MRI Reconstruction with Interpretable Pixel-Wise Operations Using Reinforcement Learning

Wentian Li,* **Xidong Feng,*** **Haotian An, Xiang Yao Ng, Yu-Jin Zhang**
Tsinghua University
Beijing, China
li-wt17@mails.tsinghua.edu.cn

Abstract

Compressed sensing magnetic resonance imaging (CS-MRI) is a technique aimed at accelerating the data acquisition of MRI. While down-sampling in k-space proportionally reduces the data acquisition time, it results in images corrupted by aliasing artifacts and blur. To reconstruct images from the down-sampled k-space, recent deep-learning based methods have shown better performance compared with classical optimization-based CS-MRI methods. However, they usually use deep neural networks as a black-box, which directly maps the corrupted images to the target images from fully-sampled k-space data. This lack of transparency may impede practical usage of such methods. In this work, we propose a deep reinforcement learning based method to reconstruct the corrupted images with meaningful pixel-wise operations (e.g. edge enhancing filters), so that the reconstruction process is transparent to users. Specifically, MRI reconstruction is formulated as Markov Decision Process with discrete actions and continuous action parameters. We conduct experiments on MICCAI dataset of brain tissues and fastMRI dataset of knee images. Our proposed method performs favorably against previous approaches. Our trained model learns to select pixel-wise operations that correspond to the anatomical structures in the MR images. This makes the reconstruction process more interpretable, which would be helpful for further medical analysis.

Introduction

Magnetic Resonance Imaging (MRI) is a non-invasive medical imaging technique that uses magnetic fields and radio waves to generate images of the organs in the human body. Compared with other forms of medical imaging, MRI typically has better resolution related to soft tissues. However, the data acquisition process of MRI is very time-consuming. The data is sampled sequentially in k-space (frequency domain) and the speed is limited by physical and physiological constraints (Lustig, Donoho, and Pauly 2007), taking about 15 minutes up to over an hour, while the patients are required to remain still. The long scan time severely limits the applicability of MRI.

*These authors contributed equally to this work

Copyright © 2020, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

To accelerate the data acquisition, compressed-sensing MRI (CS-MRI) (Lustig et al. 2008) is a practical solution, as no additional hardware is required. CS-MRI reduces traversals in the k-space. In order to apply the Compressed Sensing technique, several conditions need to be satisfied. The image should have a sparse representation in a known transform domain, and the aliasing artifacts due to k-space down-sampling should be incoherent in that transform domain. Although the conditions are not always satisfied for MRI, CS-MRI still receives much attention in the research field and exhibits fairly promising results. There are different sampling schemes and patterns in k-space, such as grid spiral, radial, etc. Cartesian down-sampling scheme (Lustig, Donoho, and Pauly 2007), which follows Gaussian distribution and retains the low frequencies, was carefully designed to mimic the property of pure randomness to make the aliasing as incoherent as possible after zero-filling the unsampled positions in the k-space. The resulted artifacts in image domain, while not totally incoherent, can be perceived as a mixture of aliasing, blur and noise.

After obtaining partially sampled k-space data, MRI reconstruction can be performed using software algorithms. Optimization-based algorithms usually enforces both sparsity in the transform domain, which is required by compressed sensing, and data consistency in frequency domain, which means that the k-space values of the reconstructed image at the sampled positions do not deviate from the originally acquired k-space data.

Recent advances of deep learning has greatly improved the reconstruction precision, measured by NMSE, PSNR, and SSIM. Most deep-learning based methods (Sun et al. 2016; Schlemper et al. 2017; Yang et al. 2017; Hammernik et al. 2018; Hyun et al. 2018; Zhu et al. 2018) directly take the image from the zero-filled k-space or the original k-space data as the input of the deep neural networks and output the reconstructed image. These models are often trained end-to-end and function as black boxes.

Interpreterbility is a challenge for deep-learning based medical image processing (Razzak, Naz, and Zaib 2018). Since the primary purpose of MRI reconstruction is to facilitate diagnosis by removing aliasing artifacts and recovering details, it may pose risk if the transformation process of the

image are hidden by the "black-box". For example, considering that drastic differences of pixel values in the anatomical structures could become very small that are hard to perceive by human eyes when severely blurred and flattened by down-sampling, it would be beneficial if users can tell whether such structures are faithfully recovered.

In this work, we formulate MRI reconstruction as image processing in a deep reinforcement learning setting, with discrete actions and continuous action parameters. The corrupted images are processed using fine-grained and understandable pixel-wise operations. Hence, the transformation from the corrupted images to reconstructed ones is transparent to users. Building on the work of pixelRL (Furuta, Inoue, and Yamasaki 2019), we assign each pixel of the input image an agent that changes the pixel value. With carefully designed action space and reward function, the agents learn to take actions (choose operations) to work on the pixel value in a proper sequence, and some of the actions have learnable parameters. Discrete action policy and continuous action parameters are learned alternately. In a full episode, the change of pixels are executed iteratively and the output of the final step is the reconstructed MR image. In addition, we also make use of the data consistency in frequency domain to further boost the performance.

