Data augmentation for deep learning based accelerated MRI reconstruction with limited data

Zalan Fabian [1] **Reinhard Heckel** [2] **Mahdi Soltanolkotabi** [1]

Abstract

Deep neural networks have emerged as very successful tools for image restoration and reconstruction tasks. These networks are often trained end-to-end to directly reconstruct an image from a noisy or corrupted measurement of that image. To achieve state-of-the-art performance, training on large and diverse sets of images is considered critical. However, it is often difficult and/or expensive to collect large amounts of training images. Inspired by the success of Data Augmentation (DA) for classification problems, in this paper, we propose a pipeline for data augmentation for accelerated MRI reconstruction and study its effectiveness at reducing the required training data in a variety of settings. Our DA pipeline, MRAugment, is specifically designed to utilize the invariances present in medical imaging measurements as naive DA strategies that neglect the physics of the problem fail. Through extensive studies on multiple datasets we demonstrate that in the low-data regime DA prevents overfitting and can match or even surpass the state of the art while using significantly fewer training data, whereas in the high-data regime it has diminishing returns. Furthermore, our findings show that DA can improve the robustness of the model against various shifts in the test distribution.

1. Introduction

In magnetic resonance imaging (MRI), an extremely popular medical imaging technique, it is common to reduce the acquisition time by subsampling the measurements, because this reduces cost and increases accessibility of MRI to patients. Due to the subsampling, there are fewer equations than unknowns, and therefore the signal is not uniquely identifiable from the measurements. To overcome this challenge there has been a flurry of activity over the last decade aimed at utilizing prior knowledge about the signal, in a research area referred to as *compressed sensing* (Candes et al., 2006; Donoho, 2006).

Compressed sensing methods reduce the required number of measurements by utilizing prior knowledge about the images during the reconstruction process, traditionally via a convex regularization that enforces sparsity in an appropriate transformation of the image. More recently, deep learning techniques have been used to enforce much more nuanced forms of prior knowledge (see Ongie et al. (2020) and references therein for an overview). The most successful of these approaches aim to directly learn the inverse mapping from the measurements to the image by training on a large set of training data consisting of signal/measurement pairs. This approach often enables faster reconstruction of images, but more importantly, deep learning techniques yield significantly higher quality reconstructions. Thus, deep learning techniques enable reconstructing a high-quality image from fewer measurements which further reduces image acquisition times. For instance, in an accelerated MRI competition known as fastMRI Challenge (Zbontar et al., 2018), all the top contenders used deep learning reconstruction techniques.

Contrary to classical compressive sensing approaches, however, deep learning techniques typically rely on large sets of training data consisting of images along with the corresponding measurement. This is also true about the use of deep learning techniques in other areas such as computer vision and Natural Language Processing (NLP) where superb empirical success has been observed. While large datasets have been harvested and carefully curated in areas such as vision and NLP, this is not feasible in many scientific applications including MRI. It is difficult and expensive to collect the necessary datasets for a variety of reasons, including patient confidentiality requirements, cost and time of data acquisition, lack of medical data compatibility standards, and the rarity of certain diseases.

A common strategy to reduce reliance on training data in

[1]Department of Electrical and Computer Engineering, University of Southern California, Los Angeles, California, USA [2]Department of Electrical and Computer Engineering, Technical University of Munich, Munich, Germany, and Department of Electrical and Computer Engineering, Rice University. Correspondence to: Mahdi Soltanolkotabi <soltanol@usc.edu>.

Proceedings of the 38^{th} International Conference on Machine Learning, PMLR 139, 2021.

classification tasks is data augmentation. Data augmentation techniques are used in classification tasks to significantly increase the performance on standard benchmarks such as ImageNet and CIFAR-10. For a comprehensive survey of image data augmentation in deep learning see (Shorten & Khoshgoftaar, 2019). More specific to medical imaging, data augmentation techniques have been successfully applied to registration, classification and segmentation of medical images. Recently, several studies (Zhao et al., 2020b; Karras et al., 2020; Zhao et al., 2020a) have demonstrated that data augmentation can significantly reduce the data needed for GAN training for high quality image generation. In a classification setting, data augmentation consists of adding additional synthetic data obtained by performing *invariant* alterations to the data (e.g. flips, translations, or rotations) which do not affect the response (i.e., the label).

In image reconstruction tasks, however, data augmentation techniques are less common and much more difficult to design because the response (the measurement) is affected by the data augmentation. For example, measurements of a rotated image are not the same as measurements from the original image. In the context of accelerated MRI reconstruction, augmentation techniques such as randomly generated undersampling masks (Liu et al., 2019) and simple random flipping (Lee et al., 2018) have been applied, and authors in Schlemper et al. (2017) note the importance of rigid transforms in avoiding overfitting. However, an effective pipeline of augmentations for training data reduction and thorough experimental studies thereof are lacking.

The goal of this paper is to explore the benefits of data augmentation techniques for accelerated MRI with limited training data. By carefully taking into account the physics behind the MRI acquisition process we design a data augmentation pipeline, which we call MRAugment [1], that can successfully reduce the amount of training data required. Our contributions are as follows:

- We perform an extensive empirical study of data augmentation in accelerated MRI reconstruction. To the best of our knowledge, this work is the first in-depth experimental investigation focusing on the benefits of data augmentation in the context of training data reduction for accelerated MRI.

- We propose a data augmentation technique tailored to the physics of the MR reconstruction problem. It is not obvious how to perform data augmentation in the context of accelerated MRI or in inverse problems in general, because by changing an image to enlarge the training set, we do not automatically get a corresponding measurement, contrary to classification problems, where the label is retained.

- We demonstrate the effectiveness of MRAugment on a variety of datasets. On small datasets we report significant improvements in reconstruction performance on the full dataset when MRAugment is applied. Moreover, on small datasets we are able to surpass full dataset baselines by using only a small fraction of the available training data by leveraging our proposed data augmentation technique.

- We perform an extensive study of MRAugment on a large benchmark accelerated MRI data set, specifically on the fastMRI (Zbontar et al., 2018) dataset. For 8-fold acceleration and multi-coil measurements (multi-coil measurements are the standard acquisition mode for clinical practice) we achieve performance on par with the state of the art with only 10% of the training data. Similarly, again for 8-fold acceleration and single-coil experiments (an acquisition mode popular for experimentation) MRAugment can achieve the performance of reconstruction methods trained on the entire dataset while using only 33% of training data.

