Reducing Uncertainty in Undersampled MRI Reconstruction with Active Acquisition

Zizhao Zhang[1,2,*] Adriana Romero[2] Matthew J. Muckley[3] Pascal Vincent[2] Lin Yang[1] Michal Drozdzal[2]
[1] University of Florida   [2] Facebook AI Research   [3] NYU School of Medicine

Abstract

The goal of MRI reconstruction is to restore a high fidelity image from partially observed measurements. This partial view naturally induces reconstruction uncertainty that can only be reduced by acquiring additional measurements. In this paper, we present a novel method for MRI reconstruction that, at inference time, dynamically selects the measurements to take and iteratively refines the prediction in order to best reduce the reconstruction error and, thus, its uncertainty. We validate our method on a large scale knee MRI dataset, as well as on ImageNet. Results show that (1) our system successfully outperforms active acquisition baselines; (2) our uncertainty estimates correlate with error maps; and (3) our ResNet-based architecture surpasses standard pixel-to-pixel models in the task of MRI reconstruction. The proposed method not only shows high-quality reconstructions but also paves the road towards more applicable solutions for accelerating MRI.

1. Introduction

Magnetic Resonance Imaging (MRI) is a commonly used scanning technique, which provides detailed images of organs and tissues within the human body. The promises of MRI, when compared to computed tomography, are the superior soft tissue contrast and the lack of ionizing radiation [49]. However, its main drawback is the slow acquisition time; MRI examinations can take as long as an hour. The acquisition is performed sequentially in *k-space* – a 2D complex-valued space that can be linked to the 2D Fourier transform of the image – at speed controlled by hardware and physiological constraints [27, 36], causing uncomfortable examination experiences and high health care costs. Therefore, accelerating MRI is a critical medical imaging problem, with the potential of substantially improving both its accessibility and the patient experience.

Reducing the number of *k*-space measurements is a standard way of speeding up the examination time. However,
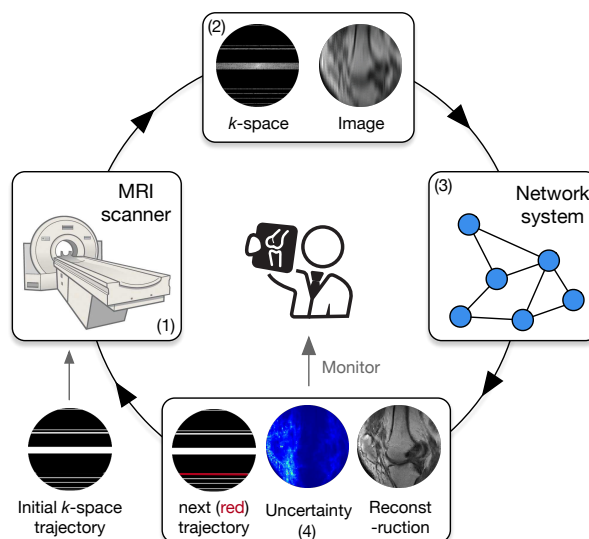
Figure 1: Overview of our proposed pipeline. A MRI scanner (1) acquires measurements given an initial trajectory. The *zero-filled* image reconstruction (2) is fed into our system (3), which outputs a reconstruction, an uncertainty map and the next suggested measurement (in red) to scan (4). These steps are repeated until the stopping criteria is met.

the images resulting from basic reconstructions from the undersampled *k*-space often exhibit blur or aliasing effects [27], making them unsuitable for clinical use. Hence, the goal of MRI reconstruction systems is to reduce the previously mentioned artifacts and recover high fidelity images.

Deep learning has recently shown great promise in MRI reconstruction with convolutional neural networks (CNNs) [13, 36, 49, 11]. Most of these methods are designed to work with a fixed set of measurements defining a sampling trajectory[1]. We argue that this sampling trajectory should be adapted on the fly, depending on the difficulty of the reconstruction. Figure 2 depicts box plots obtained by applying a reconstruction network to a large dataset for

---

[1]Throughout the paper, we use horizontal Cartesian acquisition trajectory, where *k*-space is acquired row-by-row and we use *measurement* to refer to a whole row of the Cartesian trajectory.

three acceleration factors, namely: $10\times$, $5\times$ and $4\times$. As shown in the figure, the $10\times$ plot exhibits the highest variance. As we introduce more measurements (by reducing the acceleration factor), the error variance decreases, highlighting the existing trade-off between acquisition speedup and reconstruction error variance when fixing the sampling trajectory. A natural way to overcome this trade-off is to define data driven sampling trajectories, via *active acquisition*[2] that adapt to reconstruction difficulty by selecting sequentially which parts of k-space to measure next.

Partial measurements naturally induce reconstruction *uncertainty*, as they might be consistent with multiple, equally plausible high fidelity reconstructions, which may or may not correspond to the reconstruction from fully observed k-space. In practice, these reconstructions could eventually mislead radiologists. Therefore, the ability to quantify and display the pixel-wise reconstruction uncertainty is of paramount relevance. On one hand, this pixel-wise uncertainty could allow radiologists to gain additional insight on the quality of the reconstruction and potentially yield a better diagnosis outcome. On the other hand, the *reduction in uncertainty* via *additional measurements* could be used as a signal to guide active acquisition.

In this paper, we propose a system for MRI reconstruction that, at inference time, actively acquires k-space measurements and iteratively refines the prediction with the goal of reducing the error and, thus, the final uncertainty (see Figure 1). To do so, we introduce a novel *evaluator network* to rate the quality gain in reconstruction of each k-space measurement. This evaluator is trained jointly with a *reconstruction network*, which outputs a high fidelity MRI reconstruction together with a pixel-wise uncertainty estimate. We explore a variety of architectural designs for the reconstruction network and present a residual-based model that exploits the underlying characteristics of MRI reconstruction. We extensively evaluate our method on a large scale knee MRI DICOM dataset and on ImageNet [4]. Our results show that (1) our evaluator consistently outperforms standard k-space active acquisition heuristics on both datasets; (2) our reconstruction network improves upon common pixel-wise prediction networks and; (3) the uncertainty predictions correlate with the reconstruction errors and, thus, can be used to trigger the halt signal to stop the active acquisition process.