Compared with previous works that used deep neural network as a black-box, our method is transparent in that every change made on the pixels is known. Importantly, we show that the actions taken are distributed according to the anatomical structures of the tissues, thus interpretable to certain extent.

Our contributions can be summarized as follows:

- We adapt reinforcement learning techniques and classical image filters into MRI reconstruction. Our method compares favorably against previous approaches.
- Through meaningful pixel-wise operations, we make the reconstruction process transparent and more interpretable to users than previous deep-learning based methods.

Related work

MRI reconstruction

Under CS-MRI, only parts of the data are sampled in k-space. Suppose the fully sampled k-space $y_{target} \in \mathbb{C}^{H \times W}$ is a complex-valued matrix with spatial size $H \times W$, then its corresponding image[1] by applying 2D Inverse Fast Fourier Transform is $x_{target} = \mathcal{F}^{-1}(y_{target})$. Cartesian sampling scheme in k-space is equivalent to applying a binary mask S of the same size to y_{target}, and the obtained down-sampled k-space data is $y = S \odot y_{target}$, where $\odot$ denotes element-wise matrix multiplication. Since the Nyquist criterion is violated, the corresponding image $x = \mathcal{F}^{-1}(y)$ is corrupted by aliasing artifacts. With under-sampled high frequencies, the image is also blurry.

In order to recover the image x_{target} from x and y, classical methods exploit sparsity constraints, in various transform domains (e.g. total variation (Block, Uecker, and Frahm 2007; Lustig, Donoho, and Pauly 2007; Yang, Zhang, and Yin 2010), discrete wavelet transform (Qu et al. 2012)), with geometric information (Qu et al. 2014; Ning et al. 2013), and with dictionary learning (Ravishankar and Bresler 2010; Caballero et al. 2014). Due to the abundance of works of classical methods, please refer to (Jaspan, Fleysher, and Lipton 2015) for a comprehensive review.

In recent years, deep learning has been effectively used for MRI reconstruction. While some of these works incorporated classical formulations (Sun et al. 2016; Yang et al. 2017; Hammernik et al. 2018), most of these methods (Wang et al. 2016; Schlemper et al. 2017; Yang et al. 2017; Sun et al. 2018; Lee et al. 2018; Han et al. 2018; Fan et al. 2018; Hyun et al. 2018; Zhang et al. 2019) can be roughly generalized as learning a mapping function from x to x_{target} with a deep neural network, with a few exceptions (Zhu et al. 2018; Han, Sunwoo, and Ye 2019) involving the mapping of y and y_{target}. A common choice for loss function is the pixel-wise disparity of x and x_{target}, usually represented by mean square error of the two images (Schlemper et al. 2017; Hyun et al. 2018). Other loss functions include GAN loss , perceptual loss (Yang et al. 2017), and frequency domain loss (Schlemper et al. 2017).

Similar to classical approaches, deep-learning based approaches also enforces data consistency in frequency domain to ensure that the values at the originally sampled positions in k-space are not distorted by reconstruction. Data consistency can be integrated as a differentiable layer (Schlemper et al. 2017; Sun et al. 2018) into the deep neural network. In this work, we follow (Zhang et al. 2019), where the sampled entries in k-space for the output of deep neural network are manually adjusted with respect to the original measurement.

Image processing by reinforcement learning

Several works adopt reinforcement learning (RL) for image processing tasks, such as denoising, color enhancement and exposure adjustment. Compared with conventional CNN-based approaches that learn image-to-image mapping, one advantage of the RL-based methods is that the operations performed on the image are transparent and are often understandable, thus providing a clearer view of the process. At the same time, owing to the exploratory nature of RL, the RL-based methods have demonstrated better generalization property (Park et al. 2018; Yu et al. 2018a).

In RL setting, Markov Decision Process (MDP) is used to depict sequential decision making problems and can be defined as the tuple $< s, a, r, P, \lambda >$. The agent observes the state s, takes action a based on policy π, transfers to another state according to the transfer probability P. For image processing task, the formulated Markov process is usually deterministic. The agent receives reward r, and λ is the discount factor for cumulative reward. The goal of RL is to maximize the expected cumulative reward. To formulate the image processing task under the framework of RL, one needs to define a corresponding MDP for that.

A common choice for state is the image itself or extracted features related to the task. In order to make the operations

[1]While the inverse Fourier transform of k-space data is technically a complex-valued matrix, we take its modulus when referring to the image. We do not differentiate their notations hereafter.

suitable for the image processing tasks, the design of the actions and the action space is critical. The action space can be discrete action space (Sahba and Tizhoosh 2003; Park et al. 2018; Yu et al. 2018a; Furuta, Inoue, and Yamasaki 2019), continuous action space (Yu et al. 2018b) (individual type of operation), or discrete action space with continuous parameters (Hu et al. 2018).

For example, (Hu et al. 2018) adopted discrete action space with continuous parameters for image retouching task. At every step, the agent selects one action, such as color adjustment, white balancing and exposure adjustment, and also estimates an optimal parameter for that action (e.g. the exposure value) to process the whole image. After a fixed number of steps, the image would become more appealing.