- We reveal additional benefits of data augmentation on model robustness in a variety of settings. We observe that MRAugment has the potential to improve generalization to unseen MRI scanners, field strengths and anatomies. Furthermore, due to the regularizing effect of data augmentation, MRAugment prevents overfitting to training data and therefore may help eliminate hallucinated features on reconstructions, an unwanted side-effect of deep learning based reconstruction.

2. Background and Problem Formulation

MRI is a medical imaging technique that exploits strong magnetic fields to form images of the anatomy. MRI is a prominent imaging modality in diagnostic medicine and biomedical research because it does not expose patients to ionizing radiation, contrary to competing technologies such as computed and positron emission tomography.

However, performing an MR scan is time intensive, which is problematic for the following reasons. First, patients are exposed to long acquisition times in a confined space with high noise levels. Second, long acquisition times induce reconstruction artifacts caused by patient movement, which sometimes requires patient sedation in particular in pediatric MRI (Vasanawala et al., 2010). Reducing the acquisition time can therefore increase both the accuracy of diagnosis and patient comfort. Furthermore, decreasing the acquisition time needed allows more patients to receive a scan using the same machine. This can significantly reduce patient cost, since each MRI machine comes with a high cost to maintain and operate.

Since the invention of MR in the 1980s there has been tremendous research focusing on reducing their acquisition

[1] Code: https://github.com/MathFLDS/MRAugment

time. The two main ideas are to i) perform multiple acquisitions simultaneously (Sodickson & Manning, 1997; Pruessmann et al., 1999; Griswold et al., 2002) and to ii) subsample the measurements, known as accelerated acquisition or compressed sensing (Lustig et al., 2008). Most modern scanners combine both techniques, and therefore we consider such a setup.

2.1. Accelerated MRI acquisition

In magnetic resonance imaging, measurements of a patient's anatomy are acquired in the Fourier-domain, also called *k-space*, through receiver coils. In the *single-coil* acquisition mode, the k-space measurement $\boldsymbol{k} \in \mathbb{C}^n$ of a complex-valued ground truth image $\boldsymbol{x}^* \in \mathbb{C}^n$ is given by

$$\boldsymbol{k} = \mathcal{F}\boldsymbol{x}^* + \boldsymbol{z},$$

where $\mathcal{F}$ is the two-dimensional Fourier-transform, and $\boldsymbol{z} \in \mathbb{C}^n$ denotes additive noise arising in the measurement process. In parallel MR imaging, multiple receiver coils are used, each of which captures a different region of the image, represented by a complex-valued sensitivity map $\boldsymbol{S}_i$. In this *multi-coil* setup, coils acquire k-space measurements modulated by their corresponding sensitivity maps:

$$\boldsymbol{k}_i = \mathcal{F}\boldsymbol{S}_i\boldsymbol{x}^* + \boldsymbol{z}_i, \quad i = 1, .., N,$$

where N is the number of coils. Obtaining fully-sampled k-space data is time-consuming, and therefore in accelerated MRI we decrease the number of measurements by undersampling in the Fourier-domain. This undersampling can be represented by a binary mask $\boldsymbol{M}$ that sets all frequency components not sampled to zero:

$$\tilde{\boldsymbol{k}}_i = \boldsymbol{M}\boldsymbol{k}_i, \;\; i = 1, .., N.$$

We can write the overall forward map concisely as

$$\tilde{\boldsymbol{k}} = \mathcal{A}\left(\boldsymbol{x}^*\right),$$

where $\mathcal{A}\left(\cdot\right)$ is the linear forward operator and $\tilde{\boldsymbol{k}}$ denotes the undersampled coil measurements stacked into a single column vector. The goal in accelerated MRI reconstruction is to recover the image $\boldsymbol{x}^*$ from the set of k-space measurements $\tilde{\boldsymbol{k}}$. Note that—without making assumptions on the image $\boldsymbol{x}^*$—it is in general impossible to perfectly recover the image, because we have fewer measurements than variables to recover. This recovery problem is known as compressed sensing. To make image recovery potentially possible, recovery methods make structural assumptions about $\boldsymbol{x}^*$, such that it is sparse in some basis or implicitly that it looks similar to images from the training set.

2.2. Traditional accelerated MRI reconstruction methods

Traditional compressed sensing recovery methods for accelerated MRI are based on assuming that the image $\boldsymbol{x}^*$ is sparse in some dictionary, for example the wavelet transform. Recovery is then posed typically as a convex optimization problem:

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \left\| \mathcal{A}\left(\boldsymbol{x}\right) - \tilde{\boldsymbol{k}} \right\|^2 + \mathcal{R}(\boldsymbol{x}),$$

where $\mathcal{R}(\cdot)$ is a regularizer enforcing sparsity in a certain domain. Typical functions used in CS based MRI reconstruction are ℓ_1-wavelet and total-variation regularizers. These optimization problems can be numerically solved via iterative gradient descent based methods.

2.3. Deep learning based MRI reconstruction methods

In recent years, several deep learning algorithms have been proposed and convolutional neural networks established new state of the art in MRI reconstruction significantly surpassing the classical baselines. Encoder-decoder networks such as the U-Net (Ronneberger et al., 2015) and its variants were successfully used in various medical image reconstruction (Hyun et al., 2018; Han & Ye, 2018) and segmentation problems (Çiçek et al., 2016; Zhou et al., 2018). These models consist of two sub-networks: the encoder repeatedly filters and downsamples the input image with learned convolutional filters resulting in a concise feature vector. This low-dimensional representation is then fed to the decoder consisting of subsequent upsampling and learned filtering operations. Another approach that can be considered a generalization of iterative compressed sensing reconstructions consists of unrolling the data flow graph of popular algorithms such as ADMM (Yang et al., 2016) or gradient descent iterations (Zhang & Ghanem, 2018) and mapping them to a cascade of sub-networks. Several variations of this unrolled method have been proposed recently for MR reconstruction, such as i-RIM (Putzky & Welling, 2019), Adaptive-CS-Net (Pezzotti et al., 2019), Pyramid Convolutional RNN (Wang et al., 2019) and E2E VarNet (Sriram et al., 2020).