To summarize, the contributions of the paper are the following:

- We introduce a reconstruction network design, which outputs both image reconstruction and uncertainty predictions, and is trained to jointly optimize for both.
- We introduce a novel evaluator network to perform active acquisition, which has the ability to recommend

---

[2]Note that, in active acquisition, the sampling trajectory would not only determine the number of measurements but also their sampling order.
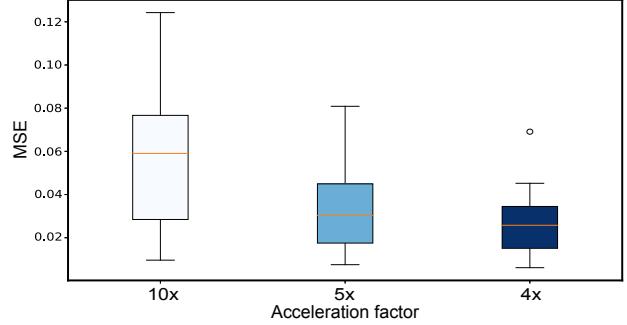

Figure 2: Box plots representing the variance of the reconstruction mean squared errors (MSE) for different acceleration factors. To obtain the plots, we apply random k-space trajectories with different acceleration factors to a set of images and feed them to a reconstruction network.

k-space trajectories for MRI scanners and reduce the uncertainty efficiently.
- We show through extensive evaluation the superior performance of the proposed approach, highlighting its practical value and paving the road towards improved practically applicable systems for accelerating MRI.

2. Related Work

MRI reconstruction. There is a vast literature tackling the problem of undersampled MRI reconstruction. State-of-the-art solutions include both signal processing techniques (e.g. Compressed Sensing (CS)) as well as machine learning ones. On one hand, CS-based MRI reconstruction has been widely studied in the literature [26, 28, 25, 31, 40]. These approaches usually result in over-smoothed reconstructions, which involve a time consuming optimization process, limiting their practical scalability. On other hand, deep learning based approaches have been introduced as a promising alternative to MRI reconstruction [42, 36, 24, 13, 35]. In [36], a cascaded CNN with a consistency layer is presented to ensure measurement fidelity in dynamic cardiac MRI reconstruction. In [13], a Unet architecture [35] is used to reconstruct brain images, while [24] proposes a recurrent inference machine for image reconstruction. Moreover, following recent trends, architectures involving image refinement mechanisms seem to be gaining increasing attention [36, 38, 24]. Although all previously-mentioned approaches are able to improve the reconstruction error, the human perception of the results is still not compelling. Therefore, recent works have also focused on exploring different training objectives such as adversarial losses [43, 8, 15] to enhance the perceptual reconstruction quality [38, 46].

Uncertainty. Significant effort has been devoted in the computer vision literature to provide uncertainty estimates [17] of predictions. There are two possible sources of uncertainty [20]: 1) model uncertainty due to an imperfect model

(epistemic uncertainty) and 2) data uncertainty due to imperfect measurements (aleatoric uncertainty). While model uncertainty can be decreased with better models, data uncertainty vanishes only with the observation of all variables with infinite precision. In medical imaging, uncertainty is often used to display probable errors [3] and has been mainly studied in the context of image segmentation [6, 22]. Segmentation errors (i.e. wrong label predictions) are often easier to detect by domain experts than reconstruction errors (i.e. shift of pixel values), which could potentially mislead diagnosis. Therefore, the study of uncertainty is crucial in the context of MRI reconstruction. In this paper, we focus on *data uncertainty*, which is caused by the partially observed *k*-space. This uncertainty can be captured by proper model parametrization, e.g. in regression tasks a Gaussian observation model is often assumed [17, 18]; this assumption can be relaxed to allow the use of arbitrary observation models as explained in [10].

Active acquisition. Previous research on optimizing *k*-space measurement trajectories from the MRI community include CS-based techniques [37, 33, 47, 9], SVD basis techniques [51, 30, 52], and region-of-interest techniques [44]. It is important to note that all these approaches work with fixed trajectories at inference time. By contrast, [23] proposed an on-the-fly eigenvalue based approach that adapts to encoding physics specific to the object. However, contrary to our approach, it requires solving an optimization problem at inference time. Moreover, since we train all the components of our pipeline jointly, our adaptive acquisition incorporates information on the image physics, the object being imaged, and the reconstruction process to select the next measurement.

3. Background and notation

Let $\mathbf{y} \in \mathbb{C}^{N \times N}$ be a complex-valued matrix representing the fully sampled *k*-space. Neglecting effects such as magnetic field inhomogeneity and spin relaxation, the image can be estimated from the *k*-space data by applying a 2D Inverse Fast Fourier Transform (IFFT) $\mathbf{x} = \mathcal{F}^{-1}(\mathbf{y})$, where $\mathbf{x} \in \mathbb{C}^{N \times N}$ is the image and $\mathcal{F}^{-1}$ is the IFFT operation. We denote the binary sampling mask defining the *k*-space Cartesian acquisition trajectory as $\mathbf{S}$ [49]. The acquired measurements are referred to as *observed* whereas the masked measurements are referred to as *unobserved*. We define the undersampled, partially observed *k*-space as $\hat{\mathbf{y}} = \mathbf{S} \odot \mathbf{y}$, where $\odot$ denotes element-wise multiplication. Thus, the basic *zero-filled* image reconstruction is obtained as $\hat{\mathbf{x}} = \mathcal{F}^{-1}(\hat{\mathbf{y}})$. Analogously, we can go from the reconstructed image to the *k*-space measurements $\hat{\mathbf{y}} = \mathcal{F}(\hat{\mathbf{x}})$, where $\mathcal{F}$ is the Fast Fourier Transform (FFT).

It is worth noting that MRI images $\mathbf{x} = \mathcal{F}^{-1}(\mathbf{y})$ are complex-valued matrices. However, most Picture Archiving and Communication Systems in hospitals do not store
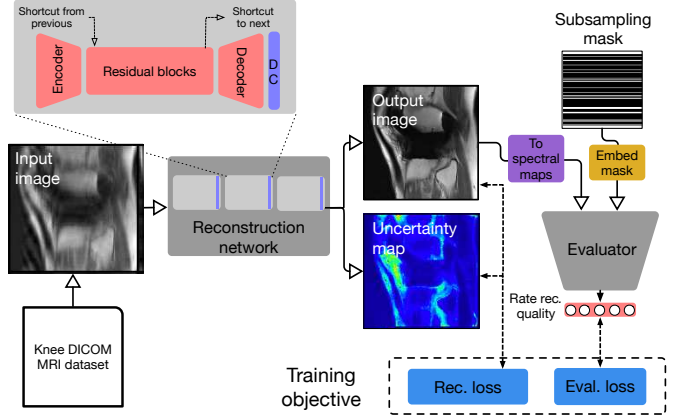

Figure 3: The training pipeline of the proposed method.

raw *k*-space measurements, but instead store the magnitude image $abs(\mathbf{x}) \in \mathbb{R}$ in the DICOM format. Therefore, we simulate *k*-space measurements by applying the FFT to the magnitude image $\mathbf{y} = \mathcal{F}(abs(\mathbf{x}))$. We do not differentiate the notation of an image in $\mathbb{R}$ or $\mathbb{C}$ hereinafter.