While most of the works operate on the whole image or image regions, pixelRL (Furuta, Inoue, and Yamasaki 2019) used pixel-wise operations, which allows fine-grained adjustment for images.

In RL setting, the agent learns to either directly or indirectly maximize the discounted cumulative reward. Reward function is designed to encourage the agent to make proper decisions, so that the processed image converges to the desired output. Some previous methods used image pairs to compute reward, and the gap between the input and target images are measured. Other methods utilized adversarial learning and took the output of the discriminator to measure the quality of the processed images. We follow the former approach.

Our Method

In this section, we present the details of our algorithm and how we design each element of MDP for MRI reconstruction.

Problem formulation

With our MDP modeling, every pixel corresponds to an agent, which models a multi-agent problem. The state is simply the input image at each time step. After observing the state $s^{(t)}$ at time step t, each agent selects one action in the action set, modifies its pixel value, and then receives a reward $r^{(t)}$ that measures how much the output pixel value is improved compared with the previous one. The image is reconstructed in an iterative way, as the agents collectively change the image. At $t = 0$, the state is the image from zero-filled down-sampled k-space, i.e. $s^{(0)} = x$. Given the input image as $s^{(t)}$ at time step t, the output image will be the next state $s^{(t+1)}$. We fix the total time step $T = 3$ in a full episode. After the final time step, the output image is adjusted by data consistency step, and that is the final output.

Actions

Building on pixelRL, we made several modifications to better fit MRI reconstruction task, which include redesigning the action set and introducing learning of the continuous action parameters.

Listed in Table 1, our action set contains 12 actions. Action 0 is to keep the pixel value intact. Action 1 to 10 are classical image filters, and they have clear visual effects and physical significance.

Table 1: Actions for MRI reconstruction.

	action	filter size	parameter
0	do nothing	-	-
1	box filter	5x5	-
2	bilateral filter	5x5	$\sigma_c = 1.0, \sigma_S = 5.0$
3	median filter	5x5	-
4	Gaussian filter	5x5	$\sigma = 0.5$
5	Laplace filter	3x3	learnable p_L
6	Sobel filter (left)	3x3	learnable p_{S1}
7	Sobel filter (right)	3x3	learnable p_{S2}
8	Sobel filter (up)	3x3	learnable p_{S3}
9	Sobel filter (down)	3x3	learnable p_{S4}
10	unsharp masking	5x5	$\sigma = 0.5$, learnable p_u
11	pixel value -= 3	-	-

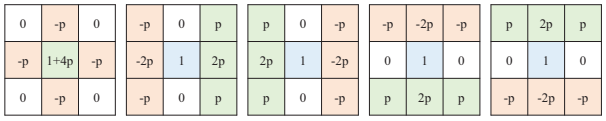

Figure 1: From left to right: Laplace filter for contrast enhancement, Sober filters for edge enhancement of four sides (left, right, up, down). Each filter has its own parameter p which is learned to control the magnitude of enhancement.

In particular, box, bilateral, median and Gaussian filters are different types of smoothing filters. They were the main operations that were utilized in pixelRL for image denoising. Although the artifacts resulted from down-sampling k-space technically are not random noise, we still find these smoothing filters helpful for MRI reconstruction and bigger kernel size (5x5) yields better visual effect.

Smoothing filters are not enough to deal with blur caused by down-sampling. Therefore, we make use of some typical sharpening filters (Action 5 to 10) for detail and edge enhancement. In view of the fact that the conventional sharpening filters like Laplace and Sobel filters are used to extract edges and thus will remove the low frequency component, we modify these filters at the center location (add the original pixel value) so that they do not filter out the mean value of the local patches. The modified filters are illustrated in Figure 1. The modified Laplace filter amplifies the contrast between the pixel on the center and its 4-neighbours, and the modified Sobel filters can enhance the edges on four sides (left, right, up, down). Action 10 is unsharp masking, which subtracts a Gaussian-blurred version of the image from itself to produce a sharpened version, and is formulated as $(1+p_u)x - p_u Gaussian(x)$, where x is the image and p_u is the amount of the Gaussian-blurred version to be subtracted. In our design, the magnitude of detail and edge enhancement depends on the current state at different time steps. That is, each of the filters has a flexible parameter to control the magnitude.

In a fully-sampled MRI image, the background originally has pixel values close to 0, whereas the aliasing effects might increase some of the pixel values up to about 10. We also observed that some low-valued pixels of anatomical struc-

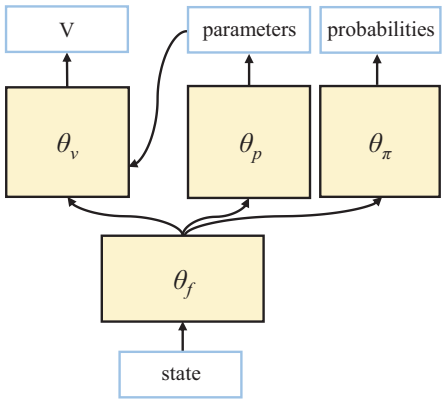

Figure 2: The architecture of our model.

tures, such as brain sulcus, are almost obscured by aliasing and are challenging to recover using the above-mentioned filters. Hence we add Action 11 to counter this. Note that Action 11 is the only action associated with the absolute pixel value, which is difficult to learn and should be set manually according to the data distribution.

