SENSE)[2] or generalized autocalibrating partially parallel acquisitions (GRAPPA).[3] Data consistency (DC) is achieved in several ways. Examples here are learned unrolled optimization schemes, image-domain networks with or without *k*-space correction, or image-domain networks, where a loss penalty on the *k*-space data is added only during training. While the majority of image-domain networks are based on UNET,[37] iterative reconstruction approaches utilize different architectures such as convolutional neural networks (CNNs), UNET,[37] or fields of experts (FoE) regularization[38] in variational networks (VNs).[18] All approaches in Table 1 except[15] rely on large training databases. Last, it can be noted that most architectures that use adversarial training use image-domain networks as generating networks and are deployed for a single-coil setting. Despite the tremendous volume of research, there remain several open questions in the field of learning MRI reconstruction.

It is still an open question how well neural networks generalize to anatomical changes, and variations in training and test domain at a large scale. Learning-based approaches are often evaluated in dedicated research settings with small, homogeneous datasets and it is hard to determine if the proposed building blocks, consisting mainly of regularization networks and DC layers, are only effective in the proposed setting or generalize to variations in the data and experimental setups. However, proposed generalization experiments in previous works have only focused on small datasets and dedicated research settings. Knoll et al[39] studied the influence of differences in signal-to-noise ratio (SNR) on knee data. Johnson et al[20] showed the influence of training specific and joint networks for different anatomies of the musculoskeletal (MSK) system, including shoulder, hip, ankle, and knee images. A first instability analysis of neural networks for image reconstruction was studied in Antun et al.[40] However, a different instability analysis was conducted for the selected approaches that were proposed for a single-coil or multi-coil setting, and with or without DC layer. Furthermore, the approaches were tested on datasets that differed in levels of SNR. This makes it challenging to draw general conclusions.

The aim of this work is to bridge the gap of the aforementioned challenges that we have observed in deep learning for parallel MRI reconstruction. We study the influence of regularization networks, DC layers, and variations in the data, in a controlled experimental setup. To the best of our knowledge, this is the first work that studies the effect of different training data configurations, including variations in anatomies and sample size, for neural network reconstructions at a large scale, using the publicly available fastMRI datasets with approximately 5400 training cases.[21] We perform an extensive evaluation of different networks with varying DC layers and regularization networks. We propose a down-up network (DUNET) as the regularization network, and we show the superior performance of the proposed DUNET compared to other state-of-the-art approaches for varying training data scenarios. These scenarios include variations in anatomy, using knee and neuro data, and variations in the number of training samples. Hence, the systematic evaluation allows us to experimentally investigate the robustness and limits of DC layers and regularization networks with respect to different acceleration factors. All experiments are performed on the fastMRI multi-coil knee and neuro dataset,[21] where a fully sampled sensitivity-combined reconstruction, with an extended set of coil sensitivity maps,[41] is used as ground truth. To reproduce our findings, we provide all our source code along with the data processing scripts for the fastMRI datasets online.

2 | THEORY

Accelerated MRI reconstruction aims at recovering a reconstruction $x \in \mathbb{C}^{N_x}$ from a set of undersampled *k*-space measurements $y \in \mathbb{C}^{N_y}$ which are corrupted by additive Gaussian noise $n \in \mathbb{C}^{N_y}$ following

$$y = Ax + n. \tag{1}$$

This inverse problem involves a linear forward operator $A: \mathbb{C}^{N_x} \rightarrow \mathbb{C}^{N_y}$ modeling the MR physics. Here, N_x and N_y define the dimensions of the reconstruction x and the *k*-space data y according to the underlying multi-coil or single-coil problem. We investigate a linear multi-coil operator $A: \mathbb{C}^{N_x} \rightarrow \mathbb{C}^{N_y 2}$ using an extended set of M coil sensitivity maps to overcome field of view (FoV) issues,[41] reconstructing $x = [x_1, \ldots, x_M]$. The dimensions are given as $N_x = N_{\mathrm{FE}} \cdot N_{\mathrm{PE}} \cdot M$ and $N_y = N_{\mathrm{FE}} \cdot N_{\mathrm{PE}} \cdot Q$, where N_{FE} and N_{PE} denote the number of frequency encoding (FE) and phase encoding (PE) lines, respectively, and Q is the number of receive coils. The final result is obtained by a RSS combination of the reconstruction along the sensitivity map dimension.

2.1 | Learning unrolled optimization

An approximate solution $\hat{x} \in \mathbb{C}^{N_x}$ to the inverse problem in Equation (1) is typically obtained by minimizing a regularized problem of form

$$\hat{x} \in \underset{x \in \mathbb{C}^{N_x}}{\arg\min} \mathscr{R}[x] + \lambda \mathscr{D}[Ax, y] \tag{2}$$

where λ balances between the regularization term $\mathscr{R}[x]$ and the DC term $\mathscr{D}[Ax, y] = \frac{1}{2}\|Ax - y\|_2^2$. While $\mathscr{R}[x]$ is fixed in classical CS approaches, we learn $\mathscr{R}[x]$ from data. A solution is obtained by alternating optimization in $\mathscr{R}$ and $\mathscr{D}$ for a fixed number of iterations T.[11,14,16,29] We define the fixed unrolled algorithm for MRI reconstruction as

$$x^{t+\frac{1}{2}} = x^t - f_{\theta^t}(x^t), \tag{3}$$

$$x^{t+1} = g(x^{t+\frac{1}{2}}, y, A) \tag{4}$$

for $0 \leq t < T$ (see Figure 1). First, we take a step along the direction of the negative gradient $-\nabla_x \mathscr{R}$, which is replaced by a regularization network $-f_\theta$ with trainable parameters θ. Hence, the regularization network naturally learns the residual. The regularization network f_θ: $\mathbb{C}^{N_x} \to \mathbb{C}^{N_x}$ has complex-valued input and output channels, represented as two-channel real-valued image, and the same network is applied separately to x_m, $m = 1, \ldots, M$. The DC layer is denoted by g. In the following, we describe regularization networks and DC layers that we use in our work in more detail.

2.1.1 | Regularization networks

The regularization network f_θ can be realized by any type of CNNs, or it can be motivated by variational methods.[18] Commonly used regularization networks are a 5-layer CNN,[14,29] UNET[21,37] or the fields-of-experts model.[18,38] In this work, we introduce DUNETs that serve as an efficient alternative to the expressive UNETs.[42]

The DUNET as shown in Figure 1 first downsamples the image by convolutions with stride 2 and then performs analysis on this coarser scale. Shifting the computation to a coarser scale is not only more memory efficient, but also does not lower the reconstruction quality at the original scale.[42,43] The core of DUNET are the multiple down-up blocks (DUBs) applied in an iterative way. This structure allows for an efficient propagation of information at different scales.[44] The outputs of the DUBs are concatenated and further analyzed by a residual convolution/activation block, followed by sub-pixel convolutions, which perform superior in terms of expressiveness and computational efficiency over upsampling convolution.[45]

2.1.2 | Data consistency

The DC term allows us to consider the physics of MR acquisition in the image reconstruction problem, and measures the similarity to the acquired k-space data. The DC term can be incorporated in the learning-based reconstruction procedure in several ways.