We make use of one of the numerous properties of FFT[3], namely Parseval's Theorem [34]. It implies that the l_2-distance between two images $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}$ is equivalent to the l_2-distance between their representation in the frequency domain, i.e. $||\mathcal{F}(\mathbf{x}^{(1)}) - \mathcal{F}(\mathbf{x}^{(2)})||_2^2 = ||\mathbf{x}^{(1)} - \mathbf{x}^{(2)}||_2^2$.

4. Method

Figure 3 illustrates our approach. The framework is composed of (1) a *reconstruction network* and (2) an *evaluator*. The goal of the reconstruction network is to produce high fidelity reconstructions from undersampled *k*-space measurements. The network takes a basic *zero-filled* image reconstruction as input and outputs an improved image reconstruction together with its uncertainty estimates. The goal of the evaluator network is to rate each corresponding *k*-space row of a reconstructed image, where the score should indicate how much it resembles true measurements. The rating score guides the measurement selection criterion: the lowest rated measurement should be acquired first.

4.1. Reconstruction network

Our reconstruction network has a cascaded backbone composed of residual networks (ResNets) [12], more precisely fully convolutional ResNets (FC-ResNets) [7, 2] followed by *data consistency* (DC) layers [36].

The DC layer [36][4] builds direct shortcut connections from the input of the network $\hat{\mathbf{x}}$ to its output $f(\hat{\mathbf{x}})$ to enforce the preservation of the observed information while estimating the reconstruction. The DC layer operates in *k*-space, and the reconstruction can be formally defined as:

---

[3]See Chapter 3.4 of [39] for the full list.
[4]We use the noiseless version of DC, which makes $\mathcal{F}(\hat{\mathbf{x}})$ fully preserved in the output, with a hard copy. See [36] for more details.
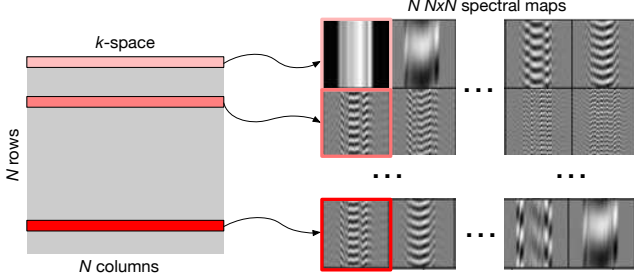
N NxN spectral maps

k-space

N rows

N columns

Figure 4: Image decomposition into N spectral maps.

$$\mathbf{r} = \mathrm{DC}(\hat{\mathbf{x}}, \mathbf{S}) = \mathcal{F}^{-1}((1-\mathbf{S})\odot\mathcal{F}(f(\hat{\mathbf{x}}))+\mathbf{S}\odot\mathcal{F}(\hat{\mathbf{x}})). \quad (1)$$

The rationale behind choosing FC-ResNet followed by DC layers as building block of our cascaded network is to learn the residual $f(\hat{\mathbf{x}}) = \mathbf{r} - \hat{\mathbf{x}}$. Thus, $f(\hat{\mathbf{x}})$ estimates the image representing the unobserved part of $\mathcal{F}(\mathbf{x})$, complementing $\mathcal{F}(\hat{\mathbf{x}})$. The rationale behind cascading the previously described building blocks is to provide intermediate deep supervision [21].

Overall, the proposed cascaded FC-ResNet (denoted c-ResNet) concatenates three identical tiny encoder-decoder networks, interleaved with DC layers. Note that this network is reminiscent of the 3D cascaded CNN proposed in [36] with minor design changes and endowed with deep supervision. To enhance the information flow between FC-ResNet modules, we add a shortcut to link residual blocks between adjacent modules (Figure 3). Hence, each module can re-use the representations of its predecessor and enhance the representations with further network capacity (see the supplementary materials for details).

4.2. Uncertainty estimates

FC-ResNet modules described in the previous section are trained to also output pixel-wise uncertainty estimates $u(\hat{\mathbf{x}})$, which we will use to trigger the halt signal to stop the active acquisition process. The additional benefit of having uncertainty estimates is that they highlight regions of the image that are likely to contain large reconstruction errors. Similarly to [10, 17], we model the uncertainty about the value of a pixel as a Gaussian centered at reconstruction mean $\mathbf{r}$ and with variance $u(\hat{\mathbf{x}})$, i.e. $\mathcal{N}(\mathbf{r}, \mathrm{diag}(u(\hat{\mathbf{x}})))$. We train our reconstruction network to maximize the average conditional log-likelihood, which amounts to minimizing:

$$\mathcal{L}_R(\hat{\mathbf{x}}, \mathbf{r}, \mathbf{x}) = \frac{1}{N^2}\sum_{i=1}^{N^2}\frac{|\mathbf{r}_i - \mathbf{x}_i|^2}{2u(\hat{\mathbf{x}})_i} + \frac{1}{2}\log(2\pi u(\hat{\mathbf{x}})_i), \quad (2)$$

where $\mathbf{x}$ is the "ground-truth" target image, $\hat{\mathbf{x}}$ is a *zero-filled* reconstruction given as input to the network, $\mathbf{r}$ is the reconstruction it outputs, and N^2 is the number of pixels.

4.3. Evaluator network

The role of the evaluator network e is to tell whether a given k-space row is likely to be a true k-space measure-

ment or to come from a reconstruction. When training the reconstruction network, we will be using the evaluator as additional regularization to encourage the reconstructed image to have phantasized k-space rows that look as if they came from the distribution of true measured rows. To be proficient in this task, the evaluator has to be able to capture small structural differences in images that define the distribution of the true, observed measurements. In our design, we leverage the idea of adversarial learning [8, 32], and train a discriminator-like evaluator to score the measurements and meanwhile encourage the reconstruction network to produce results that match the true measurement distribution.

