A Bayesian Deep CNN Framework for Reconstructing k-t-Undersampled Resting-fMRI

Karan Taneja
Electrical Engineering
IIT Bombay

Prachi H. Kulkarni
Electrical Engineering
IIT Bombay

S. N. Merchant
Electrical Engineering
IIT Bombay

Suyash P. Awate
Computer Science and Engineering
IIT Bombay

***Abstract*—Undersampled reconstruction in resting functional magnetic resonance imaging (R-fMRI) holds the potential to enable higher spatial resolution in brain R-fMRI without increasing scan duration. We propose a novel approach to reconstruct k-t undersampled R-fMRI relying on a deep convolutional neural network (CNN) framework. The architecture of our CNN framework comprises a novel scheme for R-fMRI reconstruction that *jointly* learns two multilayer CNN components for (i) explicitly *filling in missing k-space data*, using acquired data in frequency-temporal neighborhoods, and (ii) *image quality enhancement* in the spatiotemporal domain. The architecture sandwiches the Fourier transformation from the frequency domain to the spatial domain between the two aforementioned CNN components, during, both, CNN learning and inference. We propose four methods within our framework, including a *Bayesian CNN* that produces *uncertainty maps* indicating the per-voxel (and per-timepoint) confidence in the blood oxygenation level dependent (BOLD) time-series reconstruction. Results on brain R-fMRI show that our CNN framework improves over the state of the art, quantitatively and qualitatively, in terms of the connectivity maps for three cerebral functional networks.**

***Index Terms*—R-fMRI reconstruction, k-t undersampling, deep convolutional neural network, k-space filling, image quality enhancement, Bayesian modeling, uncertainty.**

I. Introduction and Related Work

Resting-state functional magnetic resonance imaging (R-fMRI) [1, 2] enables the estimation of functional connectivity [3] in subjects who may be unable to perform explicit tasks during fMRI. While typical resting-state blood-oxygen-level-dependent (BOLD) signal time-series comprise frequencies less than 0.1 Hz [4], typical R-fMRI uses much higher temporal sampling rates to overcome large physiological fluctuations and noise that corrupt the weak signal, at the cost of spatial resolution. Cerebral cortical studies acquire R-fMRI with large (8–64 mm^3) voxels [5], when the cortex is 3–4 mm thick. To increase spatial resolution [6], within the same scan time, *undersampled reconstruction* methods are vital.

Some methods speedup R-fMRI acquisition using advanced pulse sequences [7, 8] and parallel imaging [9]. Other methods undersample in k-space and reconstruct using prior models like low-rank [10] or sparsity [6, 11, 12]. While non-Cartesian k-space undersampling [13, 14] can lead to artifacts, we undersample line encodes in k-space with temporal undersampling. Recent methods [15, 16] do k-t undersampled reconstruction using robust dictionary priors on the R-fMRI signal. In contrast, we propose a novel *convolutional neural network* (CNN) *framework* to reconstruct R-fMRI from k-t undersampled data (the first such approach, to the best of our knowledge).

Recent reconstruction methods use deep neural networks to reconstruct undersampled (spatial) structural MRI [17]–[23], but *not* (spatiotemporal) fMRI. [18] uses a CNN with a consistency loss coupled with a sparsity loss. [19] uses a CNN with data-consistency and data-sharing layers for dynamic MRI, but *not* R-fMRI that has much weaker signals and 10–$20\times$ more timepoints. [20] maps the zero-filled inverse-Fourier transformed image (low quality) to a reconstructed image (higher quality) using a UNet based architecture. [21] uses an encoder-decoder framework where the encoder maps the measured data to a low-dimensional manifold that feeds into the decoder. Model-based deep learning (MoDL) [22, 24] employs a CNN in an end-to-end framework for reconstructing structural MRI and diffusion MRI starting with a zero-filled Fourier-inverse reconstruction and then using an iterative procedure for spatiotemporal reconstruction. In contrast, our framework is a one-shot end-to-end framework that first fills in the missing k-space values and then enhances the image in the spatiotemporal domain. [25] explores Bayesian modeling for reconstructing cardiac MRI. Robust artificial-neural-networks for k-space interpolation (RAKI) [26, 27] extends GRAPPA using per-subject nonlinear-CNN learning using ACS data. Unlike these methods, we propose a compact CNN framework for zero-shot *spatiotemporal* R-fMRI reconstruction from k-t undersampled data with end-to-end learning.