Another line of work, inspired by the deep image prior (Ulyanov et al., 2018) focuses on using the inductive bias of convolutional networks to perform reconstruction without any training data (Jin et al., 2019; Darestani & Heckel, 2020; Heckel & Soltanolkotabi, 2020; Heckel & Hand, 2019; Van Veen et al., 2018). Those methods do perform significantly better than classical un-trained networks, but do not perform as well as neural networks trained on large sets of training data.

3. MRAugment: a data augmentation pipeline for MRI

In this section we propose our data augmentation technique, MRAugment, for MRI reconstruction. We emphasize that data augmentation in this setup and for inverse problems in general is substantially different from DA for classification problems. For classification tasks, the label of the augmented image is trivially the same as that of the original image, whereas for inverse problems we have to generate both an augmented target image and the corresponding measurements. This is non-trivial as it is critical to match the noise statistics of the augmented measurements with those in the dataset.

We are given training data in the form of fully-sampled MRI measurements in the Fourier domain, and our goal is to generate new training examples consisting of a subsampled k-space measurement along with a target image. MRAugment is model-agnostic in that the generated augmented training example can be used with any machine learning model and therefore can be seamlessly integrated with existing reconstruction algorithms for accelerated MRI, and potentially beyond MRI.

Our data augmentation pipeline, illustrated in Figure 1, generates a new example consisting of a subsampled k-space measurement $\tilde{\boldsymbol{k}}_a$ along with a target image $\bar{\boldsymbol{x}}_a$ as follows. We are given training examples as fully-sampled k-space slices, which we stack into a single column vector $\boldsymbol{k} = \mathrm{col}(\boldsymbol{k}_1, \boldsymbol{k}_2, ..., \boldsymbol{k}_N)$ for notational convenience. From these, we obtain the individual coil images by applying the inverse Fourier transform as $\boldsymbol{x} = \mathcal{F}^{-1}\boldsymbol{k}$. We generate augmented individual coil images with an augmentation function $\mathcal{D}$, specified later, as $\boldsymbol{x}_a = \mathcal{D}(\boldsymbol{x})$. From the augmented images, we generate an undersampled measurement by applying the forward model as $\tilde{\boldsymbol{k}}_a = \mathcal{A}(\boldsymbol{x}_a)$. Both $\boldsymbol{x}$ and $\boldsymbol{x}_a$ are *complex-valued*: even though the MR scanner obtains measurements of a real-valued object, due to noise the inverse Fourier-transform of the measurement is complex-valued. Therefore the augmentation function has to generate complex-valued images, which adds an extra layer of difficulty compared to traditional data augmentation techniques pertaining to real-valued images (see Section 3.1 for further details). Finally, the real-valued ground truth image is obtained by combining the coil images $\boldsymbol{x}_{a,i}$ by pixel-wise root-sum-squares (RSS) followed by center-cropping $\mathcal{C}$:

$$\bar{\boldsymbol{x}}_a = \mathcal{C}\left(\mathrm{RSS}(\boldsymbol{x}_a)\right) = \mathcal{C}\left(\sqrt{\sum_{i=1}^{N} |\boldsymbol{x}_{a,i}|^2}\right).$$

In the following subsections we first argue why we generate individual coil images with the augmentation function, then discuss the design of the augmentation function $\mathcal{D}$ itself.

3.1. Data augmentation needs to preserve noise statistics

As mentioned before, we are augmenting complex-valued, noisy images. This noise enters in the measurement process when we obtain the fully-sampled measurement of an image

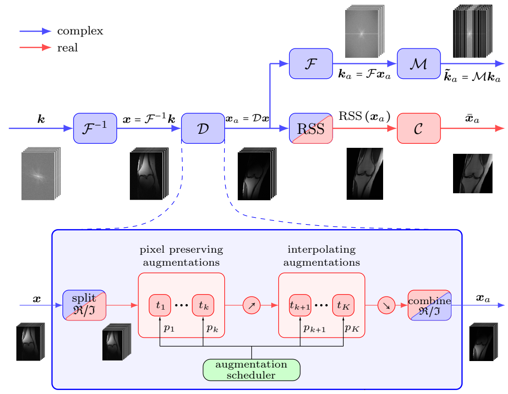


Figure 1. Flowchart of MRAugment, our data augmentation pipeline for MRI.

$\boldsymbol{x}^*$ as $\boldsymbol{k} = \mathcal{F}\boldsymbol{x}^* + \boldsymbol{z}$, and is well approximated by i.i.d complex Gaussian noise, independent in the real and imaginary parts of each pixel (Nishimura, 1996). Therefore, we can write $\boldsymbol{x} = \boldsymbol{x}^* + \boldsymbol{z}'$ where $\boldsymbol{z}'$ has the same distribution as $\boldsymbol{z}$ due to $\mathcal{F}$ being unitary. Since the noise distribution is characteristic to the instrumental setup (in this case the MR scanner and the acquisition protocol), assuming that the training and test images are produced by the same setup, it is important that the augmentation function preserves the noise distribution of training images as much as possible. Indeed, a large mismatch between training and test noise distribution leads to poor generalization (Knoll et al., 2019).

Let us demonstrate why it is non-trivial to generate augmented measurements for MRI through a simple example. A natural but perhaps naive approach for data augmentation is to augment the real-valued target image $\bar{\boldsymbol{x}}$ instead of the complex valued $\boldsymbol{x}$. This would allow us to directly obtain real augmented images from a real target image just as in typical data augmentation. However, this approach leads to different noise distribution in the measurements compared to the test data due to the non-linear mapping from individual coil images to the real-valued target and works poorly. Experiments demonstrating the weakness of this naive approach of data augmentation can be found in Section 4.5.

