Highly Scalable Image Reconstruction using Deep Neural Networks with Bandpass Filtering

Joseph Y. Cheng, Feiyu Chen, Marcus T. Alley, John M. Pauly, Member, IEEE, and Shreyas S. Vasanawala

Abstract—To increase the flexibility and scalability of deep neural networks for image reconstruction, a framework is proposed based on bandpass filtering. For many applications, sensing measurements are performed indirectly. For example, in magnetic resonance imaging (MRI), data are sampled in the frequency domain. The introduction of bandpass filtering enables leveraging known imaging physics while ensuring that the final reconstruction is consistent with actual measurements to maintain reconstruction accuracy. We demonstrate this flexible architecture for reconstructing subsampled datasets of MRI scans. The resulting high subsampling rates increase the speed of MRI acquisitions and enable the visualization rapid hemodynamics.

Index Terms—Magnetic resonance imaging (MRI), Compressive sensing, Image reconstruction - iterative methods, Machine learning, Image enhancement/restoration(noise and artifact reduction).

I. INTRODUCTION

CONVOLUTIONAL neural network (CNN) is a powerfully flexible tool for computer vision and image processing applications. Conventionally, CNNs are trained and applied in the image domain. With the fundamental elements of the network as simple convolutions, CNNs are simple to train and fast to apply. The intensive processing can be easily reduced by focusing on localized image patches. CNNs can be trained on smaller images patches while still allowing for the networks to be applied to the entire image without any loss of accuracy.

For applications where image data are indirectly collected, this scalability and flexibility of CNNs are lost. As a specific example, we focus our proposed approach on magnetic resonance imaging (MRI) where the data acquisition is performed in the frequency domain, or k-space domain. For MRI, data at only a single k-space location can be measured at any given time; this process results in long acquisition times. Scan times can be reduced by simply subsampling the acquisition. Being able to reconstruct MR images from vastly subsampled acquisitions has significant clinical impact by increasing the speed of MRI scans and enabling visualization of rapid hemodynamics [1]. Using advanced reconstruction algorithms, images can be reconstructed with negligible loss in image quality despite high subsampling factors (> 8 over Nyquist). To achieve this performance, these algorithms exploit the data acquisition model with the localized sensitivity profiles of high-density receiver coil arrays for “parallel imaging” [2]–[6]. Also, image sparsity can be leveraged to constrain the reconstruction problem for compressed sensing [7]–[9]. With the use of nonlinear sparsity priors, these reconstructions are performed using iterative solvers [10]–[12]. Though effective, these algorithms are time consuming and are sensitive to tuning parameters which limit their clinical utility.

We propose to use CNNs for image reconstruction from subsampled acquisitions in the spatial-frequency domain, and transform this approach to become more tractable through the use of bandpass filtering. The goal of this work is to enable an additional degree of freedom in optimizing the computation speed of reconstruction algorithms without compromising reconstruction accuracy. This hybrid domain offers the ability to exploit localized properties in both the spatial and frequency domains. More importantly, if the sensing measurement is in the frequency domain, this architecture enables simple parallelization and allows for scalability for applying deep learning algorithms to higher and multi-dimensional space.

II. RELATED WORK

Deep neural networks have been designed as a compelling alternative to traditional iterative solvers for reconstruction problems [13]–[20]. Tuning parameters for conventional solvers, such as regularization parameters and step sizes, are learned during training of these networks which increases the robustness of the final image reconstruction algorithm. Adjustable parameters, such as learning rates, only need to be determined and set during the training phase. Also, these networks have a fixed structure and depth, and the networks are trained to converge after this fixed depth. This set depth limits the computational complexity of the reconstruction with little to no loss in image quality. Further, computational hardware devices are optimized to rapidly perform the fundamental operations in a neural network.