Apart from reconstruction problems, some recent work uses CNN models in other applications of fMRI. [28] and [29] use 3D-CNN for classification of autism spectrum disorder and diagnosis of Schizophrenia respectively. [30] learns a spatio-temporal network to predict patterns of the default-mode-network map in an fMRI scan.

This paper makes several contributions. We propose a *three-stage CNN architecture*, with *end-to-end learning*, where (i) the first stage uses a CNN learned to fill in missing k-space data using acquired data in frequency-temporal neighborhoods, (ii) then includes a Fourier inverse to transform the data to the spatial domain, and (iii) finally uses a CNN learned for image quality enhancement in the spatiotemporal domain. We propose *Bayesian* deep learning with *uncertainty* estimation as

[0]Karan Taneja and Prachi H. Kulkarni contributed equally. This document is a preprint and was accepted at International Conference on Pattern Recognition (ICPR) 2020. Final version copyrighted by IEEE can be downlaoded from their website when available.

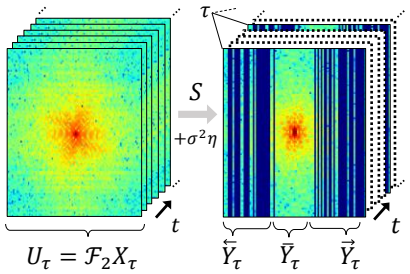

Fig. 1: **R-fMRI Undersampling Scheme in k-space + time.**

well as a loss function to make CNN learning *robust* to large physiological fluctuations typical in R-fMRI. Results on brain R-fMRI show that our CNN framework leads to significantly improved functional-network estimates over the state of the art.

II. Methods

A. Mathematical Notation

On the spatiotemporal domain of a subject undergoing R-fMRI, let the random field X model the R-fMRI BOLD signals, along with the MRI phase component that makes the resulting signal complex-valued, in a 2D brain slice within the (transaxial) acquisition plane. Let the 2D image comprise r rows and c columns, over T timepoints. Let X_t be the $r \times c$ image at timepoint $t \in [1, T]$. Let $X_v \in \mathbb{C}^T$ denote the time-series at voxel v. Let $X_{tv} \in \mathbb{C}$ be the BOLD signal, along with the complex phase component, at voxel v within X_t. Let the mathematical operator $\mathcal{F}_2$ be the 2D discrete Fourier transform (DFT). Let mathematical operators $\Re$ and $\Im$ extract the real and imaginary parts of a complex-valued variable. Let 3D random field U be the k-space representation of image X, where $U_t := \mathcal{F}_2 X_t$ is the representation at timepoint t. Let $U_{tf} \in \mathbb{C}$ be the k-space representation at frequency f within U_t. Note that U does *not* model the observed / acquired data that may be corrupted because of the noise introduced while measuring the k-space signals.

B. R-fMRI Undersampling Scheme in k-space and Time

We propose to undersample the R-fMRI acquisition in both k-space and time (Figure 1). Let the mathematical operator $\mathcal{S}_t$ model the k-space subsampling pattern during acquisition at timepoint t; $\mathcal{S}_t$ can also model temporal undersampling when *no* k-space values are acquired at t. Let Y model the acquired undersampled and corrupted data, which is a subsampled version of U corrupted with independent zero-mean Gaussian noise of variance σ^2. To undersample in time, we acquire k-space data Y_τ for a regularly-sampled subset of timepoints $\tau \in [1, T]$ with integer spacing $\Delta T \geq 2$, i.e., $\{\tau = 1, \tau = 1 + \Delta T, \cdots, \tau = T\}$ (Figure 1). This paper uses $\Delta T := 4$, resulting in $4\times$ temporal undersampling. Subsequently, for those timepoints τ for which k-space data is acquired, we propose to undersample k-space as follows. First, within the central low-frequency region of the k-space, we acquire full readout lines, producing data $\overline{Y_\tau}$ (Figure 1). Second, we undersample the remaining k-space regions by acquiring only a subset of the readout lines at locations uniformly randomly drawn within those regions, producing data $\overleftarrow{Y_\tau}$ and $\overrightarrow{Y_\tau}$ in the lines sampled on either side of $\overline{Y_\tau}$ (Figure 1). For such timepoints τ, this paper undersamples k-space by $2\times$. Thus, for frequencies f where data is acquired, $Y_{\tau f} := U_{\tau f} + \sigma^2 \eta$, where η is a standard complex-valued normal random variable.