The first step of the evaluator decomposes the output image reconstruction $\mathbf{r} \in \mathbb{C}^{N \times N}$ into N *spectral maps*, each one corresponding to a *single* k-space row. To obtain these spectral maps, we first transform $\mathbf{r}$ into the k-space representation $\mathbf{y} = \mathcal{F}(\mathbf{r})$. Then, we mask out all the k-space rows except of the i-th one using a binary mask $\hat{\mathbf{S}}^{(i)}$. The i-th spectral map of a reconstruction output is obtained as $\mathbf{M}(\mathbf{r})^{(i)} = \mathcal{F}^{-1}(\hat{\mathbf{S}}^{(i)} \odot \mathcal{F}(\mathbf{r}))$. Analogously $\mathbf{M}(\mathbf{x})^{(i)}$ denotes the i-th spectral map of the ground truth reconstruction [5]. This process is depicted in Figure 4. Moreover, it embeds the acquisition trajectory $\mathbf{S}$ into a 6D vector. Finally, both the spectral maps and the trajectory embedding as a 3D tensor are fed to a CNN, whose full architectural details are provided in the supplementary material.

We train the evaluator so that it assigns a high value to spectral maps that correspond to actually observed rows of the k-space and a low value to the unobserved ones. The simplest approach would be to train a discriminator to distinguish between observed and unobserved rows. However, we found that such strategy does not work well: the evaluator tends to output polarized scores (close to 0 or 1), making it hard to use to rank unobserved measurements. Instead, we decompose both the ground truth image $\mathbf{x}$ and the reconstruction output $\mathbf{r}$ into spectral maps and train the evaluator network $e(\mathbf{r}, \mathbf{S})$ to fit target scores given by the following kernel:

$$t(\mathbf{r}, \mathbf{x})_i = \exp(-\gamma||\mathbf{M}(\mathbf{r})^{(i)} - \mathbf{M}(\mathbf{x})^{(i)}||_2^2), \quad (3)$$

where γ is a scalar hyper-parameter. Specifically e is trained to minimize the following objective:

$$\mathcal{L}_E^E(\mathbf{r}, \mathbf{x}, \mathbf{S}) = \sum_i^N |e(\mathbf{r}, \mathbf{S})_i - t(\mathbf{r}, \mathbf{x})_i|^2, \quad (4)$$

where $e(\mathbf{r}, \mathbf{S})_i$ is the score of measurement i. Note that $t(\mathbf{r}, \mathbf{x})_i$ is close to 1 when $\mathbf{M}(\mathbf{r})_i$ is similar to $\mathbf{M}(\mathbf{x})_i$ and is close to 0 otherwise[6]. Note that the DC layer always ensures that $\mathbf{M}(\mathbf{r})_i$ is equal to $\mathbf{M}(\mathbf{x})_i$ for the observed rows of the k-space. Hence $t_i \equiv 1$ for observed measurements.

---

[5]Note that using the linearity of the Fourier transform, one could write: $\mathbf{r} = \sum_i^N \mathbf{M}(\mathbf{r})^{(i)}$.

[6]Thus, t_i can be viewed as an energy function [48] we expect to minimize by updating the parameters of the reconstruction network.

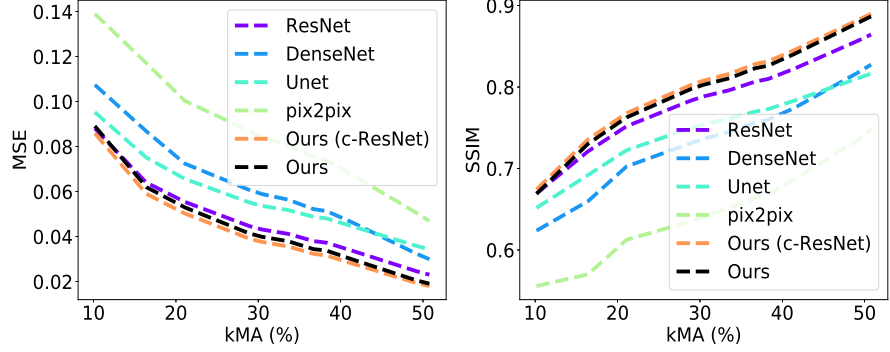| Method	MSE	SSIM
pix2pix	0.100	0.61
FC-DenseNet	0.072	0.70
Unet	0.065	0.72
ResNet	0.055	0.75
Ours (c-ResNet)	**0.050**	**0.77**
Ours	0.052	0.76

Table 1: MSE /SSIM at kMA = 21%.

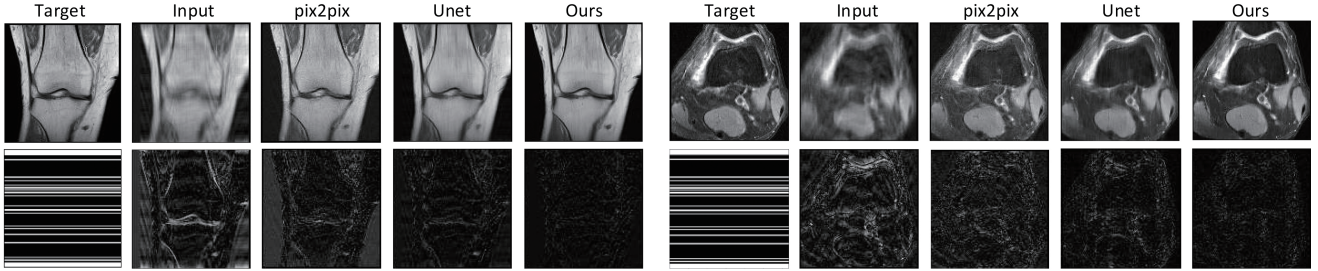Figure 5: Plots depicting MSE and SSIM for different kMA values.

Figure 6: Qualitative comparison of different reconstruction networks, including reconstruction results and error maps (normalized for improved visualization). The binary image below target is the sampling trajectory with 25% kMA.

4.4. Joint adversarial training

Following the principle of adversarial training, the evaluator network is used to update the reconstruction network using the following objective:

$$\mathcal{L}_E^R(\mathbf{r}, \mathbf{S}) = \sum_i^N |e(\mathbf{r}, \mathbf{S})_i - 1|^2, \quad (5)$$

which encourages the reconstruction network to produce reconstructions that can obtain high evaluator scores $e(\mathbf{r}, \mathbf{S})$. Overall, the reconstruction network is trained with the following objective:

$$\mathcal{L}(\mathbf{R}, \mathbf{x}, \mathbf{S}) = \frac{1}{K} \sum_{k=1}^K \mathcal{L}_R^k(\mathbf{r}^{k-1}, \mathbf{r}^k, \mathbf{x}) + \beta \mathcal{L}_E^R(\mathbf{r}^K, \mathbf{S}), \quad (6)$$

where $\mathbf{R} = [\mathbf{r}^0, ..., \mathbf{r}^K]$, $\mathbf{r}^0 = \hat{\mathbf{x}}$, $\mathbf{r}^k$ for $k \geq 1$ is the output of the k-th cascading block, β is a hyper-parameter controlling the influence of the evaluator loss on the global objective and K is the number of cascaded FC-ResNets in the reconstruction network.