In contrast, if we augment the individual coil images $\boldsymbol{x}$ directly with a linear function $\mathcal{D}$, which is our main focus here, we obtain the augmented k-space data

$$\boldsymbol{k}_a = \mathcal{F}\mathcal{D}\boldsymbol{x} = \mathcal{F}\mathcal{D}(\boldsymbol{x}^* + \boldsymbol{z}') = \mathcal{F}\mathcal{D}\boldsymbol{x}^* + \mathcal{F}\mathcal{D}\boldsymbol{z}',$$

where $\mathcal{F}\mathcal{D}\boldsymbol{x}^*$ represents the augmented signal and the noise $\mathcal{F}\mathcal{D}\boldsymbol{z}'$ is still additive complex Gaussian. A key observation is that in case of transformations such as translation, horizontal and vertical flipping and rotation the noise distribution is exactly preserved. Moreover, for general linear transformations the noise is still Gaussian in the real and imaginary parts of each pixel.

To elaborate further, in the multi-coil case our augmentation pipeline applies transformations to the underlying object modulated by the different coil sensitivity maps. In particular, the fully sampled measurement of the ith coil in the image domain takes the form

$$x_i = S_i x^* + z_i', \tag{3.1}$$

where $z_i' = \mathcal{F}^{-1} z_i$ is i.i.d Gaussian noise obtained via a unitary transform of the original measurement noise. Assuming linear augmentations, the augmented coil image from MRAugment can be written as

$$x_{a,i} = \mathcal{D}(S_i x^* + z_i') = \mathcal{D} S_i x^* + \mathcal{D} z_i', \tag{3.2}$$

where the additive noise is still Gaussian. As seen in (3.2), MRAugment transforms images modulated by the coil sensitivities, therefore the sensitivitiy maps are also indirectly augmented. However, the models we experimented with had no issues learning the proper mapping from augmented measurements with transformed sensitivity maps as our experimental results show.

It is natural to ask if data augmentation would be possible by directly augmenting the object, before the coil sensitivities are applied. If the sensitivity maps are known or are estimated a priori, one may recover the object from the various coils as

$$x = \sum_{j=1}^{N} S_j^* x_j = \sum_{j=1}^{N} S_j^* (S_j x^* + z_j') = x^* + \sum_{j=1}^{N} S_j^* z_j',$$

where S_j^* is the complex conjugate of S_j and $\sum_{j=1}^{N} S_j^* S_j = I$ due to typical normalization (Sriram et al., 2020). Then, we can apply the augmentation as

$$x_a = \mathcal{D} x = \mathcal{D}(x^* + \sum_{i=1}^{N} S_j^* z_j') = \mathcal{D} x^* + \mathcal{D} \sum_{j=1}^{N} S_j^* z_j'.$$

Finally, we obtain the augmented coil images as

$$x_{a,i} = S_i x_a = S_i \mathcal{D} x^* + S_i \mathcal{D} \sum_{j=1}^{N} S_j^* z_j'. \tag{3.3}$$

Comparing (3.3) with (3.2), one may see that now the augmentation is directly applied to the ground truth signal bypassing the coil sensitivities. However, comparing this result in (3.3) with the original unaugmented coil images in (3.1) reveals that the additive noise in (3.3) has a very different distribution from the original i.i.d Gaussian, even worse noise on different augmented coil images are now correlated. Finally, the sensitivity maps are typically not known and need to be estimated before we can apply this augmentation technique, which can introduce additional inaccuracies in the augmentation pipeline.

This discussion motivates our choice to i) augment complex-valued images directly derived from the original k-space measurements, ii) consider simple transformations which preserve the noise distribution and iii) augment individual coil images as in (3.2). Next we overview the types of augmentations we propose in line with these key observations.

3.2. Transformations used for data augmentation

We apply the following two types of image transformations $\mathcal{D}$ in our data augmentation pipeline:
Pixel preserving augmentations, that do not require any form of interpolation and simply result in a permutation of pixels over the image. Such transformations are vertical and horizontal flipping, translation by integer number of pixels and rotation by multiples of $90°$. As we pointed out in Section 3.1, these transformations do not affect the noise distribution on the measurements and therefore are suitable for problems where training and test data are expected to have similar noise characteristics.
General affine augmentations, that can be represented by an affine transformation matrix and in general require resampling the transformed image at the output pixel locations. Augmentations in this group are: translation by arbitrary (not necessarily integer) coordinates, arbitrary rotations, scaling and shearing. Scaling can be applied along any of the two spatial dimensions. We differentiate between isotropic scaling, in which the same scaling factor s is applied in both directions ($s > 1$: *zoom-in*, $s < 1$: *zoom-out*) and anisotropic scaling in which different scaling factors (s_x, s_y) are applied along different axes.

Figure 2 provides a visual overview of the types of augmentations applied in this paper. Numerous other forms of transformations may be used in this framework such as exposure and contrast adjustment, image filtering (blur, sharpening) or image corruption (cutout, additive noise). However, in addition to the noise considerations mentioned before that have to be taken into account, some of these transformations are difficult to define for complex-valued images and may have subtle effects on image statistics. For instance, brightness adjustment could be applied to the magnitude image, the real part only or both real and imaginary parts, with drastically different effects on the magnitude-phase relationship of the image. That said, we hope to incorporate additional augmentations in our pipeline in the future after a thorough study of how they affect the noise distribution.

3.3. Scheduling and application of data augmentations

With the different components in place we are now ready to discuss the scheduling and application of the augmentations, as depicted in the bottom half of Figure 1. Recall that MRAugment generates a target image $\bar{\boldsymbol{x}}_a$ and corresponding undersampled k-space measurement $\tilde{\boldsymbol{k}}_a$ from

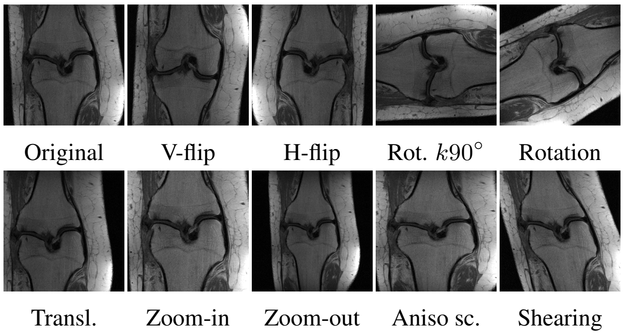


Figure 2. Transformations used in MRAugment applied to a ground truth slice.

a full k-space measurement. Which augmentation is applied and how frequently is determined by a parameter p, the common parameter determining the probability of applying a transformation to the ground truth image during training, and the weights $\mathcal{W} = (w_1, w_2, ..., w_K)$ pertaining to the K different augmentations, controlling the weights of transformations relative to each other. We apply a given transformation t_i with probability $p_i = p \cdot w_i$. The augmentation function is applied to the coil images, specifically the same transformation is applied with the same parameters to the real and imaginary parts $(\Re\{\boldsymbol{x}_1\}, \Im\{\boldsymbol{x}_1\}, \Re\{\boldsymbol{x}_2\}, \Im\{\boldsymbol{x}_2\}, ..., \Re\{\boldsymbol{x}_N\}, \Im\{\boldsymbol{x}_N\})$ of coil images. If a transformation t_i is sampled (recall that we select them with probabilities p_i), we randomly select the parameters of the transformation from a pre-defined range (for example, rotation angle in $[0, 180^\circ]$). To avoid aliasing artifacts, we first upsample the image before transformations that require interpolation. Then the result is downsampled to the original size.