Reward

Pixel values of input and target images are used to compute reward.

At time step t, the pixel[2] x changes from $x^{(t)}$ to $x^{(t+1)}$ after taking action, and then the reward is $r^{(t)} = |x^{(t)} - x_{target}| - |x^{(t+1)} - x_{target}|$. The reward is positive when the pixel gets closer to its target value, otherwise is zero or negative.

Intuitively, we are more interested in the output of the final time step rather than intermediate results. We set the discount factor $\gamma = 1$, so that the discounted cumulative reward of the full episode is $R^{(0)} = \Sigma_{t=0}^{t=T-1}\gamma^t r^{(t)} = |x^{(0)} - x_{target}| - |x^{(T)} - x_{target}|$, which is the improvement regarding the input and desired output (target). Maximizing the cumulative rewards summed up over the whole image is analogous to minimizing the L_1 error between the final output image and the target image.

Model

The whole architecture of our model is illustrated in Figure 2. Our model consists of two modules, discrete policy module and continuous parameter module. The former outputs probability distribution of discrete actions, determining which action the agent is going to take, while the latter outputs the continuous action parameters for the filters.

Considering that both modules take the state as input, we build these two modules with a shared feature extractor, denoted as θ_f. We alternately update the model for the two modules using two reinforcement learning algorithms, while keeping the value branch θ_v of both modules consistent. The value branch outputs the value for each pixel's state. Unlike (Hu et al. 2018), our design does not need the operations to be differentiable for discrete-continuous mixed action learning.

We make use of fully convolution network (FCN) (Long, Shelhamer, and Darrell 2015) to represent the modules and the feature extractor. The spatial resolution of the network remains constant throughout the layers. Using convolution layers with dilation, the receptive field for the agent is theoretically the whole image.

Discrete policy module In this part, we utilize advantage actor-critic (A2C) (Sutton and Barto 2018) to train the module and to realize end-to-end mapping from image to all pixels' actions. The discrete policy module has a policy branch θ_π and a value branch θ_v with a shared feature extractor θ_f. The policy branch outputs the probabilities of the discrete actions with softmax operation.

The loss function (for one agent at certain time step) for the discrete part consists of three terms as follows:

$$L_\pi = -\log \pi(a|s;\theta_\pi,\theta_f)(R - V(s;\theta_v,\theta_f)) \quad (1)$$

$$L_{value} = \|R - V(s;\theta_v,\theta_f)\|^2 \quad (2)$$

$$L_{entropy} = \Sigma_a \pi(a|s;\theta_\pi,\theta_f)\log \pi(a|s;\theta_\pi,\theta_f) \quad (3)$$

The first term corresponds to the policy branch, the gradient of which represents the estimation for the expectation of gradient of the trajectory's negative cumulative reward. The second term corresponds to the value branch, aiming at minimizing temporal-difference error for accurate estimation of value of the state. The third term is the negative entropy loss, which helps increase the diversity of output action probabilities and encourages action exploration. Note that we omit the superscript $^{(t)}$ for convenience and the loss above will be computed at each time step and summed for the episode.

The total loss is the weighted sum of the three terms

$$L_{discrete} = L_\pi + \lambda_1 L_{value} + \lambda_2 L_{entropy} \quad (4)$$

The loss is averaged for all the agents.

In addition to that, there are two modifications that we have incorporated into the A2C:

- We add the output of continuous parameters module to the input of θ_v since the value of state should also account for the value of continuous parameters for filters.
- During the training of discrete policy module, we freeze the parameters of continuous parameters module θ_p and update other parameters $(\theta_f, \theta_v, \theta_\pi)$ through backpropogation.

Continuous parameters module With global pooling layer, the output action parameters are shared for all the pixels from the same image. In order to restrict the action parameters in a reasonable range, we use Sigmoid function to constrain the output of the module to a fixed range of [0, 1], and map it linearly using a predefined scale. For Laplace and Sobel filters, the scale is set to 0.2, so that the action parameters never exceed 0.2. The scale is set to 1 for unsharp masking. Since the FCN is randomly initialized, at the beginning of training, the output of Sigmoid function will be close to 0.5.

[2]Both pixel value and the whole image can be denoted as x without causing confusion.