C. CNN Framework to Reconstruct k-t Undersampled R-fMRI

1) **CNN Input:** We pass the *zero-filled* k-space data matrices, for all timepoints τ for which at least some k-space data is acquired, as the input to the CNN. Thus, for frequencies f where data is missing, we set $Y_{\tau f} := \mathbf{0}$ and pass the resulting (zero-filled temporally-undersampled) Y as the input to the CNN. This strategy enables the framework to adapt to different instances of random undersampling patterns, as long as the undersampling pattern in test-set also belongs to the same distribution of random patterns used during training.

2) **CNN Architecture – Stage 1:** The first stage of the CNN framework (Figure 2(a)) takes the sequence of zero-filled k-space data matrices $\{Y_\tau\}_{\tau=1,1+\Delta T,\cdots,T}$, which continues to be undersampled in time, and learns a nonlinear mapping $\phi(\cdot;\alpha)$, parameterized by CNN weights α, to fill the missing k-space data within each Y_τ. The mapping $\phi(\cdot;\alpha)$ uses a combination of two mappings $\overleftarrow{\phi}(\cdot;\overleftarrow{\alpha})$ and $\overrightarrow{\phi}(\cdot;\overrightarrow{\alpha})$, where the set of weights $\alpha := \overleftarrow{\alpha} \cup \overrightarrow{\alpha}$. The mappings $\overleftarrow{\phi}(\cdot;\overleftarrow{\alpha})$ and $\overrightarrow{\phi}(\cdot;\overrightarrow{\alpha})$ take as arguments the temporally-undersampled zero-filled data $\overleftarrow{Y}_\tau$ and $\overrightarrow{Y}_\tau$, respectively, and map those to produce (i) $\overleftarrow{\phi}(\overleftarrow{Y}_\tau;\overleftarrow{\alpha})$ and (ii) $\overrightarrow{\phi}(\overrightarrow{Y}_\tau;\overrightarrow{\alpha})$. The combination of $\overline{Y}_\tau$ along with the mapped outputs $\overleftarrow{\phi}(\overleftarrow{Y}_\tau;\overleftarrow{\alpha})$ and $\overrightarrow{\phi}(\overrightarrow{Y}_\tau;\overrightarrow{\alpha})$ produce an estimate of the full k-space data $\phi(Y_\tau;\alpha)$ at the sampled timepoints as shown in Figure 2(a). The mappings $\overleftarrow{\phi}(\cdot)$ and $\overrightarrow{\phi}(\cdot)$ models the complex-valued input $\overleftarrow{Y}_\tau$ as a 2-channel matrix with channels $\Re\overleftarrow{Y}_\tau$ and $\Im\overleftarrow{Y}_\tau$. Figure 2(d) visually depicts the details of the CNN architecture of this stage. The output of this stage is $\Phi(Y;\alpha) := \{\phi(Y_\tau;\alpha) : \forall\tau\}$.