A critical question is how to schedule p over training in order to obtain the best model. Intuitively, in initial stages of training no augmentation is needed, since the model can learn from the available original training examples. As training progresses the network learns to fit the original data points and their utility decreases over time. We find schedules starting from $p = 0$ and increasing over epochs to work best in practice. The ideal rate of increase depends on both the model size and amount of available training data.

4. Experiments

In this section we explore the effectiveness of MRAugment in the context of accelerated MRI reconstruction in various regimes of available training data sizes on various datasets. We start with providing a summary of our main findings, followed by a detailed description of the experiments. Additional reconstructions and more experimental details can be found in the supplementary material.

In the low-data regime (up to $\approx 4k$ images), data augmentation very significantly boosts reconstruction performance. The improvement is large both in terms of raw SSIM and visual reconstruction quality. Using MRAugment, fine details are recovered that are completely missing from reconstructions without DA. This suggests that DA improves the value of reconstructions for medical diagnosis, since health experts typically look for small features of the anatomy. This regime is especially important in practice, since large public datasets are extremely rare.

In the moderate-data regime ($\approx 4k - 15k$ images) MRAugment still achieves significant improvement in reconstruction SSIM. We want to emphasize the significance of seemingly small differences in SSIM close to the state of the art and invite the reader to visit the fastMRI Challenge Leaderboard that demonstrates how close the best performing models are.

In the high-data regime (more than $15k$ images) data augmentation has diminishing returns. It does not notably improve performance of the current state of the art, but it does not degrade performance either. Our experiments in the latter two regimes however strongly suggest that data augmentation combined with much larger models may lead to significant improvement over the state of the art, even in the high-data regime. However, without larger models it is expected that in a regime of abundant data, DA does not improve performance. For the models and problem considered here, this is around $15k$ images. We hope to investigate the effectiveness of MRAugment combined with such larger models in our future work.

Additional benefits of data augmentation include improved robustness under shifts in test distribution, such as improved generalization to new MRI scanners and field strengths. Furthermore, we observe that data augmentation can help to eliminate hallucinations by preventing overfitting to training data.

4.1. Experimental setup

We use the state-of-the-art End-to-End VarNet model (Sriram et al., 2020), which is as of now one of the best performing neural network models for MRI reconstruction. We measure performance in terms of the structural similarity index measure (SSIM), which is a standard evaluation metric for medical image reconstruction. We study the performance of MRAugment as a function of the size of the training set. We construct different subsampled training sets by randomly sampling volumes of the original training dataset and adding all slices of the sampled volumes to the new subsampled dataset. For all experiments, we apply random masks by undersampling whole k-space lines in the phase encoding direction by a factor of 8 and including 4% of lowest frequency adjacent k-space lines in order to be consistent with baselines in (Zbontar et al., 2018). For both the baseline experiments and for MRAugment, we generate a new random mask for each slice on-the-fly while training by uniformly sampling k-space lines, but use the same fixed mask for each slice within the same volume on the validation set (different across volumes). This technique is standard for models trained on the fastMRI dataset and not specific to our data augmentation pipeline. For augmentation probability scheduling we use

$$p(t) = \frac{p_{max}}{1 - e^{-c}}(1 - e^{-tc/T}),$$

where t is the current epoch, T denotes the total number of epochs, $c = 5$ and $p_{max} = 0.55$ unless specified otherwise. This schedule works resonably well on datasets of various size that we have studied and has not been fine-tuned to individual experiments. Ablation studies on the effect of the scheduling function is deferred to the supplementary.

4.2. Low-data regime

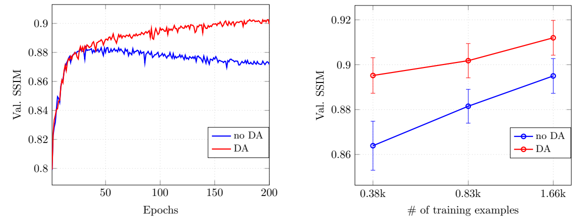
For the low-data regime we work with two different datasets, the Stanford 2D FSE dataset and the 3D FSE Knee dataset described below and demonstrate significant gains in reconstruction performance.

Stanford 2D FSE dataset. First, we perform experiments on the Stanford 2D FSE (Cheng) dataset, a public dataset of 89 fully-sampled MRI volumes of various anatomies including lower extremity, pelvis and more. We use $80\% - 20\%$ training-validation split, randomly sampled by volumes. We generate 5 random splits in order to minimize variations in reconstruction metrics due to validation set selection and report the mean validation SSIM over 5 runs along with the standard errors.

We plot a training curve of validation SSIM with and without data augmentation in Figure 3a. The regularizing effect of data augmentation prevents overfitting to the training set and improves reconstruction SSIM on the validation dataset even in case of training $4\times$ longer than in the baseline experiments without data augmentation. Figure 3b compares mean validation SSIM when the model is trained in different data regimes from 25% to 100% of all training data. MRAugment leads to significant improvement in reconstruction SSIM and this improvement is consistent across different train-val splits and training set sizes. We achieve higher mean SSIM using only 25% of the training data with MRAugment than training on the full dataset without DA. On the full dataset, we improve reconstruction SSIM from 0.8950 to 0.9120, and MRAugment achieves even larger gains in the lower data regime. Figure 4 provides a visual comparison of a reconstructed slice emphasizing the benefit of data augmentation.