Three main obstacles limit the use of CNNs for general image reconstruction. First, previously proposed networks do not explicitly enforce that the output will not deviate from the measured data [17], [18]. Without a data consistency step, deep networks may create or remove critical anatomical and pathological structures, leading to erroneous diagnosis. Second, if the measurement domain is not the same domain as where the CNN is applied (such as in the image domain) and a data consistency step is used, the training and inference can no longer be patch based. If only a small image patch is used, known information in the measurement domain (k-space domain for MRI) is lost. As a result, CNNs must be trained and applied on

J.Y. Cheng, M.T. Alley, and S.S. Vasanawala are with the Department of Radiology, Stanford University, Stanford, CA, USA email: jycheng@stanford.edu
F. Chen and J.M. Pauly are with the Department of Electrical Engineering, Stanford University, Stanford, CA, USA.
fixed image dimensions and resolutions [14]–[16], [19], [20]. This limitation increases memory requirements and decreases speed of training and inference. Lastly, parallelization of the training and inference of the CNN is not straightforward: specific steps within the CNN (such as transforming from k-space domain to image domain) require gathering all data before proceeding. To address these limitations, we introduce a generalized neural network architecture.

Here, we develop an approach for image reconstruction with deep neural networks applied to patches of data in the frequency domain. In other words, a bandpass filter is used to select and isolate the reconstruction to small localized patches in the frequency space, or k-space. Previously, Kang et al demonstrated effective de-noising with CNNs in the Wavelet domain for low-dose CT imaging [21]. Here, we extend that concept to be applicable to any frequency band, and we explicitly leverage the physical imaging model. With contiguous patches of k-space, we maintain the ability to apply the data acquisition model which enables a network architecture to enforce consistency with the measured data. Also, by selecting small patches of k-space domain, the input dimensions of the networks are reduced which decreases memory footprint and increases computational speed. Thus, the possible resolutions are not limited by the computation hardware or the acceptable computation duration for high-speed applications. Lastly, each k-space patch can be reconstructed independently which enables simple parallelization of the algorithm and further increases computational speed. With the described method, deep neural networks can be applied and trained on images with high dimensions (> 256) and/or multiple dimensions (3+ dimensions) for a wide range of applications.

III. METHOD

A. Reconstruction Overview

Training and inference are performed on localized patches of k-space as illustrated in Fig. 1. For the i-th localized k-space patch, data acquisition can be modeled as:

$$u_i = M_i A_i \left( e^{j2\pi(k_i \cdot x)} * y_i \right). \tag{1}$$

The imaging model is represented by A which transforms the desired image y_i to the measurement domain. For MRI, this imaging model consists of applying the sensitivity profile maps S and applying the Fourier transform F to transform the image to the k-space domain. Sensitivity maps S are independent of the k-space patch location and can be estimated using conventional algorithms, such as ESPIRiT [6]. Since S is set to have the same image dimensions as the k-space patch, S is faster to estimate and have a smaller memory requirement in this bandpass formulation. This imaging model is illustrated in Fig. 2.

Matrix M_i is then applied to mask out the missing points in the selected patch u_i. When selecting the k-space patch of u_i with its center pixel at k-space location k_i, a phase is induced in the image domain. To remove the impact of this phase when solving the inverse problem, the phase is modeled separately as

$$e^{j2\pi(k_i \cdot x)}$$

where x is the corresponding spatial location of each pixel in y_i, and j = \sqrt{-1}. This phase is applied through an element-wise multiplication, denoted as *.

With any standard algorithm for inverse problems [10]–[12], y_i from (1) can be estimated as \(\hat{y}_i\) through a least-squares formulation with a regularization function \(R(y_i)\) and parameter \(\lambda\):

$$\hat{y}_i = \arg\min_{y_i} \left[ \left\| W \left[ M_i A_i \left( e^{j2\pi(k_i \cdot x)} * y_i \right) - u_i \right] \right\|_2^2 + \lambda R(y_i) \right]. \tag{2}$$

In (2), we introduce a windowing function W to avoid Gibbs ringing artifacts when the patch dimension is too small (< 128). The model A includes sensitivity maps S that can be considered as a element-wise multiplication in the image domain or a convolution in the k-space domain. This window function also accounts for the wrapping effect of the k-space convolution when applying S in the image domain.

In our approach, we develop an approach for image reconstruction with deep neural networks applied to patches of data in the frequency domain. In other words, a bandpass filter is used to select and isolate the reconstruction to small localized patches in the frequency space, or k-space. Previously, Kang et al demonstrated effective de-noising with CNNs in the Wavelet domain for low-dose CT imaging [21]. Here, we extend that concept to be applicable to any frequency band, and we explicitly leverage the physical imaging model. With contiguous patches of k-space, we maintain the ability to apply the data acquisition model which enables a network architecture to enforce consistency with the measured data. Also, by selecting small patches of k-space domain, the input dimensions of the networks are reduced which decreases memory footprint and increases computational speed. Thus, the possible resolutions are not limited by the computation hardware or the acceptable computation duration for high-speed applications. Lastly, each k-space patch can be reconstructed independently which enables simple parallelization of the algorithm and further increases computational speed. With the described method, deep neural networks can be applied and trained on images with high dimensions (> 256) and/or multiple dimensions (3+ dimensions) for a wide range of applications.