We train the full model end-to-end, by alternating the reconstruction and evaluator networks' updates as in the standard adversarial training fashion [8]. We use the Adam solver ($\beta_1 = 0.5$, $\beta_2 = 0.999$) [19] with an initial learning rate of 0.0006 for 50 epochs. The learning rate is then linearly decreased per epoch for another 50 epochs, until it reaches 0. For all experiments, we set $\beta = 0.1$, $K = 3$ and $\gamma = 100$. All models are trained using 6 Tesla P100 GPUs, with a batch size of 48 per GPU.

4.5. Active acquisition

As illustrated in Figure 1, at *inference time*, the evaluator scores $e(\mathbf{r}, \mathbf{S})$ are used to select the next unobserved measurement to acquire. Then, the input image is updated accordingly and the process iterates until all measurements are acquired or a stopping criteria is met, e.g. a low global uncertainty score.

5. Experiments

In this section, we provide an in depth analysis of all the components of the proposed active acquisition pipeline. All experiments are conducted on a large scale Knee DICOM dataset from [45] as well as on ImageNet [5].

The Knee DICOM dataset is composed of 10k volumes. In our experiments, we use a subset of the data set and slice images from three axials at close-to-central positions of volumes, resulting in 11049 training images and 5048 test images. Among the training images, 10% are used for validation for hyperparameter search. We report results on the test set. All images are resized to have resolution 128×128. Volumes are from different machines and they have different intensity range. We standardize each image using mean and standard deviation computed on the corresponding volume.

In order to evaluate the quality of reconstruction on a downstream classification task, we use the ImageNet dataset [5]. We pre-process the dataset in order to have gray scale images of 128×128 pixels. Since we can not apply
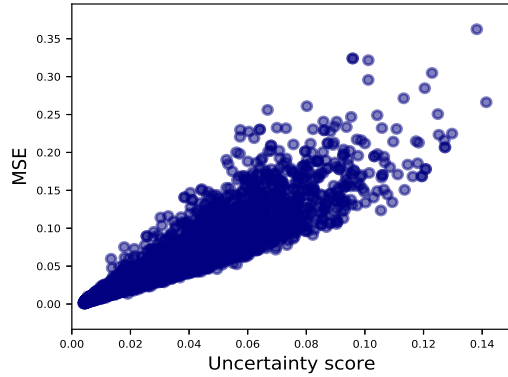
Figure 7: Correlation plot between MSE and the mean uncertainty score, each dot represents one image.

any off-the-shelf RGB pre-trained classification model, we train a ResNet50 [12] on the pre-processed images[7].

The training acquisition trajectory **S** is obtained following the Cartesian sampling by fixing 10 low frequency measurements in top and bottom rows and randomly sampling from the remaining ones until a desired number of measurements is obtained. In our experimental setup, the desired number of measurements is randomly chosen between 13 and 47. To evaluate the system, we characterize the acquisition trajectory **S** with the number of observed k-space measurements w.r.t. the total number of possible measurements as $\text{kMA} = \frac{\text{\# of acquired measurements}}{\text{\# of all possible measurements}}$ [8]. Since, acquisition time in MRI is proportional to the number of measurements acquired, the acceleration factor is computed as $\frac{1}{\text{kMA}}$. Thus, the lower kMA the higher acceleration factor (e.g. 25% kMA implies a speedup of 4x).

In the remainder of the section, we analyze the different components of our model, highlighting the obtained competitive results and its practical values.

5.1. Comparison of reconstruction architectures

In this subsection, we build two variants of our reconstruction architecture: (1) a vanilla c-ResNet trained by removing both the uncertainty estimates and the evaluator to minimize the mean squared error (MSE); and (2) a c-ResNet trained within the whole pipeline as described in Section 3. We compare these architectures to state-of-the-art deep learning models, commonly used in the MRI literature (Unet [13] and ResNet defined in CycleGAN [50]) and in dense prediction problems (FC-DenseNet103 [16], pix2pix [15, 43]). Note that pix2pix includes additional adversarial losses. We use MSE and Structural Similarity Index (SSIM) [41] as evaluation metrics.

---

[7]We use the following implementation: https://github.com/pytorch/examples/tree/master/imagenet

[8]For DICOM data, we define the number of all possible measurements to be $N/2$ - the true degrees of freedom of our data due to the Fourier Transform's conjugate symmetry property. See supplementary material for details.

For the sake of fair comparison, we add a DC layer to all models. Moreover, we found that that batch normalization (BN) [14] works poorly for undersampled MRI reconstruction, whereas instance normalization (IN) [1] is an important operation to improve results. Our findings are aligned with the recent work of [29], which suggests that IN learns features that are invariant to appearance changes, while BN better preserves content related information. Therefore, we endow all models with IN instead of BN and tune them to improve performance.

Table 1 reports MSE and SSIM performance for all above-mentioned models at kMA = 21% ($\sim$ 5x speedup). We observe that ResNet-based architectures outperform Unet and FC-DenseNet. As shown in the table, our vanilla reconstruction network (Ours (c-ResNet)) outperforms all above-mentioned pixel-wise baselines in terms of MSE and SSIM. Our full method (Ours) also optimizes uncertainty estimates and evaluator to perform active acquisition, which hinders the direct optimization of MSE and thereby results in a slight performance drop. Similarly, the weak performance of pix2pix could be explained by its discriminator.

Figure 5 depicts the MSE and SSIM performance metrics as a function of kMA. To validate the models, we create multiple validation sets by varying number of observed measurements from 10% to 50% kMA. All results were obtained with a single model trained on random acquisition trajectories with kMA varying from 10% to 37%. From these experiments, we observe the same trend as reported before, namely ResNet-based architectures being better suited to perform undersampled MRI reconstruction, for all kMA values. Moreover, we can observe that all the tested models scale gracefully to unseen kMAs, namely from 38% to 50%. Finally, we illustrate some qualitative results in Figure 6.