3) **CNN Architecture – Stage 2:** The framework now takes the temporal sequence of estimated full-k-space 2D matrices $\phi(Y_\tau;\alpha)$ through 2D *inverse DFTs* $\mathcal{F}_2^{-1}$ to produce low-quality R-fMRI reconstructions $\mathcal{F}_2^{-1}\phi(Y_\tau;\alpha)$ at timepoints τ (*not* all $t \in [1, T]$). Let the mathematical operator $\mathcal{F}^{-1}$ model the sequence of inverse 2D DFTs across τ. The resulting temporally-undersampled R-fMRI is $\mathcal{F}^{-1}\Phi(Y;\alpha)$ (Figure 2(b)). The CNN then *temporally upsamples* $\mathcal{F}^{-1}\Phi(Y;\alpha)$ to estimate the R-fMRI images at timepoints (other than τ) for which k-space data was entirely missing. We use linear interpolation, along the temporal dimension, to keep the computational cost low. Let the mathematical operator $\mathcal{U}$ model the temporal upsampling. The resulting low-quality R-fMRI image $\mathcal{U}\mathcal{F}^{-1}\Phi(Y;\alpha)$ feeds into the third stage of the CNN framework.

4) **CNN Architecture – Stage 3:** The third stage (Figure 2(c)) of the CNN takes the low-quality complex-valued R-fMRI image $\mathcal{U}\mathcal{F}^{-1}\Phi(Y;\alpha)$ through a multilayer nonlinear mapping $\Psi(\cdot;\beta)$, parameterized by weights β, to map it to a *posterior* PDF on the high-quality reconstructed R-fMRI images. The CNN models this PDF in a factored form over all voxels v and timepoints t. These factors are parameterized by univariate Gaussians with real-valued means $\Psi^M(\mathcal{U}\mathcal{F}^{-1}\Phi(Y;\alpha);\beta)$ and

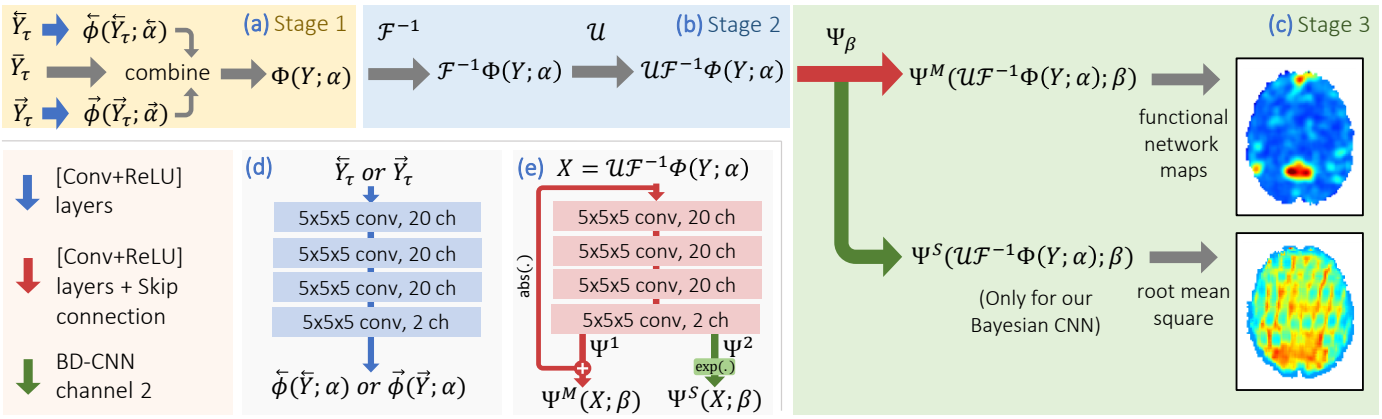

Fig. 2: **Our Convolutional Neural Network Framework.** The end-to-end CNN framework has 3 stages where (i) the first stage fills in missing k-space data, (ii) the second stage applies the inverse Fourier transform and performs temporal interpolation and (iii) the last stage performs spatiotemporal image quality enhancement. All variants are described in Section II-E.