III. METHOD

A. Reconstruction Overview

Training and inference are performed on localized patches of k-space as illustrated in Fig. 1. For the i-th localized k-space patch, data acquisition can be modeled as:

$$u_i = M_i A_i \left( e^{j2\pi(k_i \cdot x)} * y_i \right). \tag{1}$$

The imaging model is represented by A which transforms the desired image y_i to the measurement domain. For MRI, this imaging model consists of applying the sensitivity profile maps S and applying the Fourier transform F to transform the image to the k-space domain. Sensitivity maps S are independent of the k-space patch location and can be estimated using conventional algorithms, such as ESPIRiT [6]. Since S is set to have the same image dimensions as the k-space patch, S is faster to estimate and have a smaller memory requirement in this bandpass formulation. This imaging model is illustrated in Fig. 2.

Matrix M_i is then applied to mask out the missing points in the selected patch u_i. When selecting the k-space patch of u_i with its center pixel at k-space location k_i, a phase is induced in the image domain. To remove the impact of this phase when solving the inverse problem, the phase is modeled separately as

$$e^{j2\pi(k_i \cdot x)}$$

where x is the corresponding spatial location of each pixel in y_i, and j = \sqrt{-1}. This phase is applied through an element-wise multiplication, denoted as *.

With any standard algorithm for inverse problems [10]–[12], y_i from (1) can be estimated as \(\hat{y}_i\) through a least-squares formulation with a regularization function \(R(y_i)\) and parameter \(\lambda\):

$$\hat{y}_i = \arg\min_{y_i} \left[ \left\| W \left[ M_i A_i \left( e^{j2\pi(k_i \cdot x)} * y_i \right) - u_i \right] \right\|_2^2 + \lambda R(y_i) \right]. \tag{2}$$

In (2), we introduce a windowing function W to avoid Gibbs ringing artifacts when the patch dimension is too small (< 128). The model A includes sensitivity maps S that can be considered as a element-wise multiplication in the image domain or a convolution in the k-space domain. This window function also accounts for the wrapping effect of the k-space convolution when applying S in the image domain.

In our approach, we develop an approach for image reconstruction with deep neural networks applied to patches of data in the frequency domain. In other words, a bandpass filter is used to select and isolate the reconstruction to small localized patches in the frequency space, or k-space. Previously, Kang et al demonstrated effective de-noising with CNNs in the Wavelet domain for low-dose CT imaging [21]. Here, we extend that concept to be applicable to any frequency band, and we explicitly leverage the physical imaging model. With contiguous patches of k-space, we maintain the ability to apply the data acquisition model which enables a network architecture to enforce consistency with the measured data. Also, by selecting small patches of k-space domain, the input dimensions of the networks are reduced which decreases memory footprint and increases computational speed. Thus, the possible resolutions are not limited by the computation hardware or the acceptable computation duration for high-speed applications. Lastly, each k-space patch can be reconstructed independently which enables simple parallelization of the algorithm and further increases computational speed. With the described method, deep neural networks can be applied and trained on images with high dimensions (> 256) and/or multiple dimensions (3+ dimensions) for a wide range of applications.
B. Network Architecture

We propose to solve the inverse problem of [2] with a deep neural network, denoted as $G(.)$ in Fig. 1. Any network architecture can be used for this purpose. To demonstrate the ability to incorporate known imaging physics, the architecture used is based on an unrolled optimization with deep priors [14]. More specifically, we structured the network architecture based on the iterative soft-shrinkage algorithm (ISTA) [23–26] as illustrated in Fig. 3. In this framework, two different blocks are repeated: 1) update block and 2) de-noising block (or soft-shrinkage block).