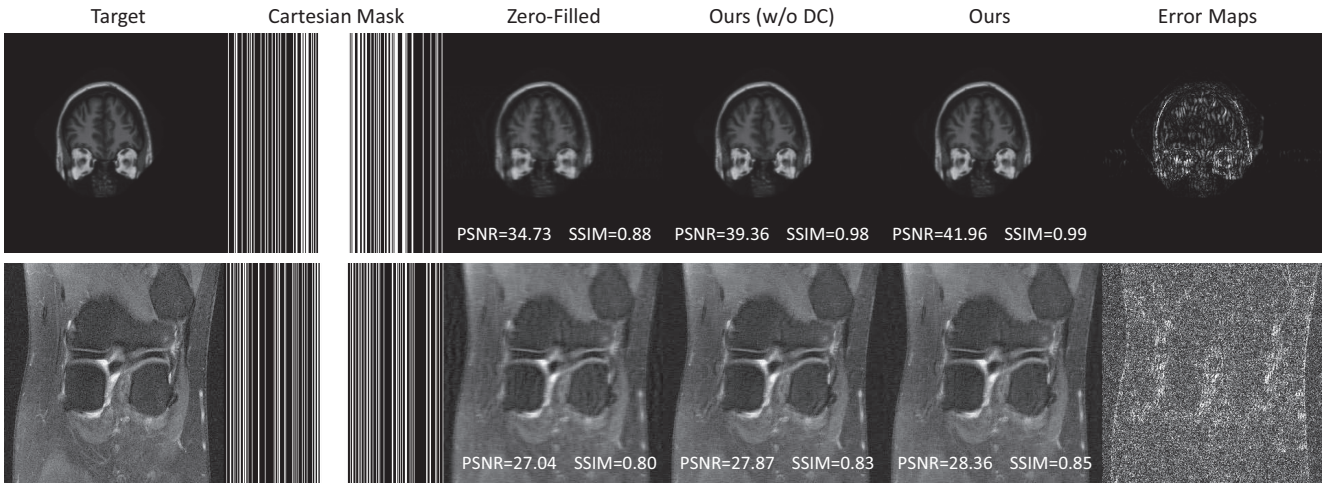

Figure 3: Results of our proposed method. The images in the first and second rows are from MICCAI dataset and fastMRI dataset respectively. The error maps compute absolute differences between our results with DC (data consistency) and are multiplied by 10 for better visualization.

Table 2: Comparison of different methods on MICCAI using 1D Cartesian Masks of different sampling ratios.

Method	10%		20%		30%		40%		50%	
	NMSE	PSNR	NMSE	PSNR	NMSE	PSNR	NMSE	PSNR	NMSE	PSNR
ZF	0.37	28.22	0.21	34.04	0.16	35.61	0.14	37.65	0.10	40.65
PANO	0.22	30.84	**0.09**	38.54	**0.08**	39.16	**0.05**	43.05	**0.04**	45.44
BM3D	0.20	31.72	0.10	39.40	0.09	39.93	0.07	43.66	0.05	45.89
DAGAN	**0.17**	**33.79**	**0.09**	39.44	**0.08**	40.20	**0.05**	**44.83**	**0.04**	47.83
Ours (w/o DC)	0.27	30.13	0.12	37.40	0.10	38.90	0.07	41.60	0.05	44.60
Ours	0.25	30.99	**0.09**	**39.52**	**0.08**	**41.07**	**0.05**	44.39	**0.04**	**47.86**

The algorithm we utilize here is a modified version of DDPG (Lillicrap et al. 2016). The loss function is

$$L_{continuous} = -\lambda_3 V(s; \theta_p) \tag{5}$$

Typically, the loss for DDPG algorithm should consist of Equation 2 and 5, and the parameters $(\theta_f, \theta_p, \theta_v)$ are supposed to be updated during training. However, in the experiment we find that the training of continuous parameters module can easily interfere with the training of discrete action module, which leads to sub-optimal result. Thus, we only update θ_p by the loss of Equation 5 to train this module.

Data Consistency

The designed reward function only considers data fidelity in the image domain. To enforce data fidelity in frequency domain, we take the output of the last time step and replace the value at the sampled positions with original data in k-space. Data consistency (DC) step can be formulated as $\hat{x} = \mathcal{F}^{-1}(S \odot y_{target} + (1 - S) \odot \mathcal{F}(x^{(T)}))$. We take the real part of $\hat{x}$ as the final result.

Experiment

Datasets

We conduct experiments on MICCAI 2013 Grand Challenge dataset[3] of T1-weighted MR images of brain tissues. We follow (Yang et al. 2017) and include 16095 training images and 50 test images. Fixed Cartesian masks with different sampling ratios are used to simulate down-sampling columns in k-space.

We also evaluate our method on the single-coil subset of fastMRI dataset (Zbontar et al. 2018) of knee images. We train our model on the training set and report our results on the first 1000 images of the validation set. The official code[4] of fastMRI is used to randomly select sampling trajectory in k-space.

Implementation details

Data preparation For MICCAI dataset, the brain tissues are located in the image center with clear background, so we apply random rotation in the range $[-10°, 10°]$, random scale in the range [0.9, 1.0] and horizontal flipping for data augmentation. We train one model for one sampling ratio. For fastMRI dataset, no data augmentation is used. Since