5.2. Uncertainty analysis

The goal of this subsection is to delve into the estimated uncertainty estimates and their correlation with the reconstruction errors. We select 512 test images, apply random random acquisition trajectories with kMA ranging from [10%, 95%], feed them to our reconstruction network and output both high fidelity reconstructions and uncertainty maps. Next, we compute the MSE between the obtained reconstructions and their corresponding ground truths as well as their mean uncertainty score. Figure 7 shows the resulting correlation plot. As it can be seen, the the mean uncertainty score correlates well with the MSE. We observe that the correlation is weaker as both MSE and uncertainty increase. These results indicate that the uncertainty estimates of our system could be useful to monitor the quality of reconstruction throughout our active acquisition process.
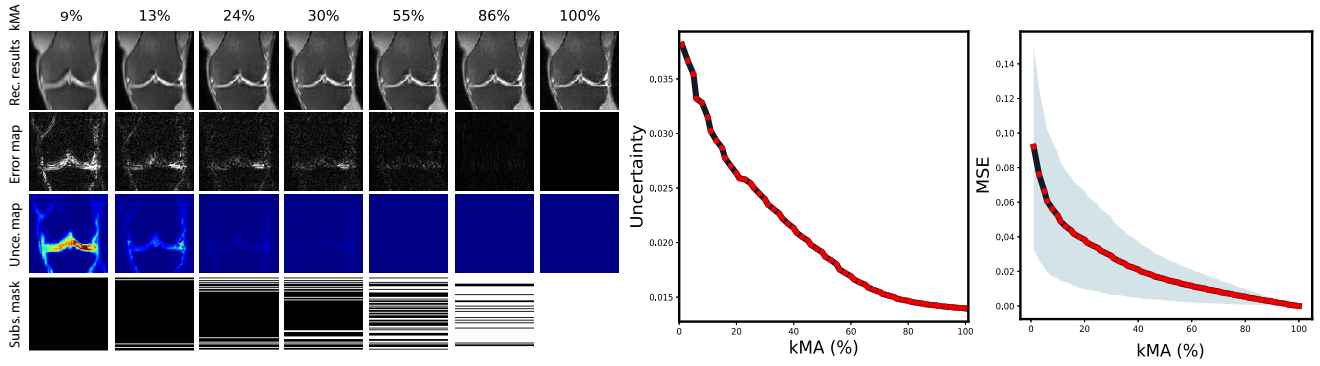
Figure 8: Simulation of *k*-space acquisition at the inference time. The left panel shows (top to bottom): reconstruction results, error maps, uncertainty maps, and sampling trajectories (in DFT coordinates). The initial mask includes 10 low-frequency rows (in white). The plots on the right monitors both MSE and the mean uncertainty value at different kMA ratios.
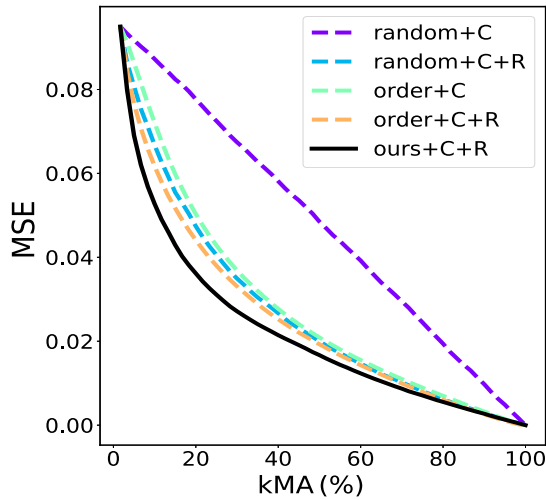

Figure 9: Comparison of different *k*-space acquisition heuristics to our model on the Knee dataset. The plot depicts MSE as a function of number of measurements.

5.3. *k*-space active acquisition analysis

Simulating the active acquisition process of an MRI scanner is straightforward. Given an input with a certain acquisition trajectory, we firstly obtain the reconstructed image. Then, we select the next unobserved row to acquire and measure it by copying it from the ground truth to the input image. After that, the updated input image is processed by our system. We iterate this process until the the stopping criteria is met or the *k*-space is fully observed.

We initialize the process with an input image resulting from the observation of 10 measurements (7.8% kMA), containing only low-frequency information. The active acquisition process is depicted in Figure 8, which contains qualitative intermediate results at different kMA values (including reconstructions, error maps, uncertainty estimates and acquisitions trajectories) as well as the progression of the mean uncertainty score and MSE on the test set. As shown in the figure, as we introduce additional measure-

ments, the reconstruction quality improves and the error and uncertainty decrease; reaching very low values around kMA = 30%. Note that the uncertainty is condensed in complex image regions, often containing high frequency information. Moreover, higher uncertainty regions appear to have higher reconstruction error values. Please refer to the supplementary video for more simulation results.

Comparison to standard active acquisition heuristics. We compare our evaluator-based approach to several baselines, including:

- Random+Copy(C): We randomly select an unobserved measurement, add it to the acquisition trajectory and compute the zero-filled reconstruction. We repeat this selection process without replacement until *k*-space is fully observed.
- Random+C+Reconstruction(R): Following Random+C selection strategy, we pass the zero-filled solution through our reconstruction network every time a measurement is added.
- Order+C: We select measurements following the low to high frequency order. Following the copy strategy, we add the measurement to the acquisition trajectory and compute the zero-filled reconstruction. We repeat this selection process until *k*-space is fully observed.
- Order+C+R: Following Order+C selection strategy, we pass the zero-filled solution through our reconstruction network every time a measurement is added.

Figure 9 analyzes the MSE as function of kMA. We observe that all methods have the same initial MSE and end up with zero MSE when all measurements are acquired. Random+C+R outperforms random+C notably, highlighting the benefit of applying the reconstruction network. However, order+C (even without any reconstruction) performs on par with random+C+R. This is not surprising, given that low frequency contains most of the information needed to reduce MSE. Finally, our method exhibits higher measurement efficiency when compared to the baselines.