The update block enforces consistency with the measured data samples. This block is critical to ensure that the final reconstructed image agrees with the measured data to minimize the chance of hallucination. More specifically, the gradient for the least-squares component in [2] is computed for the $m$-th image estimate $y_i^m$:

$$\nabla y_i^m = B_i^*B_i y_i^m - B_i^*W_i u_i.$$  

(3)

Matrix $B_i$ applies the forward model $A_i$ for patch $i$ along with phase $e^{j2\pi (k_i \cdot x)}$, k-space subsampling operation with matrix $M_i$, and weighting $W$:

$$B_i y_i^m = W M_i A_i \left( e^{j2\pi (k_i \cdot x)} * y_i^m \right).$$  

(4)

The adjoint of $B_i$ is denoted as $B_i^*$. Original k-space measurements (network input) are denoted as $u_i$. The gradient $\nabla y_i^m$ from (3) is used to update the current estimate as

$$y_i^{m+1} = y_i^m + t \nabla y_i^m.$$  

(5)

Different algorithms can be used to determine the step size $t$. For a fixed number of iterations, the optimal step size $t$ must be determined. Here, we initialize the step size $t$ to -2, and we learn a different step size for each iteration as $t^m$ to increase model flexibility.

The de-noising block consists of a number of 2D convolutional layers to effectively de-noise $y_i^{m+1}$. The input image consists of 2 channels, since the real and imaginary components for complex data $y_i^{m+1}$ are treated as 2 separate channels. This tensor is passed through an initial convolutional layer with $3 \times 3$ kernels that expands the data to 128 feature maps. The data tensor is then passed through 5 layers of repeated $3 \times 3$ convolutional layers with 128 feature maps. A final $3 \times 3$ convolutional layer combines the 128 feature maps back to 2 channels. For a residual-type structure, the input to the de-noising block is added to the output. Batch normalization [27] and Rectified Linear Unit (ReLU) layers are used after each convolutional layer except the last one. Linear activation is applied at the last layer to ensure that the sign of the data is preserved. Convolutional layers are applied using circular convolutions. MR data are acquired in the frequency domain, and the Fourier transform operator assumes that the object of interest is repeated in the image domain. The final tensor with 2 channels is then converted to complex data as the updated image $y_i^{m+1}$.

The two blocks, update and de-noising, are repeated as “iterations.” Convolutional layer weights in the de-noising block can be kept constant for each repeated block or varied. In our experiments, $W$ was designed as a rectangle convolved with a gaussian window for a stopband of 10 pixels. Input k-space data were first zero-padded with 10 pixels before the patch-based reconstruction to account for the stopband.

Incorporating a strong prior in the form of regularization has been demonstrated to enhance high image quality despite high subsampling factors. In compressed sensing, the sparsity of the image in a transform domain, such as spatial Wavelets or finite differences, can be exploited to enable subsampling factors over 8 times Nyquist rates [9, 22]. Even though our problem formulation is similar to applying Wavelet transforms, directly enforcing sparsity in that domain may not be the optimal solution, and regularization parameters for each k-space location must be tuned. Thus, instead of solving (2) using a standard algorithm, we will be leveraging deep neural networks. The idea is that these networks can be trained to rapidly solve the many small inverse problems in a feed-forward fashion. Based on the input k-space patch, the network should be sufficiently flexible to adapt to solve the corresponding inverse problem. This deep learning approach can be considered as learning a better de-noising operation for each specific bandpass-filtered image for a stronger image prior.

After different frequency bands are reconstructed, the k-space patches are gathered to form the final image. The setup allows for flexibility in choosing patch dimensions and amount of overlap between each patch. These parameters were explored in our experiments. In the areas of overlap, outputs were averaged for the final solution.
The block and de-noising block. The estimate \( O \) once, each Fourier transform requires the most computationally expensive operation. For the conventional approach of reconstructing the entire 2D image at most dimensions of \( N \) are performed with smaller image dimensions which significantly reduces computation. For example, given initial image dimensions of \( N_y \times N_z = 256 \times 256 \), we can perform the reconstruction as solving the inverse problem for \( 64 \times 64 \) patches. In this case, we reduce the order of computation for the Fourier transform by over 28 fold. In practice, many other factors contribute to the reconstruction time, including reading/writing and data transfer. The proposed framework provides a powerful degree of freedom to optimize for faster reconstructions.