[3]https://my.vanderbilt.edu/masi/workshops/

[4]https://github.com/facebookresearch/fastMRI

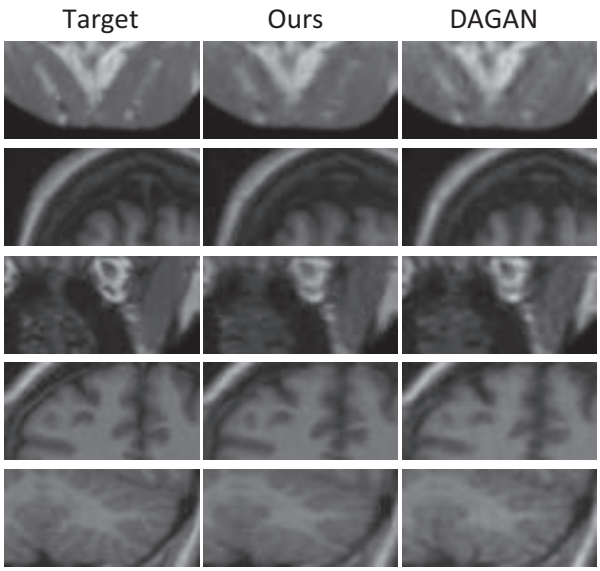

Figure 4: Comparison of our method and DAGAN. Image patches of size 40x80 are cropped for visualization.

Table 3: Results on fastMRI using random 40% Mask.

Method	NMSE	PSNR	SSIM
ZF	0.17	28.6	0.83
Unet	0.14	**30.4**	0.87
Ours (w/o DC)	0.15	29.7	0.86
Ours	**0.14**	30.3	**0.88**

the images are of various sizes, we crop the 320x320 center region and then do the down-sampling. In addition, images are normalized and rescaled into range [0, 255].

Training We set batch size to 24. Adam and SGD optimizers are used for the discrete and continuous module respectively. The learning rate is 1×10^{-3} initially and will decay linearly to 5×10^{-4}. The network will firstly conduct warm-up training in which only discrete policy module is trained. The whole number of training iterations is 10k and the alternation between two modules happens every 2 iterations. Empirically, we set the loss weights $\lambda_1 = 0.25$, $\lambda_2 = 0.1$, and $\lambda_3 = 0.5$. Our code is available[5].

Evaluation metrics Normalized mean square error (NMSE), Peak Signal-to-Noise Ratio (PSNR) and the Structural Similarity Index (SSIM) (Ledig et al. 2017) are used for evaluating MRI reconstruction.

Quantitative and qualitative results

Some results of our proposed method are shown in Figure 3.

In Table 2, we list zero-filling (ZF) as the baseline and several previous best-performing methods evaluated in (Yang et al. 2017) on MICCAI dataset, namely TV (Lustig, Donoho, and Pauly 2007), PANO (Qu et al. 2014), BM3D (Eksioglu 2016) and DAGAN (Yang et al. 2017). Without data consistency, our method can already improve the corrupted images by a fairly large margin. With data consistency, our models perform on par with DAGAN, which verifies that data fidelity in both image and frequency domain is indispensable.

As can be seen in Figure 4, the output images of our method are slightly smoother than that of DAGAN. It can be observed in the first two rows that the output of DAGAN contains a few regions of fluctuating intensity that are originally flat in the target images.

The results on fastMRI dataset are listed in Table 3. This dataset is more difficult because the target images contain noise and differ drastically in intensity range. After reconstruction, the errors on the background are still non-negligible. Nonetheless, our reconstruction method improves the images on all three metrics. We also implemented Unet (Zbontar et al. 2018) for comparison, which has much larger capacity than our model. We found that the performance of Unet is similar to ours for 40% mask and is inferior for larger sampling ratios.

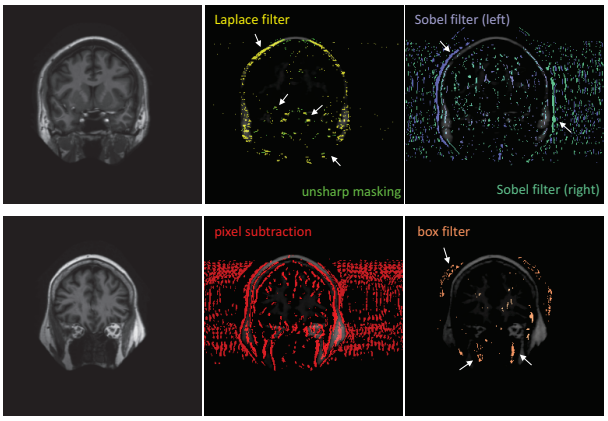

Figure 5: Example of action distribution on pixels. The target images are shown in the left column. We color the pixel if its corresponding agent takes certain action and the action changes the pixel value. Arrows indicates the regions to be highlighted. The input images are overlaid for visualization. Better viewed in color.

Learned action distribution

We display the learned action distribution on pixels in Figure 5. The trained agent is able to select actions in relation to the anatomical structures of brain tissues. Contrast-enhancing filters (Laplace filter and unsharp masking) are mainly employed on image regions corresponding to skull and Cerebrospinal fluid, where high intensity is to be restored. The Sobel filters for vertical-line enhancement show up along the left and right sides of the skull in image. It is noteworthy that whether Sobel filter increases or decreases the pixel value depends on the intensity distribution of its neighboring pixels. When properly located, Sobel filter can also reduce artifacts in background. Normally, pixel value subtraction is the most chosen action, effectively removing aliasing in the

[5]https://github.com/wentianli/MRI_RL

Figure 6: Distribution of pixel value subtraction taken at different time steps.