One possibility is to perform a gradient step[18] related to the DC term $\mathscr{D}[Ax, y]$

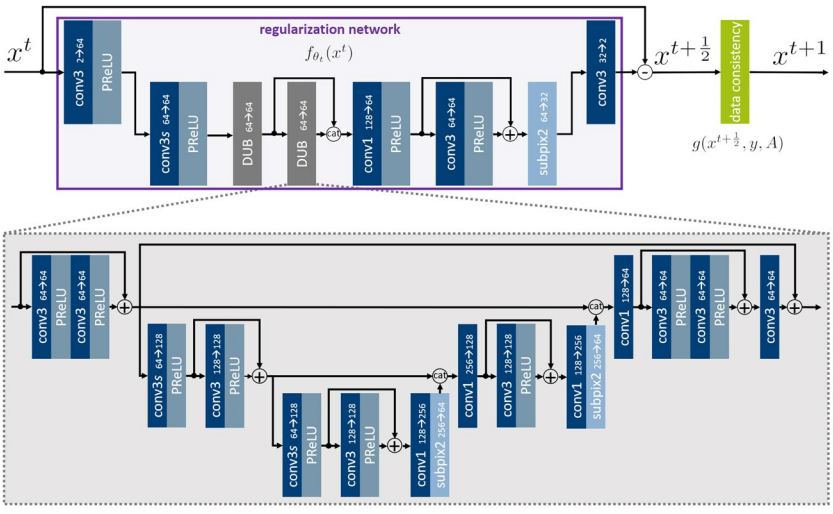

FIGURE 1 Structure of the architecture, including a regularization network and a data consistency (DC) layer. We present a novel down-up network (DUNET) as regularization network. The core of the DUNET are the down-up blocks (DUB), which are applied sequentially at a coarser scale rather than the original input. The outputs of the DUBs are concatenated and processed further with convolutions, activation functions and subpixel upsampling to get back to the original scale. Residual connections are added whenever possible. The convolution blocks (conv) additionally depict the kernel size, the number of input features and output features, and the stride if applicable

$$g_{\mathrm{GD}}(x^{t+\frac{1}{2}}) = x^{t+\frac{1}{2}} - \lambda^t A^*(Ax^{t+\frac{1}{2}} - y), \tag{5}$$

where A^* denotes the adjoint operator of A.

Instead of gradient descent (GD), DC can be modeled by the proximal mapping (PM)[14,29]

$$g_{\mathrm{PM}}(x^{t+\frac{1}{2}}) = \arg\min_x \frac{1}{2}\|x - x^{t+\frac{1}{2}}\|_2^2 + \frac{\lambda}{2}\|Ax - y\|_2^2. \tag{6}$$

This is especially feasible if the PM is easy to compute and a closed-form solution exists. If no closed-form solution exists, or a solution is intractable to compute, as this is typically the case for parallel MRI involving coil sensitivity maps, the PM can be solved numerically using a conjugate gradient optimizer as presented in Ref. [14].

To avoid the extensive computations of the PM, Duan et al[16] proposed a variable splitting (VS) scheme. To review VS, we first introduce the sensitivity-weighted multi-coil operator for the qth coil as $A_q = M\mathcal{F}C_q$. The operator $C_q: \mathbb{C}^{N_x} \to \mathbb{C}^{N_x}$ applies the qth pre-computed coil sensitivity map to x, for $q = 1, \ldots, Q$. This is followed by a Fourier Transform (FT) $\mathcal{F}: \mathbb{C}^{N_x} \to \mathbb{C}^{N_x}$. The operator $M: \mathbb{C}^{N_x} \to \mathbb{C}^{N_y}$ realizes the Cartesian sampling pattern and masks out k-space lines that where not acquired. VS divides the problem defined in Equation (6) in two sub-problems by using a coil-wise splitting variable $z_q \in \mathbb{C}^{N_x}$

$$z_q^{t+1} = \arg\min_{z_q} \frac{\lambda}{2}\sum_{q=1}^{Q}\|M\mathcal{F}z_q - y_q\|_2^2 + \frac{\alpha}{2}\sum_{q=1}^{Q}\|z_q - C_q x^{t+\frac{1}{2}}\|_2^2 \tag{7}$$

$$g_{vs}(x^{t+\frac{1}{2}}) = \arg\min_x \frac{\alpha}{2}\sum_{q=1}^{Q}\|z_q^{t+1} - C_q x\|_2^2 + \frac{\beta}{2}\|x - x^{t+\frac{1}{2}}\|_2^2, \tag{8}$$

where $\alpha > 0$ and $\beta > 0$ balance the influence of the soft constraints. Solving these sub-problems yields the following closed-form solution

$$z_q^{t+1} = \mathcal{F}^{-1}\left(\left(\lambda M^*M + \alpha I\right)^{-1}\left(\alpha\mathcal{F}C_q x^{t+\frac{1}{2}} + \lambda M^* y_q\right)\right) \tag{9}$$

$$g_{vs}(x^{t+\frac{1}{2}}) = \left(\beta I + \alpha\sum_{q=1}^{Q}C_q^*C_q\right)^{-1}\left(\beta x^{t+\frac{1}{2}} + \alpha\sum_{q=1}^{Q}C_q^* z_q^{t+1}\right). \tag{10}$$

Here, I denotes the identity matrix and * the adjoint operation.

All presented DC layers, that is, GD, PM, and VS, ensure soft DC to the measurement data y, representing image reconstruction networks. By setting $\lambda = 0$ in Equation (5), DC is omitted and we achieve a pure residual network performing a post-processing task.

3 | METHODS

This section provides an overview of the used datasets and data processing as well as network setup and training. Specific details on the networks and data processing are given in the source code repository.

3.1 | fastMRI datasets

All our experiments were performed on the fastMRI knee and neuro dataset.[21] Training was performed on the multi-coil training data, testing was performed on the multi-coil validation data. The number of training and testing samples are denoted by N_{train} and N_{test}, respectively. The knee dataset consists of two different sequences:

- Coronal proton-density weighted with fat-saturation (PDFS): $N_{\mathrm{train}} = 489$, $N_{\mathrm{test}} = 99$
- Coronal proton-density weighted without fat-saturation (PD): $N_{\mathrm{train}} = 484$, $N_{\mathrm{test}} = 100$

The neuro dataset consists of four different sequences:

- Axial FLAIR (AXFLAIR): $N_{\mathrm{train}} = 344$, $N_{\mathrm{test}} = 107$
- Axial T_1 (AXT1): $N_{\mathrm{train}} = 498$, $N_{\mathrm{test}} = 169$
- Axial T_1 with contrast agent (AXT1POST): $N_{\mathrm{train}} = 949$, $N_{\mathrm{test}} = 287$
- Axial T_2 (AXT2): $N_{\mathrm{train}} = 2678$, $N_{\mathrm{test}} = 815$

For details on the sequence parameters, we refer to the original publication.[21]

3.2 | Data processing

We defined the target as the sensitivity-weighted coil-combined image of the fully sampled data. We estimated two sets ($M = 2$) of sensitivity maps according to soft SENSE [41] to account for any field-of-view issues or other obstacles in the data. The number of auto-calibration lines (ACLs) needed for sensitivity map estimation varied according to the acceleration factor and was set to 30 ACLs for $R = 4$ and 15 ACLs for $R = 8$ for the training and validation set. These numbers were motivated by examining the number of given low frequencies in the test and challenge dataset. The data were normalized by a factor obtained from the low frequency scans by taking the median value of the 20% largest magnitude values, to account for outliers.

We also make use of foreground masks to stabilize training. Foreground masks were extracted semi-automatically for the knee dataset,[46] and by thresholding the RSS combination of the sensitivity maps for the neuro dataset.