positive real-valued standard deviations $\Psi^S(\mathcal{U}\mathcal{F}^{-1}\Phi(Y;\alpha);\beta)$. Figure 2(e) visually depicts the details of architecture of this stage. The complex-valued image $\mathcal{U}\mathcal{F}^{-1}\Phi(Y;\alpha)$, modeled as a two-channel (real and imaginary) image, passes through a sequence of convolution+ReLU layers to produce the output $\Psi(\cdot;\beta)$. Output $\Psi(\cdot;\beta)$ has two channels which are denoted by $\Psi^1(\cdot;\beta)$ and $\Psi^2(\cdot;\beta)$, each one is real-valued and is of the same size as the R-fMRI image X. We use Ψ^1 to obtain the magnitude of the output R-fMRI reconstruction and Ψ^2 to obtain the standard deviation (at each voxel at each time step) corresponding to the R-fMRI reconstruction. We introduce a skip-connection mapping that takes the magnitudes of the complex values in the input $\mathcal{U}\mathcal{F}^{-1}\Phi(Y;\alpha)$ and adds them to $\Psi^1(\cdot;\beta)$ producing $\Psi^M(\cdot;\beta)$. Thus, $\Psi^1(\cdot;\beta)$ effectively models the residual mapping. To ensure that the standard deviations in $\Psi^S(\cdot;\beta)$ are positive valued, we model these by an element-wise exponentiation of the values produced by $\Psi^2(\cdot;\beta)$. The convolution kernels in $\Psi^1(\cdot;\beta)$ and $\Psi^2(\cdot;\beta)$ have the same design as those used in $\Phi(\cdot;\alpha)$. The output of the Bayesian-modeling based CNN is the factored Gaussian PDF on reconstructed R-fMRI images given by the means in $\Psi^M(\mathcal{U}\mathcal{F}^{-1}\Phi(Y;\alpha);\beta)$ and the standard deviations in $\Psi^S(\mathcal{U}\mathcal{F}^{-1}\Phi(Y;\alpha);\beta)$.

For a CNN *without Bayesian modeling*, i.e., in the typical style, we propose to eliminate the mapping $\Psi^2(\cdot;\beta)$ such that $\Psi^M(\cdot;\beta)$ is the same as the R-fMRI reconstructed image $\Psi(\cdot;\beta)$ finally output by the CNN framework.

D. CNN Loss Functions

We construct the training set as follows. We start with fully-sampled high-quality R-fMRI BOLD signals for N slices $\{X^n\}_{n=1}^N$ across several subjects. Given X^n, we generate acquired k-space data Y^n by (i) undersampling X^n in time to give $\{X^n_\tau : \forall\tau\}$, followed by (ii) undersampling each $\mathcal{F}_2 X^n_\tau$ in k-space and introducing noise to give $Y^n_{\tau f}$. We use $\{(X^n, Y^n)\}_{n=1}^N$ for learning. We design loss functions to penalize the mismatch between the original image X^n and the (PDF on) reconstructed R-fMRI image output by the CNN. We propose three loss functions as follows.

1) **Mean-Squared-Error Loss:** The typical formulation for the optimization problem in this context is $\arg\min_{\alpha,\beta}\sum_{n=1}^N \|X^n - \Psi(\mathcal{U}\mathcal{F}^{-1}\Phi(Y^n;\alpha);\beta)\|_2^2$. We find that we can improve the stability of the learning and lead to faster convergence by also penalizing the error between the inverse DFT of the k-space filled outputs $\mathcal{F}_2^{-1}\Phi(Y^n_\tau;\alpha)$ and X^n_τ for the subset of timepoints τ where k-space data was acquired leading to the modified optimization problem:

$$\arg\min_{\alpha,\beta}(1-\lambda)\sum_{n=1}^N \|X^n - \Psi(\mathcal{U}\mathcal{F}^{-1}\Phi(Y^n;\alpha);\beta)\|_2^2 + \lambda\sum_{n=1}^N\sum_{\tau}\|X^n_\tau - \mathcal{F}_2^{-1}\Phi(Y^n_\tau;\alpha)\|_2^2, \quad (1)$$

where $\lambda \in (0,1)$ balances the two terms and is a free parameter that we train using cross validation.