In the proposed design, we can further accelerate the reconstruction on two fronts. First, the reconstruction of each individual k-space patch can be performed independently. This property enables parallelization of the reconstruction process. The entire reconstruction can be performed in the time it takes to reconstruct a single patch which further highlights the savings from applying the Fourier transform on smaller image dimensions. Second, conventional iterative approaches to solve (2) require an unknown number of iterations for convergence and the need to empirically tune regularization parameters. With the proposed ISTA-based network, the number of iterations is fixed, and the network is trained to converge in the given number of steps.

IV. EXPERIMENT SETUP

With Institutional Board Review approval and informed consent, abdominal images were acquired using gadolinium-contrast-enhanced MRI with GE MR750 3T scanners. Both 20-channel body and 32-channel cardiac coil arrays were used. Free-breathing T1-weighted scans were collected from 301 pediatric patient volunteers using a 1–2 minute RF-spoiled gradient-recalled-echo sequence with pseudo-random Cartesian view-ordering and intrinsic motion navigation [28], [29]. Each scan acquired a volumetric image with a minimum dimension of \( 224 \times 180 \times 80 \). Data were fully sampled in the \( k_x \) direction (spatial frequency in \( x \)) and were subsampled in the \( k_y \) and \( k_z \) directions (spatial frequencies in \( y \) and \( z \)). The raw imaging data were first compressed from the 20 or 32 channels to 6 virtual channels using a singular-value-decomposition-based compression algorithm [30]. Images were modestly subsampled with a reduction factor of 1 to 2, and images were first reconstructed using compressed-sensing-based parallel imaging. Sensitivity maps for parallel imaging were estimated using ESPIRiT [6]. Compressed sensing regularization was applied using spatial wavelets [9]. Image artifacts from respiratory motion were suppressed by weighting measurements according to the degree of motion corruption [28], [31].

For training, all volumetric data were first transformed into the hybrid \( (x, k_y, k_z) \)-space. Each \( x \)-slice was considered as a separate training example. Data were divided by patient: 229 patients for training (44,006 slices), 14 patients for validation (2,688 slices), and 58 patients for testing (11,135 slices). Seventy two different sampling maps were generated using pseudo-random poisson-disc sampling [9] with reduction factors ranging from 2 to 9 with a fully sampled calibration region of \( 20 \times 20 \) in the center of the frequency space. Both uniform and variable-density sampling maps were generated. Sensitivity maps for the data acquisition model were estimated from k-space data in the calibration region using ESPIRiT [6]. As suggested in Ref. [6], 2 sets of ESPIRiT maps were used which resulted in the input and output of the de-noising block as a tensor with 4 channels: 2 ESPIRiT maps with complex data that were separated into 2 real and 2 imaginary channels. Since these 2 maps were highly correlated, we maintained the use of 128 feature maps in the de-noising block.

Each training example was normalized by the square root of the total energy in the center \( 5 \times 5 \) block of k-space data. The example was then scaled by \( 10^5 \) so that maximum pixel values in the image domain for a \( 64 \times 64 \) patch will be on the order of 100. The Adam optimizer [32] was used with \( \beta_1 = 0.9, \beta_2 = 0.999 \), and a learning rate of 0.01 to minimize the \( l_1 \) error of the output compared to the ground truth. For each training step, a batch of random data examples were selected, and random k-space subsampling masks were applied. Afterwards, the training examples were randomly cropped to the desired k-space patch dimension.

We evaluated three main features: 1) number of iteration blocks in the ISTA-based network, 2) dimensions of each k-space patch, and 3) amount of overlap between neighboring patches. First, we evaluated the impact of the number of iteration blocks in the ISTA-based network by training and applying different networks with 2, 4, 8, and 12 iteration blocks. Second, separate networks were trained for different patch dimensions: \( 32 \times 32, 48 \times 48, 64 \times 64, \) and \( 80 \times 80 \). The weights for each network were then applied to reconstruct images with varying size patches to evaluate how well the...
weights generalize. Additionally, we evaluated the reconstruction time as function of patch dimension. Assuming the ability to parallelize an unlimited number of patches, a single patch of varying dimensions was reconstructed 50 times, and the average inference time was reported. Third, the amount of overlap between neighboring reconstructed patches was evaluated. For simplicity, we used a constant 50% overlap in the $k_z$ dimension and varied the amount of overlap in the $k_y$ dimension. If unspecified, experiments were performed using a patch size of $64 \times 64$ with a 50% overlap, a variable-density subsampling with reduction factors of $R = 5.4 \pm 0.2$, and ISTA-based network with 4 iterations. The final reconstructed k-space image is transformed to the image domain by applying the adjoint imaging model $A^\dagger$.