Stanford Fullysampled 3D FSE Knees dataset. The Stanford Fullysampled 3D FSE Knees dataset (Sawyer et al., 2013) consists of 20 fully-sampled k-space volumes of knees. We use the same methodology to generate training and validation splits and evaluate results as in case of the Stanford 2D FSE dataset.

This dataset has significantly less variation compared to the Stanford 2D FSE dataset. Consequentially, we observe strong overfitting early in training if no data augmentation is used (Figure 5a). However, applying data augmentation successfully prevents overfitting. Furthermore, in accordance with observations on the Stanford 2D FSE dataset, data augmentation significantly boosts reconstruction SSIM across different data regimes (Figure 5b).



(a) Validation SSIM vs. epochs. MRAugment prevents the model from overfitting to training data.

(b) Validation SSIM vs. number of training images. Mean and standard error over 5 train/val splits is depicted.

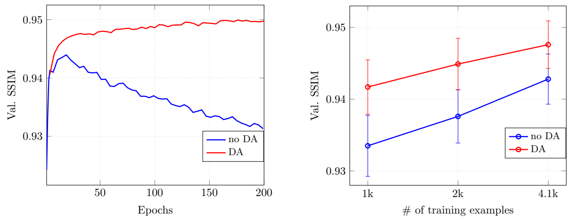
Figure 3. Experimental results on the Stanford 2D FSE dataset.

Figure 4. Visual comparison of reconstructions on the Stanford 2D FSE dataset with and without data augmentation.

4.3. High-data regime

Next, we perform an extensive study on the fastMRI dataset (Zbontar et al., 2018), the largest publicly available fully-sampled MRI dataset with competitive baseline models, that allows us to investigate the utility of MRAugment across a wide range of training data regimes. More specifically, we use the fastMRI knee dataset, for which the original training set consists of approximately $35k$ MRI slices in 973 volumes and we subsample to 1%, 10%, 33% and 100% of the original size. We measure performance on the original (fixed) validation set separate from the training set.

Single-coil experiments. For single-coil acquisition we are able to exactly match the performance of the model trained on the full dataset using only a third of the training data as depicted on the left in Fig. 7. Moreover, with only 10% of the training data we achieve comparable SSIM to the model



(a) Validation SSIM vs. training epochs. We observe strong overfitting without data augmentation.

(b) Validation SSIM vs. number of training images. Mean and standard error over 5 train/val splits is depicted.

Figure 5. Experimental results on the Stanford Fullysampled 3D FSE dataset.

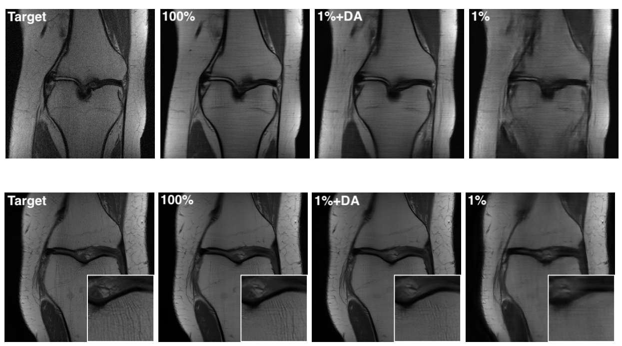


Figure 6. Visual comparison of single-coil (top row) and multi-coil (bottom-row) reconstructions using varying amounts of training data with and without data augmentation. We achieve reconstruction quality comparable to the state of the art but using 1% of the training data. Without DA fine details are completely lost.

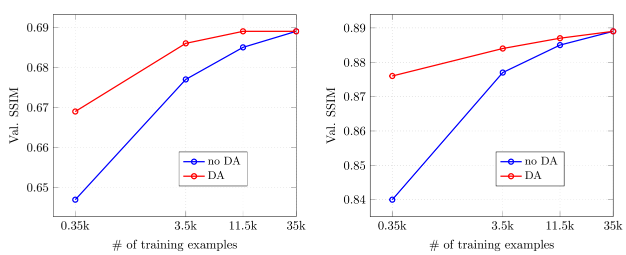


Figure 7. Single-coil (left) and multi-coil (right) validation SSIM vs. # of training images.

trained on the full dataset. The visual difference between reconstructions with and without data augmentation becomes striking in the low-data regime. As seen in the top row of Fig. 6, the model without DA was unable to reconstruct any of the fine details and the results appear blurry with strong artifacts. Applying MRAugment greatly improves reconstruction quality both in a quantitative and qualitative sense, visually approaching that obtained from training on the full dataset but using hundred times less data.

Multi-coil experiments. As depicted on the right in Fig. 7 for multi-coil acquisition we closely match the state of the art while significantly reducing training data. More specifically, we approach the state of the art SSIM within 0.6% using 10% of the training data and within 0.25% with 33% of training data. As seen in the bottom row of Fig. 6, when using only 1% of the training data we successfully reconstruct fine details comparable to that obtained from training on the full dataset, while high frequencies are completely lost without DA.

Finally, we perform ablation studies on the fastMRI dataset and demonstrate that both pixel preserving and interpolating transformations individually improve reconstruction SSIM. Furthermore their effect is complementary: the best results are obtained by adding all transformations to the pipeline. Moreover, we investigate the effect of the data augmentation scheduling function and demonstrate that exponential scheduling results in better performance compared to a constant augmentation probability. We also evaluate a range of different augmentation schedules and show that both significantly lower or higher probabilities lead to poorer reconstruction SSIM. All ablation experiments can be found in the supplementary material.

4.4. Model robustness

In this section we investigate further potential benefits of data augmentation in scenarios where training examples from the target data distribution are not only scarce as studied before, but unavailable. Distribution shifts can have a detrimental effect on a variety of reconstruction methods (Darestani et al., 2021).

To this end, we explore how data augmentation impacts generalization to new MRI scanner models not available in training time. Different MRI scanners may use different field strenghts for acquisition, and higher field strength typically correlates with higher SNR. Approximately half of the volumes in the fastMRI knee dataset have been acquired by a $1.5T$ scanner, whereas the rest by three different $3T$ scanners. We perform the following experiments:

- $3T \to 3T$: We train and validate on volumes acquired using $3T$ scanners. Volumes in the validation set have been imaged by a $3T$ scanner not in the training set.
- $3T \to 1.5T$: We train on all volumes acquired by $3T$ scanners and validate on the $1.5T$ scanner.
- $1.5T \to 3T$: We train on all volumes acquired by the $1.5T$ scanner and validate on all other $3T$ scanners.