2) **Robust Loss:** Because the R-fMRI image X can get corrupted with heavy-tailed physiological noise, we can replace the usual mean-squared-error penalty (stemming from a Gaussian model on the residuals) by the p-th power of the Frobenius norm, where $p \leq 2$ is a free parameter. In more general terms, the extra parameter p allows the training to adapt to a non-Gaussian PDF for the residual magnitudes between the CNN outputs and the fully-sampled images X used for training. Thus, we propose the CNN learning formulation as

$$\arg\min_{\alpha,\beta}(1-\lambda)\sum_{n=1}^N \|X^n - \Psi(\mathcal{U}\mathcal{F}^{-1}\Phi(Y^n;\alpha);\beta)\|_{2,\epsilon}^p + \lambda\sum_{n=1}^N\sum_{\tau}\|X^n_\tau - \mathcal{F}_2^{-1}\Phi(Y^n_\tau;\alpha)\|_{2,\epsilon}^p, \quad (2)$$

where $\lambda \in (0,1)$ balances the two terms, and $\|\cdot\|_{2,\epsilon}$ is the ϵ-regularized norm defined for a matrix A as $\|A\|_{2,\epsilon}^p := \sum_{v=1}^V(\|A_v\|_2^2+\epsilon)^{p/2}$, where $\epsilon := 10^{-5}$ is a small constant that

makes the function differentiable. $\lambda \in (0,1)$ and $p \in (0,2)$ are free parameters that we train using cross validation.

3) **Loss Based on Bayesian Modeling:** Assuming that the ground truth X was drawn from a factored Gaussian PDF with means in $\Psi^M(\mathcal{U}\mathcal{F}^{-1}\phi(Y;\alpha);\beta)$ and standard deviations in $\Psi^S(\mathcal{U}\mathcal{F}^{-1}\phi(Y;\alpha);\beta)$, the posterior probability density of the ground truth is given by

$$\mathcal{P}(X|Y) = \prod_{v=1}^{V}\prod_{t=1}^{T} G(X_{vt};[\Psi^M(\mathcal{U}\mathcal{F}^{-1}\Phi(Y;\alpha);\beta)]_{vt}, [\Psi^S(\mathcal{U}\mathcal{F}^{-1}\Phi(Y;\alpha);\beta)]_{vt}), \quad (3)$$

where $G(\cdot;a,b)$ is the Gaussian PDF (without robustness) with mean $a \in \mathbb{R}$ and standard deviation $b \in \mathbb{R}^+$, and the notation $[\cdot]_{vt}$ denotes the value of the argument at voxel v and timepoint t. Thus, we formulate the learning problem as maximizing the posterior of the observed training set. Taking the negative log likelihood of the objective function, the learning problem reduces to

$$\arg\min_{\alpha,\beta} \sum_{n=1}^{N}\sum_{v=1}^{V}\sum_{t=1}^{T} \frac{(X^n_{vt} - [\Psi^M(\mathcal{U}\mathcal{F}^{-1}\Phi(Y^n;\alpha);\beta)]_{vt})^2}{(\delta + [\Psi^S(\mathcal{U}\mathcal{F}^{-1}\Phi(Y^n;\alpha);\beta)]_{vt})^2} + 2\log(\delta + [\Psi^S(\mathcal{U}\mathcal{F}^{-1}\Phi(Y^n;\alpha);\beta)]_{vt}), \quad (4)$$

where $\delta := 10^{-5}$ is a small constant to avoid numerical errors during learning. During training, the per-voxel per-timepoint standard deviations tend to be higher for those spatiotemporal locations (vt) where the quality of the predictions are poorer leading to larger residual magnitudes $|X^n_{vt} - [\Psi^M(\mathcal{U}\mathcal{F}^{-1}\Phi(Y^n;\alpha);\beta)]_{vt}|$. Indeed, in the aforementioned case, a larger standard deviation keeps the first penalty term smaller. On the other hand, the second penalty term keeps a check on the standard deviations getting very large. Thus, the standard deviations can lend themselves to be interpreted as a measure of *uncertainty* (i.e., lack of confidence) in the estimates of the reconstructed R-fMRI image values. Note that robust loss and Bayesian loss make different assumptions on the distribution of voxel magnitudes and we explore them separately as different methods.