When applicable, results were compared with the subsampled input data that was reconstructed by directly applying $A^\dagger$. Also, state-of-the-art compressed-sensing reconstructions with parallel imaging and spatial Wavelet regularization were performed for comparison. Reconstructions were evaluated using peak-signal-to-noise-ratio (PSNR), root-mean-square-error normalized by the norm of the reference (NRMSE), and structural-similarity metric (SSIM).

The proposed method was implemented in Python with TensorFlow[34]. Sensitivity map estimation with ESPRIT, compressed sensing reconstruction, and generation of poisson-disk sampling masks were performed using the Berkeley Advanced Reconstruction Toolbox (BART[35]).

V. RESULTS

Representative tests results for different frequency bands are shown in Fig. 4. The final results are comparable with compressed sensing in Fig. 5.

The impact of the number of iteration block on reconstruction performance is summarized in Table I. When more iteration blocks were used in the proposed bandpass network, the reconstruction performance improved with higher PSNR, lower NRMSE, and higher SSIM. The most gains were seen going from 2 iteration blocks with SSIM values of 0.83 and 0.85 to 4 iteration blocks with SSIM values of 0.87 and 0.88 for both uniform and variable-density sampling, respectively. With 12 iteration blocks, the bandpass network performed similarly to compressed sensing. To evaluate other components of the network, 4 iterations were used to balance between performance and depth.

The impact of patch dimensions is shown in Fig. 6. Reconstruction performance improved for larger patch dimensions during inference. By training the bandpass network specifically for smaller patch dimensions ($32 \times 32$), the reconstruction performance was best for smaller patch dimensions during inference. However, maximum PSNR and SSIM were lower and minimum NRMSE was higher for this bandpass network compared to the bandpass network trained and applied with larger patch dimensions. For all cases, the trained network can be applied to a small range of different patch dimensions. The $(64 \times 64)$-trained bandpass network had improved performance in terms of SSIM for inference on $70 \times 70$ patches, but performance began to degrade for inference on patches larger than $80 \times 80$. The $(48 \times 48)$-trained network had similar performance to the $(64 \times 64)$-trained network but with maximum SSIM shifted towards smaller patch dimensions.

The impact of patch dimension on reconstruction time is
Fig. 5. Representative results. Test data (upper right) were subsampled with uniform ((a) and (b)) and variable-density ((c) and (d)) sampling masks (lower right) to generate the input data (left column). Images were reconstructed with a network trained on the entire image (second column) and with the proposed deep network with bandpass filtering (third column). Compressed-sensing-based parallel imaging reconstructions are also displayed (fourth column).

Fig. 6. Performance as a function of patch dimension. The bandpass network was separately trained for specific patch dimensions: $32 \times 32$ (long-dash green), $48 \times 48$ (short-dash purple), $64 \times 64$ (blue), and $80 \times 80$ (dash-dot orange). During testing, weights trained were applied for varying patch dimensions. Compressed sensing (dot black) does not use the patch dimension and is plotted for reference.

Fig. 7. Reconstruction (inference) time for a single patch as a function of patch dimension on a NVIDIA Titan X card. A single patch of the specified dimensions was reconstructed 50 times with a ISTA-based neural network built with 4 iterations, and the average inference time is plotted. The total reconstruction time increased quadratically with respect to patch dimension.

summarized in Fig. 7. In this plot, average inference time to reconstruct a single patch for 50 runs is plotted with respect to the patch size. The main advantage of the approach is its ability to parallelize the reconstruction. If the entire image was considered as a single patch, the average time to reconstruct a single $512 \times 512$ image was 395 ms. A single $64 \times 64$ patch
was reconstructed with the trained CNN in 17 ms — a 23-fold speed up in reconstruction time. With enough computation resources, this gain can be realized by reconstructing the $512 \times 512$ image as $64 \times 64$ k-space patches.