3.3 | Training setup

All networks were trained using a combined ℓ_1 and structural similarity index (SSIM)[47,48] content loss $\mathscr{L}$ between the reference x_{ref} and the reconstruction $x_{\text{rec}} = x^T$

$$\begin{aligned}\mathscr{L}(x_{\text{rec}}, x_{\text{ref}}) &= 1 - \text{SSIM}(m \odot |x_{\text{rec}}|, m \odot |x_{\text{ref}}|) \\ &+ \gamma_{\ell_1} \ell_1 (m \odot |x_{\text{rec}}|, m \odot |x_{\text{ref}}|),\end{aligned} \tag{11}$$

where $\odot$ is the pixel wise product and $|\cdot|$ denotes the RSS reconstruction to combine the individual output channels $[x_1, x_2]$. This loss formulation also involves a binary foreground mask m to focus the network training on the image content and not on the background. The parameter $\gamma_{\ell_1} = 10^{-5}$ is chosen empirically to match the scale of the two losses and is motivated by the fastMRI challenge requirements. Although we aim for maximizing the SSIM scores in testing, a combined loss is beneficial to stabilize training.[24,47]

We used the ADAM optimizer[49] with learning rate 0.0001, default momentum (0.9,0.999) and learning rate scheduling every 15 epochs by $\gamma_{\text{lr}} = 0.5$. We use a progressive training scheme, starting with 2 cascades in the first 2 epochs and increasing the number of cascades with every epoch, up to a total number of $T = 10$. We trained all network architectures for 60 epochs. To overcome the huge graphics processing unit (GPU) memory consumption during training, we randomly extracted patches of size 96 in FE direction.[29] Training was performed using an NVIDIA Quadro RTX 6000 (24 GB) and took approximately 12 days for a single network. Testing was performed using an NVIDIA Titan Xp (12 GB). We report average reconstruction times of the network architectures along with the number of trainable network parameters in Supporting Information Table S1.

We trained the networks on all contrasts and the acceleration factors $R = 4$ and $R = 8$ simultaneously as we want to examine how the different architectures respond to a generalized training setup. This allows us to have a general network that can be applied to any anatomy and acceleration factor. Hence, we did not aim for the best scores on a benchmark, although we would expect improvements with fine-tuning the networks for specific conditions.

3.4 | Experimental setup

We systematically investigate how state-of-the art architectures and the proposed DUNETs with varying DC layers perform on variations in training data. We study the domain shift problem for image reconstruction experimentally, that is, we study how the different architectures can deal if training and test data do not come from the same data cohort.

3.4.1 | Network architectures

The DUNETs have $N_f = 64$ base features, resulting in a total number of 3 372 985 network parameters. We implement three different DC layers, that is, GD, PM, and VS. We compared the DUNETs to three state-of-the-art architectures. First, we omitted DC and implemented a residual UNET based on[21] with $N_f = 64$ base features and kernel size 3, corresponding to 3 357 827 parameters. Second, we investigated MoDL,[14] which has a 5-layer CNN with $N_f = 64$ base features and 3×3 filter kernels as regularization network, and PM for DC. The total number of network parameters was 113 155. We omitted batch normalization as this resulted in instable trainings. Third, we investigated VNs, which can be interpreted as unrolled GD scheme with a regularization network that derives from the fields-of-experts model. Following,[18] we learned $N_f = 48$ filter kernels of size 11×11 and trainable linear activation functions with 31 nodes. The filter kernels are projected on the zero-mean and ℓ_2 norm-ball constraint after each parameter update. The total number of parameters is 131 051. The parameters are not shared over the cascades, following the original publication.

The regularization parameter λ was not trained as we experienced instabilities during training the PM-DUNET and MoDL.[14] For these two architectures, we initialized $\lambda = 10$, otherwise $\lambda = 1$. For the VS networks, we experimented with different settings for the parameters α and β, and we set $\alpha = \beta = 0.1$ empirically for our experiments.

3.4.2 | Training data

Training and evaluation was performed on the fastMRI multicoil knee and neuro training and validation set, respectively. We performed following base experiments:

- *knee 100*: Training on 100% knee data ($N_{\text{train}} = 973$)
- *neuro*: Training on 100% neuro data ($N_{\text{train}} = 4412$)

To study the influence on the number of samples and joint training of knee and neuro data, we performed an ablation study as follows:

- *knee 50*: Training with 50% knee data ($N_{\text{train}} = 487$)
- *knee 25*: Training with 25% knee data ($N_{\text{train}} = 244$)
- *joint 100*: Joint knee and neuro training with 18% of all data, samples equal 100% of knee data ($N_{\text{train}} = 968$)
- *joint 50*: Joint knee and neuro training with 9% of all data, samples equal 50% of knee data ($N_{\text{train}} = 486$)
- *joint 25*: Joint knee and neuro training with 4.5% of all data, samples equal 25% of knee data ($N_{\text{train}} = 240$)

- *joint uni 100*: Joint knee and neuro training with uniform distribution of contrasts, samples equal 100% of knee data ($N_{\text{train}} = 974$)
- *joint uni 50*: Joint knee and neuro training with uniform distribution of contrasts, samples equal 50% of knee data ($N_{\text{train}} = 484$)
- *joint uni 25*: Joint knee and neuro training with uniform distribution of contrasts, samples equal 25% of knee data ($N_{\text{train}} = 243$)

We would like to note here that we performed two different sets of joint knee and neuro training, one with uniform and one with non-uniform distribution of the contrasts. As pointed out in Section 3.1, the number of samples differ for the knee and neuro set, and also for the different contrasts in the neuro dataset. While the uniform datasets contain the same number of samples from each available contrast, the non-uniform datasets contain only a fraction of samples such that the distribution of contrasts in the reduced dataset corresponds to the distribution of contrasts in the full dataset, which is a common scenario in clinical practice.

For quantitative evaluation, we report the SSIM. All experiments are visualized as ranked lists[50] with descending SSIM, where the bars are colored regarding the training dataset and the quantitative values are printed next to the bars. Qualitative examples are displayed for the best and worst performing training dataset for the individual models. Difference maps are provided as supporting information.

4 | RESULTS

We plotted ranked lists for networks evaluated on knee and neuro data. The performance of networks trained only on knee and neuro data is depicted in Figure 2 (knee, $R = 4$), Figure 3 (knee, $R = 8$), Figure 4 (neuro, $R = 4$), and Figure 5 (neuro, $R = 8$). Results for trainings on different fractions of knee and joint training data are illustrated in Supporting Information Figure S1 (knee, $R = 4$), Supporting Information Figure S2 (knee, $R = 8$), Supporting Information Figure S3 (neuro, $R = 4$), and Supporting Information Figure S4 (neuro, $R = 8$). For an acceleration factor of $R = 4$, we observe that all post-processing UNETs perform inferior than the worst performing reconstruction network, independent of the number and type of training samples. This effect is more prominent on the neuro data compared to the knee data. For the neuro data and $R = 4$, the best post-processing UNET, trained on neuro data, achieves an SSIM an 0.9291 and the worst reconstruction method, VS-DUNET trained on knee 25 data, achieves an SSIM of 0.9460. For the knee data and $R = 4$, the best post-processing UNET, trained on knee 100 data, achieves an SSIM an 0.9142 and the worst reconstruction method, MoDL trained on neuro data, achieves an SSIM of 0.9153. For an acceleration factor of $R = 8$, the post-processing UNET trained and evaluated on the same data outperforms some image reconstruction networks which were trained and evaluated on different data.

The ranked lists also show a substantial performance gain at accelerations (4/8) in terms of SSIM of the best performing reconstruction DUNET compared to state-of-the-art reconstruction (0.0109/0.0284) and compared to post-processing methods (0.0376/0.0644) for neuro data and a performance gain to state-of-the-art reconstruction (0.0079/0.0266) and post-processing methods (0.0233/0.0415) for knee data. We observe larger performance gain for neuro data compared to knee data, especially for $R = 8$.

Figure 6 shows results for an example coronal PDw scan with fat saturation and $R = 4$. The top row shows the best performing network, corresponding to knee 100 data. We already observe an anatomy change of the UNET in the interchondylar notch which is correctly depicted in all reconstruction networks. The DUNETs have the least artifacts and appear most homogenous compared to VN and MoDL, supported by the difference image in Supporting Information Figure S5. The bottom row shows the worst performing networks. For MoDL, VN, and UNET, networks trained only with neuro data perform worst. This is different for the DUNETs where the joint 25 dataset led to the worst results.