Table 1 summarizes our results. Data augmentation consistently improves reconstruction SSIM on unseen scanner models. Similarly to our main experiments, the improvement is especially significant in the low-data regime. We observe that DA provides the greatest benefit when training on $1.5T$ scanners and testing on $3T$ models. We hypothesize that data augmentation can hinder the model from

2% train	no DA	DA
$3T \to 3T$	0.8646	**0.9049**
$3T \to 1.5T$	0.8241	**0.8551**
$1.5T \to 3T$	0.8174	**0.8913**

100% train	no DA	DA
$3T \to 3T$	0.9177	**0.9185**
$3T \to 1.5T$	0.8686	**0.8690**
$1.5T \to 3T$	0.9043	**0.9062**

Table 1. Scanner transfer results using 2% (top) and 100% (bottom) of training data.

overfitting to the higher noise level present on $1.5T$ acquisitions during training thus resulting in better generalization on the lower noise $3T$ volumes.

Experiments suggesting additional promising properties of models trained with MRAugment, such as improved generalization on unseen anatomies and robustness to hallucinations can be found in the supplementary.

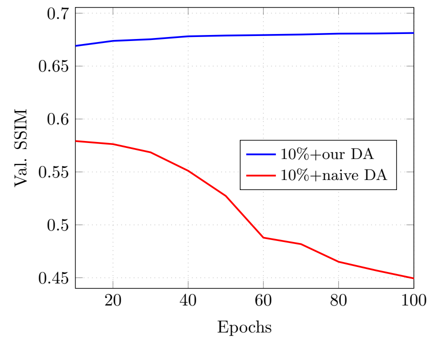
4.5. Naive data augmentation

We would like to emphasize the importance of applying DA in a way that takes into account the measurement noise distribution. When applied incorrectly, DA leads to significantly worse performance than not using any augmentation.

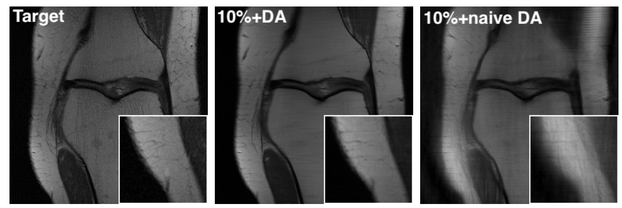
We train a model using 'naive' data augmentation without considering the measurement noise distribution as described in Section 3.1, by augmenting real-valued target images. We use the same exponential augmentation probability scheduling for MRAugment and the naive approach. As Fig. 8a demonstrates, reconstruction quality degrades over training using the naive approach. This is due to the fact that as augmentation probability increases, the network sees less and less original, unaugmented images, whereas the poorly augmented images are detrimental for generalization due to the mismatch in train and validation noise distribution. On the other hand, MRAugment clearly helps and validation performance steadily improves over epochs. Fig. 8b provides a visual comparison of reconstructions using naive DA and our data augmentation method tailored to the problem. Naive DA reconstruction exhibits high-frequency artifacts and low image quality caused by the mismatch in noise distribution. These drastic differences underline the vital importance of taking a careful, physics-informed approach to data augmentation for MR reconstruction.

5. Conclusion

In this paper, we develop a physics-based data augmentation pipeline for accelerated MR imaging. We find that MRAugment yields significant gains in the low-data regime which can be beneficial in applications where only little training data is available or where the training data changes quickly. We also demonstrate that models trained with data augmentation are more robust against overfitting and shifts in the test distribution. We believe that this work opens up many interesting venues for further research with respect to data augmentation for inverse problems in the low data regime.



(a) Naive data augmentation degrades generalization performance.



(b) Naive data augmentation that does not preserve measurement noise distribution leads to significantly degraded reconstruction quality.

Figure 8. Experimental results comparing MRAugment with naive data augmentation.

First, learning efficient data augmentation from the training data has been investigated in prior literature (Cubuk et al., 2019; Lemley et al., 2017; Tran et al., 2017) and would be a natural extension of our method. Second, finding the optimal augmentation strength throughout training is challenging, therefore an adaptive scheme that automatically schedules the augmentation probability would potentially further improve upon our results. Such a method is proposed in Karras et al. (2020) for discriminator augmentation in GAN training, where the augmentation strength is adjusted based on an discriminator overfitting heuristic with great success. Finally, combining our technique with a generative framework such as AmbientGAN (Bora et al., 2018) that generates high quality samples of a target distribution from noisy partial measurements, could be potentially used to synthesize fully-sampled MRI data from few noisy k-space measurements.

Acknowledgments

M. Soltanolkotabi is supported by the Packard Fellowship in Science and Engineering, a Sloan Research Fellowship in Mathematics, an NSF-CAREER under award #1846369, DARPA Learning with Less Labels (LwLL) and FastNICS programs, and NSF-CIF awards #1813877 and #2008443. R. Heckel is supported by the IAS at TUM, the DFG (German Research Foundation), and by the NSF under award IIS-1816986.

References

Bora, A., Price, E., and Dimakis, A. G. Ambientgan: Generative models from lossy measurements. In *International Conference on Learning Representations*, 2018.

Candes, E. J., Romberg, J. K., and Tao, T. Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, 59(8):1207–1223, 2006.

Cheng, J. Y. Stanford 2D FSE. http://mridata.org/list?project=Stanford 2D FSE.

Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., and Ronneberger, O. 3D U-Net: Learning dense volumetric segmentation from sparse annotation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 424–432. Springer, 2016.

Cubuk, E. D., Zoph, B., Mane, D., Vasudevan, V., and Le, Q. V. Autoaugment: Learning augmentation strategies from data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 113–123, 2019.

Darestani, M. Z. and Heckel, R. Can un-trained neural networks compete with trained neural networks at image reconstruction? *arXiv:2007.02471 [cs, eess, stat]*, 2020.