The impact of overlap between neighboring patches is summarized in Fig. 8. Loss of performance was noted if the amount of overlap is less than the stopband of the window function. In this case, either part of the k-space was not reconstructed, or errors near the stopband of the window function were accentuated. Above an overlap threshold of around 15%, NRMSE and SSIM were relatively independent to changes in amount of patch overlap. Fewer patches can be reconstructed by minimizing the amount of overlap between neighboring patches. Conversely, more patch overlap yielded negligible gains. Therefore, for computational efficiency and without any loss in accuracy, the patch size should be set to the minimal size needed to account for the window stopband.

The effect of subsampling factor ($R$) on reconstruction performance is shown in Fig. 9. The bandpass network with $64 \times 64$ patches, 50% overlap, and 4 iteration blocks was trained with both uniform and variable-density subsampling ($R = 2–9$). Overall, higher subsampling factors resulted in lower PSNR and SSIM and higher NRMSE. The proposed method performed comparably to compressed sensing and with a network trained specifically on the full image. Slight discrepancies may be the result of an imbalance of subsampling factors and patterns during training. Similar trends were observed for uniform subsampling (not shown) with minor loss in performance for the same subsampling factors as seen in Table I.

**VI. DISCUSSION**

We introduced the use of bandpass filtering to enable parallelization of the image reconstruction while maintaining the use of the data acquisition model. We developed and demonstrated this approach in a deep-learning framework. The data-driven strategy with deep learning eliminates the need to engineer priors by hand for each frequency band and enables generalization of this approach to different applications.

An unrolled network based on ISTA was used as the core network. The setup can be easily adapted for more sophisticated network architecture. Also, the training can include loss functions that correlate better with diagnostic image quality such as with a generative adversarial network [18], [36]–[38]. For simplicity, we chose to implement the network for complex numbers as 2 separate channels, and we were able to demonstrate high image quality. The network can be further improved by considering complex data in each operation of the neural network [39], [40].

An advantage of the network structure is the ability to include more sophisticated imaging models. For example, non-Cartesian sampling trajectories offer the ability to reduce MRI scan durations even before subsampling the acquisition. To demonstrate this flexibility, we applied the bandpass network to hybrid Cartesian MRI. More specifically, we applied our approach to wave-encoded imaging [41]–[43]. In this case, sinusoids were used for the k-space sampling trajectory. We adapted the imaging model A to include an operator that grids the non-Cartesian sampling onto a Cartesian grid [41], [42]. Multi-slice 2D T2-weighted single-shot fast-spin-echo abdominal scans were acquired from 137 patient volunteers on a 3T scanner with a subsampling factor of 3.2. Data were divided as 104 patients (5005 slices) for training, 8 patients (383 slices) for validation, and 25 patients (1231 slices) for testing. Due to T2 signal decay and patient motion, fully sampled datasets cannot be obtained; thus, the ground truth was obtained through a compressed-sensing reconstruction for wave encoding [42]. Though the ground truth was biased towards the compressed-sensing reconstruction, we still demonstrated the ability of the bandpass network to reconstruct more general imaging models. In Fig. 10, the bandpass network method was able to recover image sharpness and yielded comparable results to compressed sensing.

The output of our proposed network had the same number of input complex data channels. This property enabled the ability to replace the estimated samples with original measurements. In doing so, the final reconstructed image will not deviate from the measured samples. Furthermore, if there are concerns with
A bandpass deep neural network architecture was developed and demonstrated here to solve the inverse problem of estimating missing measurements of subsampled MRI datasets. The main advantages of the bandpass network were leveraged when the division of data into localized patches was performed in the measurement domain. The highly scalable and flexible architecture can be adapted for other applications in MRI, such as detection and correction of corrupt measurements on a patch-by-patch basis. Additionally, this approach can be adapted for other applications, such as super-resolution or image de-noising. Working in the hybrid frequency-spatial space offers unique image-processing properties that can be further investigated.

ACKNOWLEDGEMENTS

We are grateful for the support from NIH R01-EB009690, NIH R01-EB019241, NIH R01-EB026136, and GE Healthcare.