Results for a selected coronal PDw scan and $R = 8$ are illustrated in Figure 7, along with difference images in

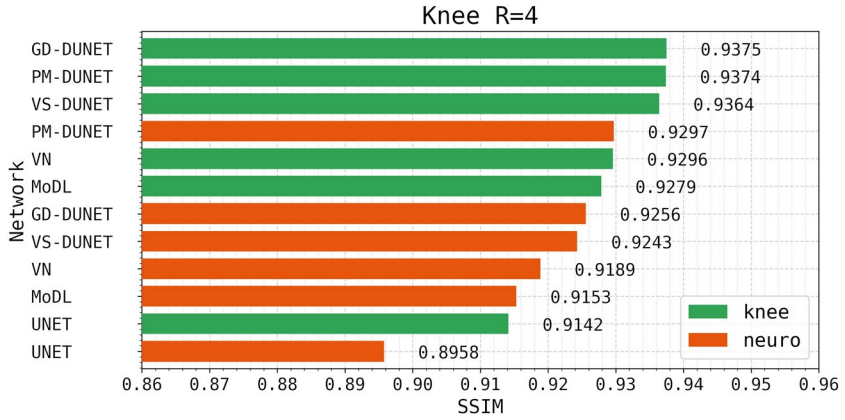

FIGURE 2 Ranked list for the fastMRI knee dataset at $R = 4$ trained with knee and neuro datasets. All reconstruction networks perform superior than the post-processing networks. GD-DUNET trained on the knee dataset performs best. PM-DUNET trained only on the neuro dataset performs better than the state-of-the-art methods

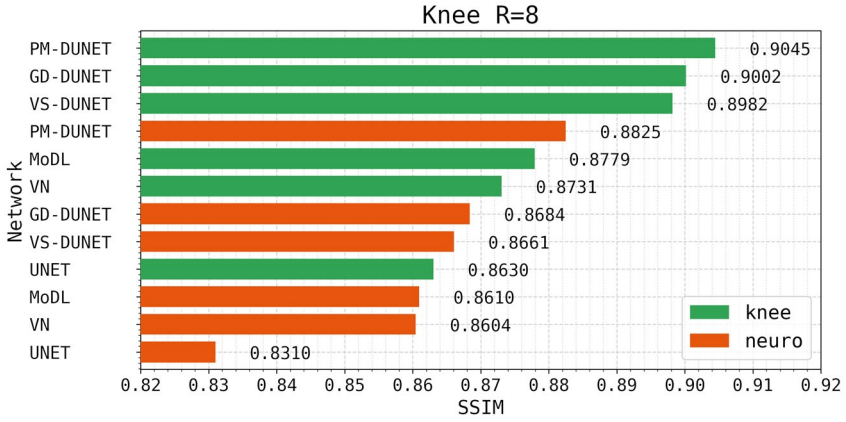

FIGURE 3 Ranked list for the fastMRI knee dataset at $R = 8$ trained with knee and neuro datasets. PM-DUNET trained on the knee dataset performs best. PM-DUNET trained only on the neuro dataset performs better than all state-of-the-art methods. Post-processing UNET trained on knee data performs superior than VN and MoDL trained only on the neuro dataset

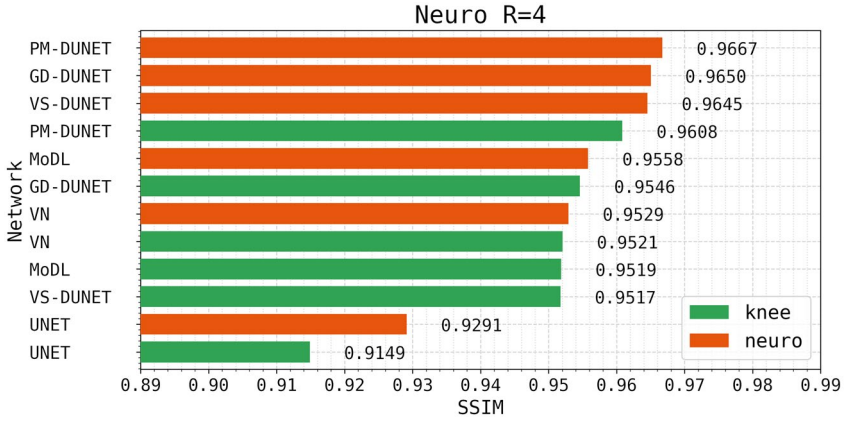

FIGURE 4 Ranked list for the fastMRI neuro dataset at $R = 4$ trained with knee and neuro datasets. All reconstruction networks perform superior than the post-processing networks. PM-DUNET trained on the neuro dataset performs best

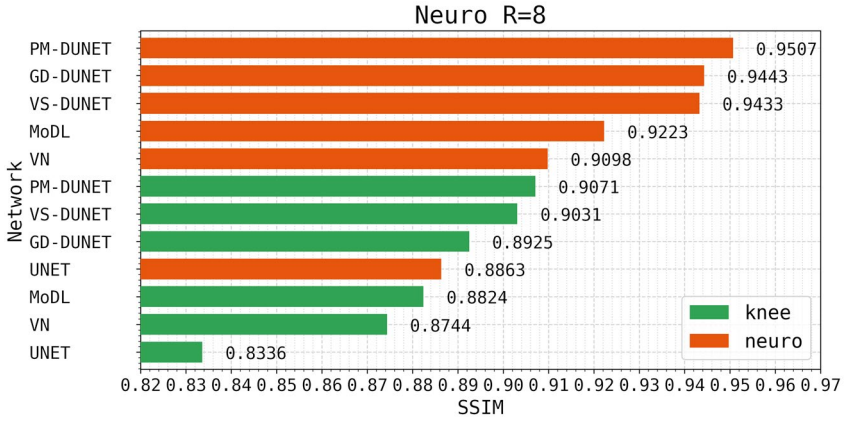

FIGURE 5 Ranked list for the fastMRI neuro dataset at R=8 trained with knee and neuro datasets. PM-DUNET trained only on the knee dataset cannot compete with other state-of-the-art approaches trained on neuro data. Post-processing UNET trained on only neuro data performs superior than many networks containing knee data only. All networks trained on knee data fail for this dataset and acceleration factor.

Supporting Information Figure S6. The UNET result appears blurry, however, it has less artifacts than MoDL and VN which have difficulties to reconstruct images at this high acceleration factor. The DUNETs are able to reconstruct the images with high quality when trained with knee data. The VN results between the best and worst performing network do not differ greatly. The UNET reconstructions appear artificial when trained on neuro data. All DUNETs show artifacts when trained on neuro data, however, the anatomy itself does not change.

Figure 8 shows example results for an axial T_1w scan and $R = 4$. The VN shows the most artifacts of the reconstruction networks. The drop in image quality between different datasets is lowest for MoDL and PM-DUNET, supported by the difference images in Supporting Information Figure S7. GD-DUNET and VS-DUNET show severe artifacts in the reconstructions when trained on knee 25 data. The post-processing UNET stays close to the zero filling solution when trained with the wrong data.