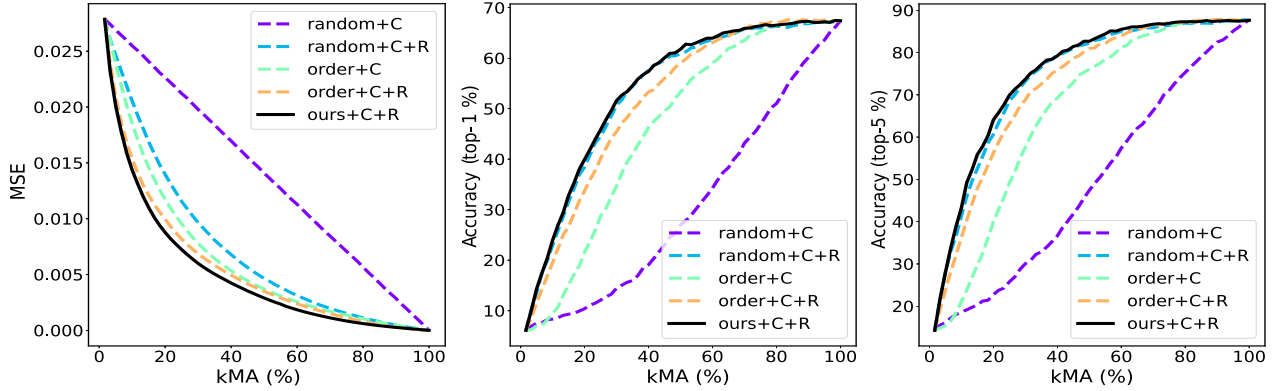**ImageNet simulation.** MSE is unable to reflect how well

Figure 10: Comparison of different *k*-space acquisition heuristics to with our pipeline on ImageNet. The plots depict MSE and accuracy (top-1&5) as a function of number of measurements.
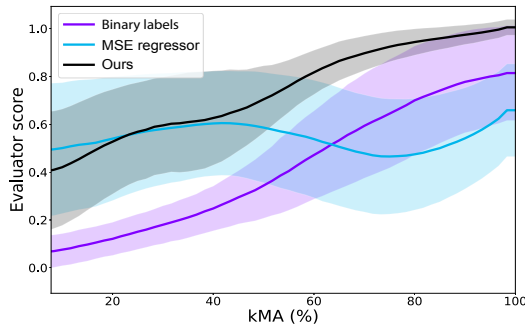

Figure 11: Evaluator score as a function of the number of measurements. We compare our evaluator design to two baselines: MSE regressor and adversarial loss trained with binary labels.

the semantic details, which may be required for diagnosis, are recovered. Since we don't have access to classification information on our knee dataset, we reach out to an auxiliary classification dataset and test our pipeline. We evaluate the method by means of MSE and top-*k* classification accuracy. Results are presented in Figure 10. The MSE results for different acquisition heuristics follow the same pattern as in the knee dataset. Interestingly, when it comes to the classification accuracy, random+C+R outperform other baselines (which were better in terms of MSE, e. g. order+C+R), achieving results comparable to our method. This experiment suggests that semantic information could exist in arbitrary high-frequency parts of images. Our method demonstrates excellent effectiveness at recovering both image quality and semantic details.

Evaluator ablation study. Finally, we compare our evaluator training strategy, described in subsection 4.3, with two alternatives. First, we train our evaluator network with binary labels (following adversarial training of image-to-image translation networks [15]), i.e. 0 for spectral maps corresponding to unobserved measurements (fake), and 1 for spectral maps corresponding to observed measurements (real). Second, we adapt the recently proposed [6] to score our spectral maps in terms of MSE. This approach trains a

regression network on top of pre-trained regression model. Note that this is different to adversarial training, since the regression network does not affect the weights of the reconstruction network. The results of the comparison are shown in Figure 11, where the scores of different evaluators are depicted as a function of kMA. Note that only the scores of spectral maps corresponding to unobserved measurements are considered. A good evaluator should produce increasing scores (up to a maximum value of 1) as the number of acquired measurements increases. Similarly, the evaluator score variance should decrease with the number of acquired measurements. As it can been observed, our method is the only one satisfying both requirements, highlighting the benefits of our evaluator design.

6. Conclusions

In this paper, we presented a novel active acquisition pipeline for undersampled MRI reconstruction, which can iteratively suggests *k*-space trajectories to best reduce uncertainty. We extensively validated our approach on a large scale knee dataset as well as on ImageNet, showing that (1) our evaluator design consistently outperforms alternative active acquisition heuristics; (2) our uncertainty estimates correlate with the reconstruction error and thus, can be used to trigger the halt signal of active acquisition at inference time; (3) our reconstruction architecture surpasses previously introduced architectures. Finally, we argued that the proposed method paves the road towards more applicable solutions for accelerating MRI, which ensure the optimal acquisition speedup while maintaining high fidelity image reconstructions with low uncertainty.

References

[1] J. L. Ba, J. R. Kiros, and G. E. Hinton. Layer normalization. *arXiv preprint arXiv:1607.06450*, 2016.

[2] A. Casanova, G. Cucurull, M. Drozdzal, A. Romero, and Y. Bengio. On the iterative refinement of densely connected representation levels for semantic segmentation. In *CVPR Workshop*, 2018.

[3] T. Ching, D. S. Himmelstein, B. K. Beaulieu-Jones, A. A. Kalinin, B. T. Do, G. P. Way, E. Ferrero, P.-M. Agapow, M. Zietz, M. M. Hoffman, et al. Opportunities and obstacles for deep learning in biology and medicine. *Journal of The Royal Society Interface*, 15(141):20170387, 2018.

[4] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR*, 2009.

[5] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. Imagenet: A large-scale hierarchical image database. In *CVPR*, pages 248–255, 2009.

[6] T. DeVries and G. W. Taylor. Leveraging uncertainty estimates for predicting segmentation quality. *arXiv preprint arXiv:1807.00502*, 2018.

[7] M. Drozdzal, E. Vorontsov, G. Chartrand, S. Kadoury, and C. Pal. The importance of skip connections in biomedical image segmentation. *MICCAI Workshop*, 2016.

[8] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In *NIPS*, pages 2672–2680, 2014.

[9] B. Gözcü, R. K. Mahabadi, Y.-H. Li, E. Ilıcak, T. Cukur, J. Scarlett, and V. Cevher. Learning-based compressive mri. *IEEE transactions on medical imaging*, 37(6):1394–1406, 2018.

[10] P. Gurevich and H. Stuke. Learning uncertainty in regression tasks by deep neural networks. *arXiv preprint arXiv:1707.07287*, 2017.

[11] Y. Han and J. C. Ye. k-space deep learning for accelerated MRI. *CoRR*, abs/1805.03779, 2018.

[12] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *CVPR*, 2016.

[13] C. M. Hyun, H. P. Kim, S. M. Lee, S. Lee, and J. K. Seo. Deep learning for undersampled mri reconstruction. *Physics in medicine and biology*, 2018.