Table 4: Ablation study for fixed or learned parameters on MICCAI using 30% 1D Cartesian Mask.

learned	DC	NMSE	PSNR	SSIM
		0.11	37.39	97.56
	✓	0.09	40.12	98.39
✓		0.10	38.90	97.88
✓	✓	**0.08**	**41.07**	**98.69**

background and reducing the pixel values at brain sulcus region of the image. Box filter smooths regions outside the skull as well as inside the spine.

The action distribution changes at different time steps. An example is illustrated in Figure 6. As the aliasing in background is gradually suppressed, the number of subtracted pixels declines.

The model spontaneously discovered the similarity between actions belonging to the same group. That is, similar actions are likely to be taken successively. By counting the different actions taken by the same agent at two adjacent timesteps and normalizing the numbers by the frequencies of the actions, we found that the most probable action pair is unsharp masking with Laplace filter, followed by bilateral filter with box filter and median filter with box filter.

Learned action parameters

To show the necessity of learning action parameters, we conducted an ablation study and trained a model with fixed action parameters. We simply removed the continuous parameters module, reduced the total training iterations (alternate update was no longer needed), as well as set p_L, p_{S1}, p_{S2}, p_{S3}, p_{S4} to 0.1 and p_u to 0.5. The result is in Table 4.

The learned parameters vary when trained under different sampling ratios. Averaged on the test images of MACAII and over the episode, we found that the learned parameters are mostly above 0.6 for 10% mask, but below 0.5 for 50% mask. It fits with our expectation that larger magnitude for contrast and edge enhancement is needed for lower sampling ratios.

Discussion

The behavior of the agents can potentially be an indicator of the reliability of the output. When the agent behaves abnormally (e.g. some details emerges in a confusing way after reconstruction), the result may be rejected by the user. The behavior can also help speculate whether the input is out of distribution. This is not achievable with black-box methods.

Our method has several limitations. First, our method cannot completely remove the aliasing in the background when the sampling ratio is very low. Second, as the designed operations are all local, the output is restricted. As the aliasing appears somewhat structural, it might be helpful to design operations that consider pixel relations in longer ranges. Third, our method becomes harder to interpret when the action of pixel value subtraction occurs in a disorganized way on undesirable areas, especially the brain tissue region of the image. A more proper action set may help alleviate this issue.

A possible solution addressing all the aforementioned limitations could be to combine our method with classical ones. Our model could potentially be used as an image enhancement module to process the output of the classical methods.

Conclusion

We proposed a deep reinforcement learning based framework for MRI reconstruction. The actions we designed are complementary and understandable to users. We mixed A2C and DDPG algorithms to learn the discrete policy and continuous action parameters by alternate update, which effectively helped the models adapt to different sampling ratios. Through experiments, we demonstrated that our method is competitive compared with previous ones and brings in interpretability. By revealing the reconstruction process, our method provides *know-how* to users. In the future, we wish to further explore *know-why* and improve the interpretability for MRI reconstruction.

References

Block, K. T.; Uecker, M.; and Frahm, J. 2007. Undersampled radial mri with multiple coils. iterative image reconstruction using a total variation constraint. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine* 57(6):1086–1098.

Caballero, J.; Price, A. N.; Rueckert, D.; and Hajnal, J. V. 2014. Dictionary learning and time sparsity for dynamic mr data reconstruction. *IEEE transactions on medical imaging* 33(4):979–994.

Eksioglu, E. M. 2016. Decoupled algorithm for mri reconstruction using nonlocal block matching model: Bm3d-mri. *Journal of Mathematical Imaging and Vision* 56(3):430–440.

Fan, Z.; Sun, L.; Ding, X.; Huang, Y.; Cai, C.; and Paisley, J. 2018. A segmentation-aware deep fusion network for compressed sensing mri. In *ECCV*.

Furuta, R.; Inoue, N.; and Yamasaki, T. 2019. Fully convolutional network with multi-step reinforcement learning for image processing. In *AAAI*.

Hammernik, K.; Klatzer, T.; Kobler, E.; Recht, M. P.; Sodickson, D. K.; Pock, T.; and Knoll, F. 2018. Learning a variational network for reconstruction of accelerated mri data. *Magnetic resonance in medicine* 79(6):3055–3071.

Han, Y.; Yoo, J.; Kim, H. H.; Shin, H. J.; Sung, K.; and Ye, J. C. 2018. Deep learning with domain adaptation for accelerated projection-reconstruction mr. *Magnetic resonance in medicine* 80(3):1189–1205.

Han, Y.; Sunwoo, L.; and Ye, J. C. 2019. k-space deep learning for accelerated mri. *IEEE transactions on medical imaging*.

Hu, Y.; He, H.; Xu, C.; Wang, B.; and Lin, S. 2018. Exposure: A white-box photo post-processing framework. *ACM Transactions on Graphics (TOG)* 37(2):26.

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

Jaspan, O. N.; Fleysher, R.; and Lipton, M. L. 2015. Compressed sensing mri: a review of the clinical literature. *The British journal of radiology* 88(1056):20150487.