Results for a selected axial T_1w post contrast scan for $R = 8$ is illustrated in Figure 9 along with the difference images in Supporting Information Figure S8. Reconstruction DUNETs trained with neuro data show the best image quality at this high acceleration factor. UNET cannot reconstruct this

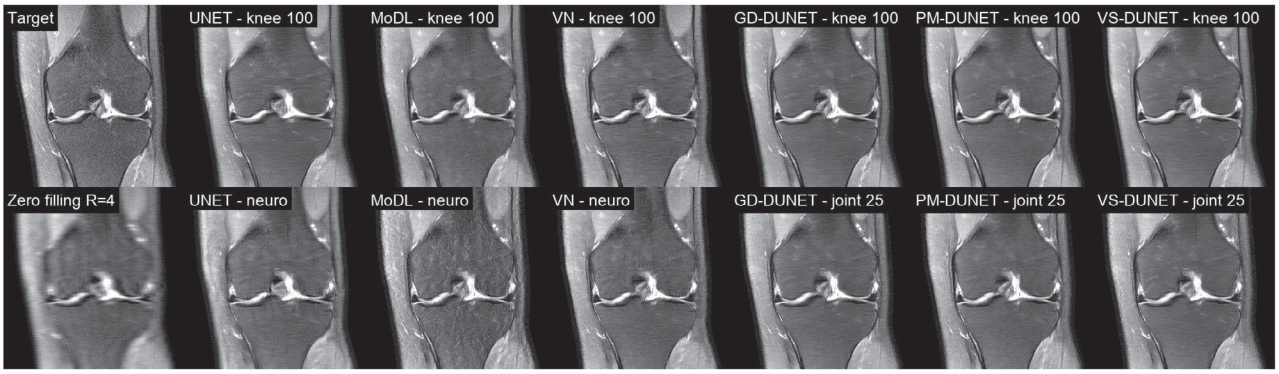

FIGURE 6 Coronal PDw with fat saturation, $R = 4$ (`file1001188.h5`, slice 21): The first column shows the target (top) and zero filling reconstruction (bottom). Columns 2-7 show the reconstruction results for the best performing training dataset (top) and worst performing training dataset (bottom). Note that the anatomy changes in the intercondylar notch for the post-processing UNET. DUNET with different DC layers perform best for all cases. MoDL shows the most performance decrease when trained on neuro data

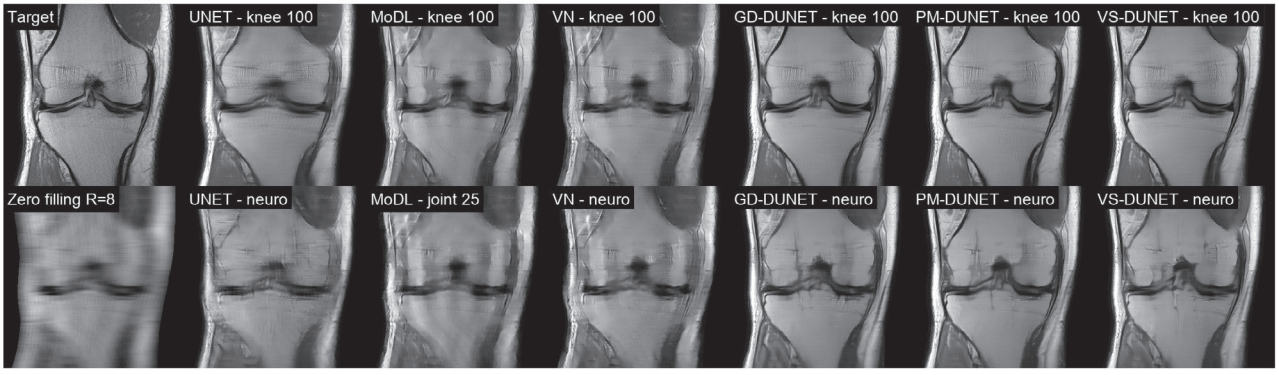

FIGURE 7 Coronal PDw, $R = 8$ (`file1000432.h5`, slice 20): The first column shows the target (top) and zero filling reconstruction (bottom). Columns 2-7 show the reconstruction results for the best performing training dataset (top) and worst performing training dataset (bottom). Both MoDL and VN fail for this high acceleration factor, while the DUNET with different DC layers are able to reconstruct the images. However, when DUNETs are trained with neuro data, the artifacts cannot be fully removed. Note the great change in anatomy for the post-processing UNET when trained with neuro data

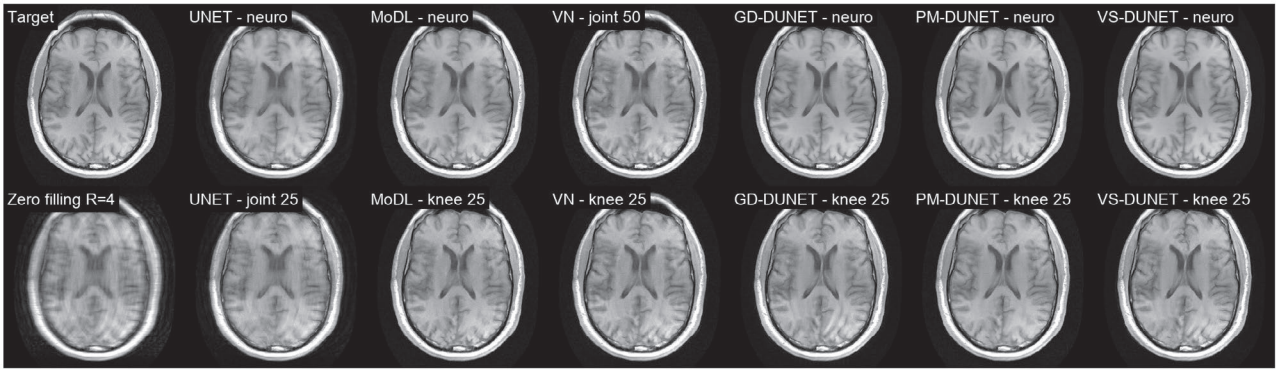

FIGURE 8 Axial T_1w, $R = 4$ (`file_brain_AXT1PRE_203_6000649.h5`, slice 2): The first column shows the target (top) and zero filling reconstruction (bottom). Columns 2-7 show the reconstruction results for the best performing training dataset (top) and worst performing training dataset (bottom). The post-processing UNET yields blurry results while the VN contains remaining artifacts. MoDL is less sharp than the reconstruction DUNETs. However, when trained with knee 25 data, GD-DUNET, VS-DUNET, and VN show severe artifacts while PM-DUNET reconstructs a reasonable result. The UNET stays similar to the zero filling solution when trained with joint 25 data

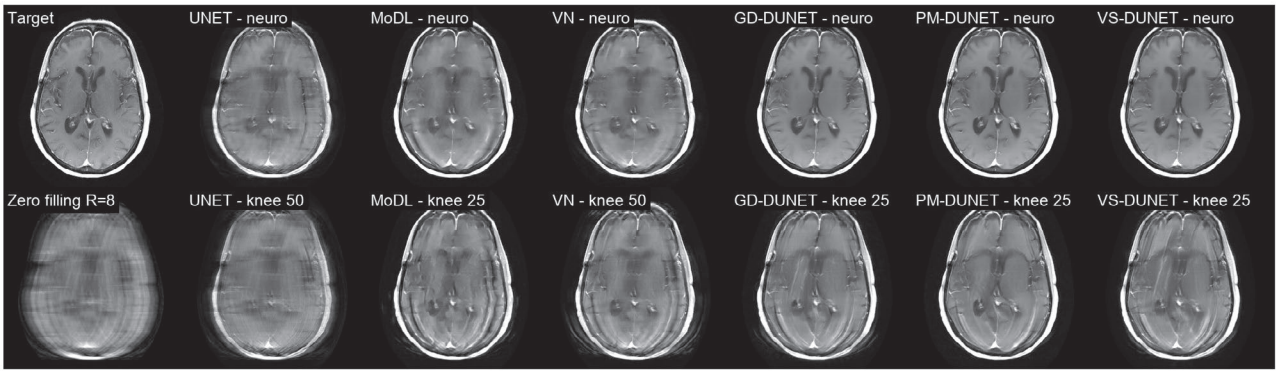

FIGURE 9 Axial T_1w post contrast, $R = 8$ (`file_brain_AXT1POST_200_6002237.h5`, slice 1): The first column shows the target (top) and zero filling reconstruction (bottom). Columns 2-7 show the reconstruction results for the best performing training dataset (top) and worst performing training dataset (bottom). The DUNETs with varying DC show the best results when trained with neuro data. However, if they are trained on knee 25 data, even the ventricles disappear and instead an artificial structure resembling a knee appears

example. MoDL and VN show severe artifacts at this acceleration factor. When trained with wrong data, the shape of the ventricles changes in all reconstruction DUNETs.