[14] S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *ICML*, 2015.

[15] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. *CVPR*, 2017.

[16] S. Jégou, M. Drozdzal, D. Vazquez, A. Romero, and Y. Bengio. The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation. In *CVPR Workshop*, pages 1175–1183, 2017.

[17] A. Kendall and Y. Gal. What uncertainties do we need in bayesian deep learning for computer vision? In *NIPS*, pages 5574–5584, 2017.

[18] A. Kendall, Y. Gal, and R. Cipolla. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In *CVPR*, 2018.

[19] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. In *ICLR*, 2014.

[20] A. D. Kiureghian and O. Ditlevsen. Aleatory or epistemic? does it matter? *Workshop on Risk Acceptance and Risk Communication*, 2007.

[21] C.-Y. Lee, S. Xie, P. W. Gallagher, Z. Zhang, and Z. Tu. Deeply-supervised nets. In *AISTATS*, volume 38, 2015.

[22] C. Leibig, V. Allken, M. S. Ayhan, P. Berens, and S. Wahl. Leveraging uncertainty information from deep neural networks for disease detection. *Scientific reports*, 7(1):17816, 2017.

[23] E. Levine and B. Hargreaves. On-the-fly adaptive k-space sampling for linear mri reconstruction using moment-based spectral analysis. *IEEE transactions on medical imaging*, 37(2):557–567, 2018.

[24] K. Lønning, P. Putzky, M. W. Caan, and M. Welling. Recurrent inference machines for accelerated mri reconstruction. In *Medical Imaging with Deep Learning*, 2018.

[25] M. Lustig, D. Donoho, and J. M. Pauly. Sparse mri: The application of compressed sensing for rapid mr imaging. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 58(6):1182–1195, 2007.

[26] M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly. Compressed sensing mri. *IEEE signal processing magazine*, 25(2):72–82, 2008.

[27] D. Moratal, A. Vallés-Luch, L. Martí-Bonmatí, and M. E. Brummer. k-space tutorial: an mri educational tool for a better understanding of k-space. *Biomedical imaging and intervention journal*, 4(1), 2008.

[28] R. Otazo, D. Kim, L. Axel, and D. K. Sodickson. Combination of compressed sensing and parallel imaging for highly accelerated first-pass cardiac perfusion mri. *Magnetic resonance in medicine*, 64(3):767–776, 2010.

[29] X. Pan, P. Luo, J. Shi, and X. Tang. Two at once: Enhancing learning and generalization capacities via ibn-net. In *ECCV*, 2018.

[30] L. P. Panych, C. Oesterle, G. P. Zientara, and J. Hennig. Implementation of a fast gradient-echo svd encoding technique for dynamic imaging. *Magnetic resonance in medicine*, 35(4):554–562, 1996.

[31] T. M. Quan, T. Nguyen-Duc, and W.-K. Jeong. Compressed sensing mri reconstruction using a generative adversarial network with a cyclic loss. *IEEE transactions on medical imaging*, 37(6):1488–1497, 2018.

[32] A. Radford, L. Metz, and S. Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. In *ICLR*, 2016.

[33] S. Ravishankar and Y. Bresler. Adaptive sampling design for compressed sensing mri. In *Engineering in Medicine and Biology Society (EMBC)*, pages 3751–3755, 2011.

[34] O. Rippel, J. Snoek, and R. P. Adams. Spectral representations for convolutional neural networks. In *NIPS*, pages 2449–2457, 2015.

[35] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *MICCAI*, pages 234–241, 2015.

[36] J. Schlemper, J. Caballero, J. V. Hajnal, A. N. Price, and D. Rueckert. A deep cascade of convolutional neural networks for dynamic mr image reconstruction. *IEEE transactions on Medical Imaging*, 37(2):491–503, 2018.

[37] M. Seeger, H. Nickisch, R. Pohmann, and B. Schölkopf. Optimization of k-space trajectories for compressed sensing by bayesian experimental design. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 63(1):116–126, 2010.

[38] M. Seitzer, G. Yang, J. Schlemper, O. Oktay, T. Würfl, V. Christlein, T. Wong, R. Mohiaddin, D. Firmin, J. Keegan, et al. Adversarial and perceptual refinement for compressed sensing mri reconstruction. In *MICCAI*, 2018.

[39] R. Szeliski. Computer vision algorithms and applications, 2011.

[40] M. Tygert, R. Ward, and J. Zbontar. Compressed sensing with a jackknife and a bootstrap. *arXiv preprint arXiv:1809.06959*, 2018.

[41] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.

[42] L. Xu, J. S. Ren, C. Liu, and J. Jia. Deep convolutional neural network for image deconvolution. In *NIPS*, pages 1790–1798, 2014.

[43] G. Yang, S. Yu, H. Dong, G. Slabaugh, P. L. Dragotti, X. Ye, F. Liu, S. Arridge, J. Keegan, Y. Guo, et al. Dagan: Deep dealiasing generative adversarial networks for fast compressed sensing mri reconstruction. *IEEE transactions on medical imaging*, 37(6):1310–1321, 2018.

[44] S.-S. Yoo, C. R. Guttmann, L. Zhao, and L. P. Panych. Real-time adaptive functional mri. *Neuroimage*, 10(5):596–606, 1999.

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

[46] P. Zhang, F. Wang, W. Xu, and Y. Li. Multi-channel generative adversarial network for parallel magnetic resonance image reconstruction in k-space. In *MICCAI*, pages 180–188, 2018.

[47] Y. Zhang, B. S. Peterson, G. Ji, and Z. Dong. Energy preserved sampling for compressed sensing mri. *Computational and mathematical methods in medicine*, 2014, 2014.

[48] J. Zhao, M. Mathieu, and Y. LeCun. Energy-based generative adversarial network. In *ICLR*, 2017.

[49] B. Zhu, J. Z. Liu, S. F. Cauley, B. R. Rosen, and M. S. Rosen. Image reconstruction by domain-transform manifold learning. *Nature*, 555(7697):487, 2018.

[50] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *ICCV*, 2017.

[51] G. P. Zientara, L. P. Panych, and F. A. Jolesz. Dynamically adaptive mri with encoding by singular value decomposition. *Magnetic Resonance in Medicine*, 32(2):268–274, 1994.

[52] G. P. Zientara, L. P. Panych, and F. A. Jolesz. Applicability and efficiency of near-optimal spatial encoding for dynamically adaptive mri. *Magnetic resonance in medicine*, 39(2):204–213, 1998.