Ledig, C.; Theis, L.; Huszár, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z.; et al. 2017. Photo-realistic single image super-resolution using a generative adversarial network. In *CVPR*.

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

Lillicrap, T. P.; Hunt, J. J.; Pritzel, A.; Heess, N.; Erez, T.; Tassa, Y.; Silver, D.; and Wierstra, D. 2016. Continuous control with deep reinforcement learning. In *ICLR*.

Long, J.; Shelhamer, E.; and Darrell, T. 2015. Fully convolutional networks for semantic segmentation. In *CVPR*.

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

Lustig, M.; Donoho, D.; and Pauly, J. M. 2007. 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.

Ning, B.; Qu, X.; Guo, D.; Hu, C.; and Chen, Z. 2013. Magnetic resonance image reconstruction using trained geometric directions in 2d redundant wavelets domain and non-convex optimization. *Magnetic resonance imaging* 31(9):1611–1622.

Park, J.; Lee, J.-Y.; Yoo, D.; and So Kweon, I. 2018. Distort-and-recover: Color enhancement using deep reinforcement learning. In *CVPR*.

Qu, X.; Guo, D.; Ning, B.; Hou, Y.; Lin, Y.; Cai, S.; and Chen, Z. 2012. Undersampled mri reconstruction with patch-based directional wavelets. *Magnetic resonance imaging* 30(7):964–977.

Qu, X.; Hou, Y.; Lam, F.; Guo, D.; Zhong, J.; and Chen, Z. 2014. Magnetic resonance image reconstruction from undersampled measurements using a patch-based nonlocal operator. *Medical image analysis* 18(6):843–856.

Ravishankar, S., and Bresler, Y. 2010. Mr image reconstruction from highly undersampled k-space data by dictionary learning. *IEEE transactions on medical imaging* 30(5):1028–1041.

Razzak, M. I.; Naz, S.; and Zaib, A. 2018. Deep learning for medical image processing: Overview, challenges and the future. In *Classification in BioApps*. Springer. 323–350.

Sahba, F., and Tizhoosh, H. R. 2003. Filter fusion for image enhancement using reinforcement learning. In *CCECE 2003-Canadian Conference on Electrical and Computer Engineering. Toward a Caring and Humane Technology (Cat. No. 03CH37436)*, volume 2, 847–850. IEEE.

Schlemper, J.; Caballero, J.; Hajnal, J. V.; Price, A. N.; and Rueckert, D. 2017. A deep cascade of convolutional neural networks for dynamic mr image reconstruction. *IEEE transactions on Medical Imaging* 37(2):491–503.

Sun, J.; Li, H.; Xu, Z.; et al. 2016. Deep admm-net for compressive sensing mri. In *NeurIPS*.

Sun, L.; Fan, Z.; Huang, Y.; Ding, X.; and Paisley, J. 2018. Compressed sensing mri using a recursive dilated network. In *AAAI*.

Sutton, R. S., and Barto, A. G. 2018. *Reinforcement learning: An introduction*. MIT press.

Wang, S.; Su, Z.; Ying, L.; Peng, X.; Zhu, S.; Liang, F.; Feng, D.; and Liang, D. 2016. Accelerating magnetic resonance imaging via deep learning. In *International Symposium on Biomedical Imaging*, 514–517. IEEE.

Yang, G.; Yu, S.; Dong, H.; Slabaugh, G.; Dragotti, P. L.; Ye, X.; Liu, F.; Arridge, S.; Keegan, J.; Guo, Y.; et al. 2017. Dagan: deep de-aliasing generative adversarial networks for fast compressed sensing mri reconstruction. *IEEE transactions on medical imaging* 37(6):1310–1321.

Yang, J.; Zhang, Y.; and Yin, W. 2010. A fast alternating direction method for tvl1-l2 signal reconstruction from partial fourier data. *IEEE Journal of Selected Topics in Signal Processing* 4(2):288–297.

Yu, K.; Dong, C.; Lin, L.; and Change Loy, C. 2018a. Crafting a toolchain for image restoration by deep reinforcement learning. In *CVPR*.

Yu, R.; Liu, W.; Zhang, Y.; Qu, Z.; Zhao, D.; and Zhang, B. 2018b. Deepexposure: Learning to expose photos with asynchronously reinforced adversarial learning. In *NeurIPS*.

Zbontar, J.; Knoll, F.; Sriram, A.; Muckley, M. J.; Bruno, M.; Defazio, A.; Parente, M.; Geras, K. J.; Katsnelson, J.; Chandarana, H.; Zhang, Z.; Drozdzal, M.; Romero, A.; Rabbat, M.; Vincent, P.; Pinkerton, J.; Wang, D.; Yakubova, N.; Owens, E.; Zitnick, C. L.; Recht, M. P.; Sodickson, D. K.; and Lui, Y. W. 2018. fastMRI: An open dataset and benchmarks for accelerated MRI.

Zhang, Z.; Romero, A.; Muckley, M. J.; Vincent, P.; Yang, L.; and Drozdzal, M. 2019. Reducing uncertainty in undersampled mri reconstruction with active acquisition. In *CVPR*.

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