5 | DISCUSSION

In this work, we investigate the performance and limits of deep neural networks with respect to different design parameters, including regularization networks, DC layers, and data variations in a controlled, experimental setup. Specifically, we compare three state-of-the architectures, namely UNET[21] (no DC), MoDL[14] (5-layer CNN, PM as DC), and VN[18] (fields of experts model, GD as DC) to our proposed DUNETs with GD, PM, and VS as DC. We deploy a challenging setup, where we train on all contrasts and acceleration factors simultaneously to study the robustness of all networks. This stands in contrast to tremendous amount of research that is conducted to improve the accuracy of deep neural networks on benchmarks,[24,31,46,51] and to overcome limitations of CS approaches, for a specific anatomy or acceleration factor. Our evaluation simulates a very common scenario in medical imaging where the source and type of images and acceleration factors might be unknown, or only limited ground truth data, but diverse test data might be available.

Up to now, the robustness of neural networks to training data has not been studied in literature, although this is a crucial part for a successful clinical translation of MRI reconstruction. Recent work focused on the robustness of sampling trajectories,[52] the robustness to noise levels and image contrast,[39] the effect of diverse MSK anatomies,[20] or performed instability analysis of neural networks with respect to image perturbations.[40] However, a drawback of these approaches is that the size of the dataset is limited, and datasets are homogeneous. Hence, the robustness of these approaches to large, inhomogeneous datasets is unknown.

Variations in DC

We first discuss the impact of different DC layers, with an expressive DUNET as regularization network. The results depict that the differences between DC layers for acceleration factor $R = 4$ are minor. We observe that the PM-DUNET performs most stable over different training datasets, independent of the number and type of training samples, especially at $R = 8$. For knee data and $R = 4$, PM-DUNET trained on neuro data even outperforms the best reported state-of-the-art method. The implicit DC step in the PM allows the network to use a larger regularization parameter λ, resulting in stronger DC. A GD layer would require a smaller stepsize, hence, more iterations to impose the same λ. However, comparing DC layers with respect to unrolled iterations is out of scope of this paper and was studied previously in Ref. 14. Indeed, our results show that GD-DUNET and VS-DUNET are more sensitive to the type and amount of training data. VS-DUNET in general performs worse than GD-DUNET, which stands in contrast to the results reported in Ref. 16. This can be explained by the inhomogeneous dataset which makes it more challenging to tune the parameters α and β.

Variations in regularization networks

Interestingly, the behavior of DC layers cannot be directly transferred to networks with a less expressive regularization network deployed in VN and MoDL. While MoDL performs superior than VN on the neuro dataset, MoDL performs inferior than VN on the knee dataset, especially for the fat saturated knee data. In our study, MoDL is more sensitive to the content of the training dataset. We believe that this is due to the small CNN regularization and a strong PM DC, as this setup cannot capture the inhomogeneity between acceleration factors, SNR levels and

anatomies. This indicates that expressive regularization networks such as the DUNETs are able to compensate for inhomogeneities in the data.

Importance of DC

The ranked list visualizations show that all networks with DC perform superior than the UNET without DC for $R = 4$, independent of type and amount of training data. It is impressive that VN and MoDL only have 4% of parameters compared to UNET without DC, but they outperform UNET substantially for $R = 4$. For $R = 8$, UNET achieves superior quantitative results compared to VN and MoDL when trained on mismatched anatomy. However, we observe changes in anatomy for the UNET, as shown in Figure 7. These findings indicated that modeling the acquisition physics in the DC is more important than a large amount of training data, for successful learning in MRI reconstruction.

Robustness to variations in data

Our results indicate that domain shift is less of an issue at $R = 4$, and more general training data settings can be used to achieve decent reconstruction quality. For $R = 8$, the increase in dataset size tend to help but blindly increasing the dataset is not enough to account for the large domain shift. This becomes clear when examining the results for DUNET trained with knee 25 data in Figure 9. We observe a structure in the brain that resembles more a knee structure than brain ventricles, and we suspect overfitting to the small knee dataset. Hence, including aligned data for training and testing is more helpful at $R = 8$, even if the dataset is small in size. Additionally, the influence of DC is limited for high acceleration factors as less information is available in k-space. This rises the concern if structures are invented by the networks if the acceleration factor is pushed too far. Hence, both theoretical studies and radiologists' evaluations are required to estimate the limits of acceleration.

Radiologists' evaluation are also required to assess image quality. For training the networks, we often use global, quantitative measures, that is, MSE and SSIM, which represent the human perceptual system poorly. These measures cannot account for important local features such as subtle pathologies, and often result in blurry images.[47,51] However, training on different loss functions, including adversarial losses and unsupervised training, is out of scope of this paper. Furthermore, we did not investigate transfer learning[39,53] or fine-tuning on individual contrasts and acceleration factors in this study, which can further improve image quality.

Our work provides first insights into the robustness of neural networks for MRI reconstruction from an experimental perspective, where we cover a large-scale evaluation with respect to variations in training data, regularization networks, and DC layers. There are also other sources of variation that are not covered in our experiments, including variations in unrolled iterations,[14] number of features, loss functions,[24] and the influence of optimizers in training.[24] Furthermore, it still remains an open question how robust networks are to variations in SNR, field strength, scanner types, and hardware from different vendors, which have to be investigated in future work.

In the present work, we focused on static 2D imaging, but the basics of DC and regularization networks can be directly translated to higher-dimensional image reconstruction, for example, dynamic imaging. However, the presented DC and regularization networks might not be sufficient to exploit complex temporal dynamics. Recent works focused, for example, on improving the regularization by exploiting spatiotemporal redundancies in X-f domain,[54] or combining MoDL with a SmooThness regularization on manifolds (SToRM) prior for dynamic imaging.[55] We believe that additionally integrating advanced DC schemes, for example, motion-corrected DC,[56] or combining the classic DC term with temporal models[57] will be key ingredients to further improve dynamic reconstruction with deep learning.

6 | CONCLUSION

Large-scale studies are indispensable to assess the application potential of neural networks in clinical workflow, where we have to deal with both limited and inhomogeneous datasets. In our work, we experimentally validate hypotheses about acceleration limits, properties, and the robustness of neural networks to sources of variation in DC, regularization networks and training data. Our findings underpin the importance of DC layers, and suggest that PM[14] together with an expressive regularization network, that is, the proposed DUNET, leads to the most stable results over a wide range of training scenarios. For low acceleration factors, general and robust networks can be learned that do not depend substantially on the type and amount of training data. For high acceleration factors, the results are impressive only if train and test domain are aligned. Although we get the impression that neural networks add more details to the reconstruction, we should be aware that they cannot recover high-frequency information that has not been captured in the acquisition process.

ACKNOWLEDGMENTS

The work was funded in part by the EPSRC Programme Grant (EP/P001009/1) and by the Intramural Research Programs of the National Institutes of Health Clinical Center (1Z01 CL040004).

DATA AVAILABILITY STATEMENT

To support the findings of our manuscript, we make all Pytorch source code available at https://github.com/khammernik/sigmanet, including the implemented state-of-the-art approaches. The data that support the findings of this study are available from the fastMRI research project at https://fastmri.org. The estimated foreground masks will be provided upon request.