REFERENCES

[1] S. S. Vasanawala, M. T. Alley, B. A. Hargreaves, R. A. Barth, J. M. Pauly, and M. Lustig, “Improved pediatric MRI imaging with compressed sensing,” Radiology, vol. 256, no. 2, pp. 607–616, aug 2010. [Online]. Available: http://pubs.rsna.org/doi/abs/10.1148/radiol. 10099128http://www.ncbi.nlm.nih.gov/pubmed/20529991
[2] M. A. Griswold, P. M. Jakob, R. M. Heidemann, M. Nittka, V. Jelus, J. Wang, B. Kiefer, and A. Hase, “Generalized autocalibrating partially parallel acquisitions (GRAPPA),” Magnetic Resonance in Medicine, vol. 47, no. 6, pp. 1202–1210, jun 2002. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/12111967
[3] M. Lustig and J. M. Pauly, “SPIRIT: Iterative self-consistent parallel imaging reconstruction from arbitrary k-space,” Magnetic resonance in medicine, vol. 64, no. 2, pp. 457–471, aug 2010. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/20665790
[4] K. P. Pruessmann, M. Weiger, M. B. Scheidegger, and P. Boesiger, “SENSE: sensitivity encoding for fast MRI,” Magnetic Resonance in Medicine, vol. 42, no. 5, pp. 952–62, nov 1999. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/10542355
[5] D. K. Sodickson and W. J. Manning, “Simultaneous acquisition of spatial harmonics (SMASH): fast imaging with radiofrequency coil arrays,” Magnetic Resonance in Medicine, vol. 38, no. 4, pp. 591–603, oct 1997. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/9324327
[6] M. Uecker, P. Lai, M. J. Murphy, P. Virtue, M. Elad, J. M. Pauly, S. S. Vasanawala, and M. Lustig, “ESPReS-an eigenvalue approach to autocalibrating parallel MRI: Where SENSE meets GRAPPA,” Magnetic Resonance in Medicine, vol. 71, no. 3, pp. 990–1001, mar 2014. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/23649942http://doi.wiley.com/10.1002/mrm.24753
[7] E. Candès, J. Romberg, T. Tao, E. Candes, J. Romberg, and T. Tao, “Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information,” IEEE Transactions on Information Theory, vol. 52, no. 2, pp. 489–509, feb 2006. [Online]. Available: http://ieeexplore.ieee.org/document/1580791/
[8] D. Donoho, “Compressed sensing,” IEEE Transactions on Information Theory, vol. 52, no. 4, pp. 1289–1306, apr 2006. [Online]. Available: http://ieeexplore.ieee.org/document/1614066/
[9] M. Lustig, D. Donoho, and J. M. Pauly, “Sparse MRI: The application of compressed sensing for rapid MR imaging,” Magnetic Resonance in Medicine, vol. 58, no. 6, pp. 1182–1195, dec 2007. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/17960013
[10] A. Beck and M. Teboulle, “A fast iterative shrinkage-thresholding algorithm for linear inverse problems,” SIAM J Imaging Sci, vol. 2, no. 1, pp. 183–202, 2009. [Online]. Available: http: //epubs.siam.org/doi/pdf/10.1137/080716542
[11] S. Boyd, “Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers,” Foundations and Trends in Machine Learning, vol. 3, no. 1, pp. 1–122, 2010. [Online]. Available: http://www.nowpublishers.com/article/Details/MAL-016
[12] T. Goldstein and S. Osher, “The Split Bregman Method for L1-Regularized Problems,” SIAM Journal on Imaging Sciences, vol. 2, no. 2, pp. 323–343, jan 2009. [Online]. Available: http: //epubs.siam.org/doi/10.1137/08070875891
[13] J. Adler and O. Oktay, “Learned Primal-dual Reconstruction,” arXiv:1707.06474 [math.OC], Jul 2017. [Online]. Available: https://arxiv.org/pdf/1707.06474.pdfhttp://arxiv.org/abs/1707.06474
[14] S. Diamond, V. Sitzmann, F. Heide, and G. Wetzstein, “Unrolled Optimization with Deep Priors,” arXiv: 1705.08041 [cs.CV], 2017. [Online]. Available: http://arxiv.org/abs/1705.08041

Fig. 10. T2-weighted 2D abdominal scan with wave sampling. The proposed technique was adapted to support wave-encoding (subsampling in middle right). Input (top left) and bandpass ConvNet output (top middle) are displayed along with the difference with the compressed sensing reconstruction (right). Enlarged images (bottom row) highlight recovery of fine details (arrows).