Darestani, M. Z., Chaudhari, A., and Heckel, R. Measuring robustness in deep learning based compressive sensing. In *International Conference on Machine Learning*, 2021.

Donoho, D. L. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, 2006.

Griswold, M. A., Jakob, P. M., Heidemann, R. M., Nittka, M., Jellus, V., Wang, J., Kiefer, B., and Haase, A. Generalized autocalibrating partially parallel acquisitions (GRAPPA). *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 47(6):1202–1210, 2002.

Han, Y. and Ye, J. C. Framing U-Net via deep convolutional framelets: Application to sparse-view CT. *IEEE Transactions on Medical Imaging*, 37(6):1418–1429, 2018.

Heckel, R. and Hand, P. Deep decoder: Concise image representations from untrained non-convolutional networks. In *International Conference on Learning Representations*, 2019.

Heckel, R. and Soltanolkotabi, M. Compressive sensing with un-trained neural networks: Gradient descent finds the smoothest approximation. In *International Conference on Machine Learning*, 2020.

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

Jin, K. H., Gupta, H., Yerly, J., Stuber, M., and Unser, M. Time-dependent deep image prior for dynamic MRI. *arXiv:1910.01684 [cs, eess]*, 2019.

Karras, T., Aittala, M., Hellsten, J., Laine, S., Lehtinen, J., and Aila, T. Training generative adversarial networks with limited data. *arXiv preprint arXiv:2006.06676*, 2020.

Knoll, F., Hammernik, K., Kobler, E., Pock, T., Recht, M. P., and Sodickson, D. K. Assessment of the generalization of learned image reconstruction and the potential for transfer learning. *Magnetic Resonance in Medicine*, 81(1):116–128, 2019.

Lee, D., Yoo, J., Tak, S., and Ye, J. C. Deep residual learning for accelerated MRI using magnitude and phase networks. *IEEE Transactions on Biomedical Engineering*, 65(9):1985–1995, 2018.

Lemley, J., Bazrafkan, S., and Corcoran, P. Smart augmentation learning an optimal data augmentation strategy. *Ieee Access*, 5:5858–5869, 2017.

Liu, F., Samsonov, A., Chen, L., Kijowski, R., and Feng, L. SANTIS: Sampling-augmented neural network with incoherent structure for MR image reconstruction. *Magnetic Resonance in Medicine*, 82(5):1890–1904, 2019.

Lustig, M., Donoho, D. L., Santos, J. M., and Pauly, J. M. Compressed sensing MRI. *IEEE Signal Processing Magazine*, 25(2):72–82, 2008.

Nishimura, D. G. *Principles of magnetic resonance imaging*. Stanford University, 1996.

Ongie, G., Jalal, A., Baraniuk, C. A. M. R. G., Dimakis, A. G., and Willett, R. Deep learning techniques for inverse problems in imaging. *IEEE Journal on Selected Areas in Information Theory*, 2020.

Pezzotti, N., de Weerdt, E., Yousefi, S., Elmahdy, M. S., van Gemert, J., Schülke, C., Doneva, M., Nielsen, T., Kastryulin, S., Lelieveldt, B. P., et al. Adaptive-CS-Net: fastMRI with adaptive intelligence. *arXiv preprint arXiv:1912.12259*, 2019.

Pruessmann, K. P., Weiger, M., Scheidegger, M. B., and Boesiger, P. SENSE: sensitivity encoding for fast MRI. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 42(5):952–962, 1999.

Putzky, P. and Welling, M. Invert to learn to invert. In *Advances in Neural Information Processing Systems*, pp. 446–456, 2019.

Ronneberger, O., Fischer, P., and Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 234–241. Springer, 2015.

Sawyer, A. M., Lustig, M., Alley, M., Uecker, P., Virtue, P., Lai, P., and Vasanawala, S. Creation of fully sampled MR data repository for compressed sensing of the knee. 2013.

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

Shorten, C. and Khoshgoftaar, T. M. A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1):1–48, 2019.

Sodickson, D. K. and Manning, W. J. Simultaneous acquisition of spatial harmonics (SMASH): fast imaging with radiofrequency coil arrays. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 38(4):591–603, 1997.

Sriram, A., Zbontar, J., Murrell, T., Defazio, A., Zitnick, C. L., Yakubova, N., Knoll, F., and Johnson, P. End-to-end variational networks for accelerated MRI reconstruction. *arXiv preprint arXiv:2004.06688*, 2020.

Tran, T., Pham, T., Carneiro, G., Palmer, L., and Reid, I. A bayesian data augmentation approach for learning deep models. *arXiv preprint arXiv:1710.10564*, 2017.

Ulyanov, D., Vedaldi, A., and Lempitsky, V. Deep Image Prior. In *Conference on Computer Vision and Pattern Recognition*, pp. 9446–9454, 2018.

Van Veen, D., Jalal, A., Soltanolkotabi, M., Price, E., Vishwanath, S., and Dimakis, A. G. Compressed sensing with deep image prior and learned regularization. *arXiv preprint arXiv:1806.06438*, 2018.

Vasanawala, S. S., Alley, M. T., Hargreaves, B. A., Barth, R. A., Pauly, J. M., and Lustig, M. Improved pediatric MR imaging with compressed sensing. *Radiology*, 256 (2):607–616, 2010.

Wang, P., Chen, E. Z., Chen, T., Patel, V. M., and Sun, S. Pyramid convolutional RNN for MRI reconstruction. *arXiv preprint arXiv:1912.00543*, 2019.

Yang, Y., Sun, J., Li, H., and Xu, Z. Deep ADMM-Net for compressive sensing MRI. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, pp. 10–18, 2016.

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

Zhang, J. and Ghanem, B. ISTA-Net: Interpretable optimization-inspired deep network for image compressive sensing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1828–1837, 2018.

Zhao, S., Liu, Z., Lin, J., Zhu, J.-Y., and Han, S. Differentiable Augmentation for Data-Efficient GAN Training. *arXiv:2006.10738*, 2020a.

Zhao, Z., Zhang, Z., Chen, T., Singh, S., and Zhang, H. Image Augmentations for GAN Training. *arXiv:2006.02595*, 2020b.

Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., and Liang, J. Unet++: A nested u-net architecture for medical image segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pp. 3–11. Springer, 2018.