ORCID

Kerstin Hammernik https://orcid.org/0000-0002-2734-1409
Jo Schlemper https://orcid.org/0000-0003-1867-1155
Chen Qin https://orcid.org/0000-0003-3417-3092
Jinming Duan https://orcid.org/0000-0002-5108-2128
Ronald M. Summers https://orcid.org/0000-0001-8081-7376
Daniel Rueckert https://orcid.org/0000-0002-5683-5889

TWITTER

Kerstin Hammernik @khammernik

REFERENCES

1. Sodickson DK, Manning WJ. Simultaneous acquisition of spatial harmonics (SMASH): fast imaging with radiofrequency coil arrays. *Magn Reson Med*. 1997;38:591-603.
2. Pruessmann KP, Weiger M, Scheidegger MB, Boesiger P. SENSE: sensitivity encoding for fast MRI. *Magn Reson Med*. 1999;42:952-962.
3. Griswold MA, Jakob PM, Heidemann RM, et al. Generalized autocalibrating partially parallel acquisitions (GRAPPA). *Magn Reson Med*. 2002 6;47:1202-1210.
4. Lustig M, Donoho D, Pauly JM. Sparse MRI: the application of compressed sensing for rapid MR imaging. *Magn Reson Med*. 2007;58:1182-1195.
5. Block KT, Uecker M, Frahm J. Undersampled radial MRI with multiple coils. Iterative image reconstruction using a total variation constraint. *Magn Reson Med*. 2007;57:1086-1098.
6. Liang D, Liu B, Wang J, Ying L. Accelerating SENSE using compressed sensing. *Magn Reson Med*. 2009;62:1574-1584.
7. Haldar JP. Low-rank modeling of local *k*-space neighborhoods (LORAKS) for constrained MRI. *IEEE Trans Med Imaging*. 2013;33:668-681.
8. Shin PJ, Larson PEZ, Ohliger MA, et al. Calibrationless parallel imaging reconstruction based on structured low-rank matrix completion. *Magn Reson Med*. 2014;72:959-970.
9. Knoll F, Clason C, Bredies K, Uecker M, Stollberger R. Parallel imaging with nonlinear reconstruction using variational penalties. *Magn Reson Med*. 2012;67:34-41.
10. Knoll F, Hammernik K, Zhang C, et al. Deep-learning methods for parallel magnetic resonance imaging reconstruction: a survey of the current approaches, trends, and issues. *IEEE Signal Process Mag*. 2020;37:128-140.
11. Hammernik K, Knoll F. Machine learning for image reconstruction. In: Rueckert D, Fichtinger G, Zhou SK, eds. *Handbook of Medical Image Computing and Computer Assisted Intervention*, Academic Press; 2018:25-64.
12. Liang D, Cheng J, Ke Z, Ying L. Deep magnetic resonance image reconstruction: inverse problems meet neural networks. *IEEE Signal Process Mag*. 2020;37:141-151.
13. Lundervold AS, Lundervold A. An overview of deep learning in medical imaging focusing on MRI. *Z Med Phys*. 2019;29:102-127.
14. Aggarwal HK, Mani MP, Jacob M. MoDL: model based deep learning architecture for inverse problems. *IEEE Trans Med Imaging*. 2019;38:394-405.
15. Akçakaya M, Moeller S, Weingärtner S, Uğurbil K. Scan-specific robust artificial-neural-networks for k-space interpolation (RAKI) reconstruction: database-free deep learning for fast imaging. *Magn Reson Med*. 2019;81:439-453.
16. Duan J, Schlemper J, Qin C, et al. VS-net: variable splitting network for accelerated parallel MRI reconstruction. In: Shen D, Liu T, Peters TM, Staib LH, Essert C, Zhou S, Yap PT, Khan A, eds. *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer International Publishing; 2019:713-722.
17. Eo T, Jun Y, Kim T, Jang J, Lee HJ, Hwang D. KIKI-net: cross-domain convolutional neural networks for reconstructing undersampled magnetic resonance images. *Magn Reson Med*. 2018;80:2188-2201.
18. Hammernik K, Klatzer T, Kobler E, et al. Learning a variational network for reconstruction of accelerated MRI data. *Magn Reson Med*. 2018;79:3055-3071.
19. Hyun CM, Kim HP, Lee SM, Lee S, Seo JK. Deep learning for undersampled MRI reconstruction. *Phys Med Biol*. 2018;63:135007.
20. Johnson PM, Muckley MJ, Bruno M, et al. Joint multi-anatomy training of a variational network for reconstruction of accelerated magnetic resonance image acquisitions. In: Knoll F, Maier A, Rueckert D, Ye JC, eds. *Machine Learning for Medical Image Reconstruction*, Springer; 2019:71-79.
21. Knoll F, Zbontar J, Sriram A, et al. fastMRI: a publicly available raw k-space and DICOM dataset of knee images for accelerated MR image reconstruction using machine learning. *Radiol Artif Intell*. 2020;2:e190007.
22. Lee D, Yoo J, Tak S, Ye J. Deep residual learning for accelerated MRI using magnitude and phase networks. *IEEE Trans Biomed Eng*. 2018;65:1985-1995.
23. Mardani M, Gong E, Cheng JY, et al. Deep generative adversarial neural networks for compressive sensing (GANCS) MRI. *IEEE Trans Med Imaging*. 2019;38:167-179.
24. Pezzotti N, Yousefi S, Elmahdy MS, et al. An adaptive intelligence algorithm for undersampled knee MRI reconstruction. *IEEE Access*. 2020;8:204825-204838.
25. Putzky P, Welling M. Invert to learn to invert. In: Wallach H, Larochelle H, Beygelzimer A, d'Alché-Buc F, Fox E, Garnett R, eds. *Advances in Neural Information Processing Systems*. Curran Associates, Inc; 2019:444-454.
26. Qin C, Schlemper J, Caballero J, Price AN, Hajnal JV, Rueckert D. Convolutional recurrent neural networks for dynamic MR image reconstruction. *IEEE Trans Med Imaging*. 2019;38:280-290.
27. Quan TM, Nguyen-Duc T, Jeong WK. Compressed sensing MRI reconstruction using a generative adversarial network with a cyclic loss. *IEEE Trans Med Imaging*. 2018;37:1488-1497.
28. Ran M, Xia W, Huang Y, et al. MD-recon-net: a parallel dual-domain convolutional neural network for compressed sensing MRI. *IEEE Trans Radiation Plasma Med Sci*. 2021;5:120-135.

29. Schlemper J, Caballero J, Hajnal JV, Price AN, Rueckert D. A deep cascade of convolutional neural networks for dynamic MR image reconstruction. *IEEE Trans Med Imaging*. 2018;37:491-503.
30. Seitzer M, Yang G, Schlemper J, et al. Adversarial and perceptual refinement for compressed sensing MRI reconstruction. In: Frangi AF, Schnabel JA, Davatzikos C, Alberola-López C, Fichtinger G, eds. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Vol. 11070. Springer International Publishing; 2018:232-240.
31. Sriram A, Zbontar J, Murrell T, Zitnick CL, Defazio A, Sodickson DK. GrappaNet: combining parallel imaging with deep learning for multi-coil MRI reconstruction. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*; 2020:14315-14322.
32. Wang P, Chen EZ, Chen T, Patel VM, Sun S. Pyramid convolutional RNN for MRI reconstruction. arXiv preprint arXiv:191200543 2019.
33. Yang Y, Sun J, Li H, Xu Z. Deep ADMM-net for compressive sensing MRI. In: Lee D, Sugiyama M, Luxburg U, Guyon I, Garnett R, eds. *Advances in Neural Information Processing Systems*. Curran Associates, Inc; 2016:10-18.
34. Yang G, Yu S, Dong H, et al. DAGAN: deep de-aliasing generative adversarial networks for fast compressed sensing MRI reconstruction. *IEEE Trans Med Imaging*. 2018;37:1310-1321.
35. Zhang P, Wang F, Xu W, Li Y. Multi-channel generative adversarial network for parallel magnetic resonance image reconstruction in K-space. 2018:180-188. https://doi.org/10.1007/978-3-030-00928-1_21
36. Zhu B, Liu JZ, Cauley SF, Rosen BR, Rosen MS. Image reconstruction by domain-transform manifold learning. *Nature*. 2018;555:487-492.
37. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells WM, Frangi AF, eds. *International Conference on Medical Image Computing and Computer Assisted Intervention*; 2015:234-241.
38. Roth S, Black MJ. Fields of experts. *Int J Comput Vis*. 2009;82:205-229.
39. Knoll F, Hammernik K, Kobler E, Pock T, Recht MP, Sodickson DK. Assessment of the generalization of learned image reconstruction and the potential for transfer learning. *Magn Reson Med*. 2019;81:116-128.
40. Antun V, Renna F, Poon C, Adcock B, Hansen AC. On instabilities of deep learning in image reconstruction and the potential costs of AI. *Proc Natl Acad Sci*. 2020;117:30088-30095.
41. Uecker M, Lai P, Murphy MJ, et al. ESPIRiT–an eigenvalue approach to autocalibrating parallel MRI: where SENSE meets GRAPPA. *Magn Reson Med*. 2014;71:990-1001.
42. Yu S, Park B, Jeong J. Deep iterative down-up CNN for image denoising. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops; 2019.
43. Zhang Y, Li K, Li K, Wang L, Zhong B, Fu Y. Image super-resolution using very deep residual channel attention networks. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Vol. 11211. LNCS; 2018:294-310.
44. Haris M, Shakhnarovich G, Ukita N. Deep back-projection networks for super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2018:1664-1673.
45. Shi W, Caballero J, Huszar F, et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural networkk. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition; 2016:1874-1883.
46. Hammernik K, Schlemper J, Qin C, Duan J, Summers RM, Rueckert D. Σ-net: ensembled iterative deep neural networks for accelerated parallel MR image reconstruction. In: Proceedings of the ISMRM & SMRT Virtual Conference & Exhibition, 2020; 2020:0987.
47. Zhao H, Gallo O, Frosio I, Kautz J. Loss functions for image restoration with neural networks. *IEEE Trans Comput Imaging*. 2016;3:47-57.
48. Hammernik K, Knoll F, Sodickson D, Pock T. L2 or not L2: impact of loss function design for deep learning MRI reconstruction. In: Proceedings of the International Society of Magnetic Resonance in Medicine; 2017:687.
49. Kingma DP, Ba J. Adam: a method for stochastic optimization. In: Proceedings of the International Conference on Learning Representations (ICLR); 2015.
50. Mylavarapu P, Yalcin A, Gregg X, Elmqvist N. Ranked-list visualization: a graphical perception study. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems CHI '19, New York, NY, USA: Association for Computing Machinery; 2019:1-12.
51. Knoll F, Murrell T, Sriram A, et al. Advancing machine learning for MR image reconstruction with an open competition: overview of the 2019 fastMRI challenge. *Magn Reson Med*. 2020;84:3054-3070.
52. Liu F, Samsonov A, Chen L, Kijowski R, Feng L. SANTIS: sampling-augmented neural network with incoherent structure for MR image reconstruction. *Magn Reson Med*. 2019;82:1890-1904.
53. Han Y, Yoo J, Kim HH, Shin HJ, Sung K, Ye JC. Deep learning with domain adaptation for accelerated projection-reconstruction MR. *Magn Reson Med*. 2018;80:1189-1205.
54. Qin C, Schlemper J, Duan J, et al. k-t NEXT: dynamic MR image reconstruction exploiting spatio-temporal correlations. In: International Conference on Medical Image Computing and Computer-Assisted Intervention Springer; 2019:505-513.
55. Biswas S, Aggarwal HK, Jacob M. Dynamic MRI using model-based deep learning and SToRM priors: MoDL-SToRM. *Magn Reson Med*. 2019;82:485-494.
56. Batchelor PG, Atkinson D, Irarrazaval P, Hill DLG, Hajnal J, Larkman D. Matrix description of general motion correction applied to multishot images. *Magn Reson Med*. 2005;54:1273-1280.
57. Velikina JV, Samsonov AA. Reconstruction of dynamic image series from undersampled MRI data using data-driven model consistency condition (MOCCO). *Magn Reson Med*. 2015;74:1279-1290.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

FIGURE S1 Ranked list for the fastMRI knee dataset at $R = 4$ trained with different configurations of knee and joint knee/neuro data. All reconstruction networks perform superior than the post-processing networks. GD-DUNET trained on the knee 100 dataset performs best

FIGURE S2 Ranked list for the fastMRI knee dataset at $R = 8$ trained with differentt different configurations of knee and joint knee/neuro data. PM-DUNET trained on the knee 100 dataset performs best. Joint training with a low number of datasets, i.e., joint 25 and joint uni 25, drops the performance of MoDL more compared to VN. MoDL outperforms VN for trainings with only knee data

FIGURE S3 Ranked list for the fastMRI neuro dataset at $R = 4$ trained with different configurations of knee and joint knee/neuro data. All reconstruction networks perform superior than the post-processing networks. UNET trained with uniformly distributed knee and neuro data (joint uni 100) performs better than the network with non-uniformely distributed knee and neuro data. The type and amount of training data is less critical for reconstruction networks at this acceleration

FIGURE S4 Ranked list for the fastMRI neuro dataset at $R = 8$ trained with different configurations of knee and joint knee/neuro data. PM-DUNET trained only on the knee dataset cannot compete with other state-of-the-art approaches containing neuro data. All networks trained only on knee data fail for this dataset and acceleration factor

FIGURE S5 Difference images (×5) to Figure 6 for coronal PDw with fat saturation, $R = 4$ (file1001188.h5, slice 21). Columns 1-6 show the difference to the target for the best performing training dataset (top) and worst performing training dataset (bottom). DUNETs with varying DC layers perform slightly better than VN and MoDL when trained with knee data. The performance decrease in MoDL when trained with neuro data is most obvious

FIGURE S6 Difference images (×5) to Figure 7 for coronal PDw, $R = 8$ (file1000432.h5, slice 20). Columns 1-6 show the difference to the target for the best performing training dataset (top) and worst performing training dataset (bottom). MoDL and VN fail for this high acceleration factor. DUNETs with varying DC outperform the other state-of-the-art approaches substantially in all cases. The overall difference for the post-processing UNET is surprising, when trained on knee data

FIGURE S7 Difference images (×5) to Figure 8 for axial T1w, $R = 8$ (file_brain_AXT1PRE_203_6000649.h5, slice 2). Columns 1-6 show the difference to the target for the best performing training dataset (top) and worst performing training dataset (bottom). The post-processing UNET has the worst reconstruction error. All DUNETs outperform the other reconstruction approaches when trained on neuro data. When trained with the wrong data, PM-DUNET yields the best results for this contrast and acceleration factor

FIGURE S8 Difference images (×5) to Figure 9 for axial T1w post contrast, $R = 8$ (file_brain_AXT1POST_200_6002237.h5, slice 1). Columns 1-6 show the difference to the target for the best performing training dataset (top) and worst performing training dataset (bottom). The post-processing UNET has the worst reconstruction error. All DUNETs outperform the other reconstruction approaches when trained on neuro data. When trained with the wrong data, PM-DUNET yields the best results for this contrast and acceleration factor

TABLE S1 Number of parameters and reconstruction times averaged over 20 knee slices for the presented network architectures. The reconstruction times are measured using an NVIDIA Titan Xp (12 GB) for a fastMRI knee dataset of size 640×368 acquired with 15 coils[21]

How to cite this article: Hammernik K, Schlemper J, Qin C, et al. Systematic evaluation of iterative deep neural networks for fast parallel MRI reconstruction with sensitivity-weighted coil combination. *Magn Reson Med*. 2021;86:1859–1872. https://doi.org/10.1002/mrm.28827