E. CNN Model Variations

We compare several models within our CNN framework for reconstructing R-fMRI images from k-t undersampled acquisitions. These are: (i) **D-CNN**: The deep CNN (D-CNN) model uses four layers each in Φ and Ψ (see Figure 2). It excludes the mapping $\Psi^S(\cdot;\beta)$. It uses the loss function described in Section II-D1. (ii) **RD-CNN**: The robust deep CNN (RD-CNN) model has the same architecture as the D-CNN. It uses the robust loss described in Section II-D2. (iii) **BD-CNN**: The Bayesian deep CNN (BD-CNN) model generalizes the D-CNN model, as shown in Figure 2, by using both $\Psi^M(\cdot;\beta)$ and $\Psi^S(\cdot;\beta)$. It uses the Bayesian-modeling based loss described in Section II-D3. (iv) **S-CNN**: The shallow CNN (S-CNN) model uses two layers in both Φ and Ψ. It excludes the mapping $\Psi^S(\cdot;\beta)$. It uses the loss function described in Section II-D1.

F. CNN Parameters and Computational Aspects

Our CNN model is very compact in terms of the number of layers in order to reduce the GPU-memory needs of storing intermediate matrices across the layers of the CNN framework. Our CNN model is also efficient in the number of parameters to reduce the computational cost during learning / optimization. The CNN framework has only around 3.3×10^5 parameters: the 4 layers in $\Phi(\cdot,\alpha)$ used for k-space filling have $2\times20\times5^3$ + $20\times20\times5^3$ + $20\times20\times5^3$ + $20\times2\times5^3$ convolution-kernel parameters and 62 bias parameters, which are about the same number as those in $\Psi(\cdot,\beta)$ used for image-quality enhancement. All the R-fMRI scans are rescaled by a constant factor to bring all inputs and outputs in the image domain within the range $[0,2]$. For all CNNs, we use the Adam optimizer [31], with a learning rate of 10^{-3} and $\beta = (0.9, 0.999)$.

III. Results and Discussion

We empirically evaluate all the methods within our CNN framework described in Section II-E, i.e., (i) S-CNN, (ii) D-CNN, (iii) RD-CNN, and (iv) BD-CNN, to reconstruct brain R-fMRI from data that is retrospectively undersampled in k-space and time. First, we compare with two recent reconstruction approaches for R-fMRI and one reconstruction scheme for fMRI (which has much higher SNR compared to R-fMRI), none of which employ neural networks: (i) **RA-DICT:** This uses the robust data-adaptive sparse dictionary modeling for R-fMRI in [15]. (ii) **WAVE:** This uses a sparse wavelet model on the spatiotemporal fMRI signal, similar to the model in [12]. (iii) **LOWRANK:** This uses a *low-rank* model on the joint k-space and temporal domain for R-fMRI, similar to the model in [10]. Further, to improve LOWRANK's performance, in the matrix of reconstructed k-space values, at the locations where the k-space data was acquired, we replace the reconstructed values by the original acquired values, as suggested in [10]. RA-DICT and WAVE both use zero-filled inverse-Fourier-transformed initialization. LOWRANK uses a zero-filled k-space initialization, as in [10]. Second, we extend a typical CNN-based dynamic-MRI reconstruction method [25] (**BDMRI**, originally proposed for cardiac MRI), for R-fMRI reconstruction. We use a 4-layer CNN (architecture same as the third stage of our BD-CNN). The input to BDMRI [25] is the Fourier-inverse of the zero-filled k-space data. We evaluate the performance of all methods using mean structural similarity (mSSIM) on three functional-network estimates, i.e., (i) dorsal attentive network (DAN), (ii) executive control network (ECN), and (iii) default mode network (DMN).

A. Results on Brain R-fMRI

To evaluate all methods, we use high-quality brain R-fMRI from the Human Connectome Project (HCP) having $2\times2\times2$ mm^3 voxels and 1.4 Hz temporal sampling rate. We evaluate input-image quality using a measure similar to tSNR [32] that we call the gray-matter temporal SNR (GM-tSNR), which we define as the average, over all the gray-matter voxels, of the ratios of (i) the root-mean-square (RMS) of the time series (because it gives a standard norm of the signal) to

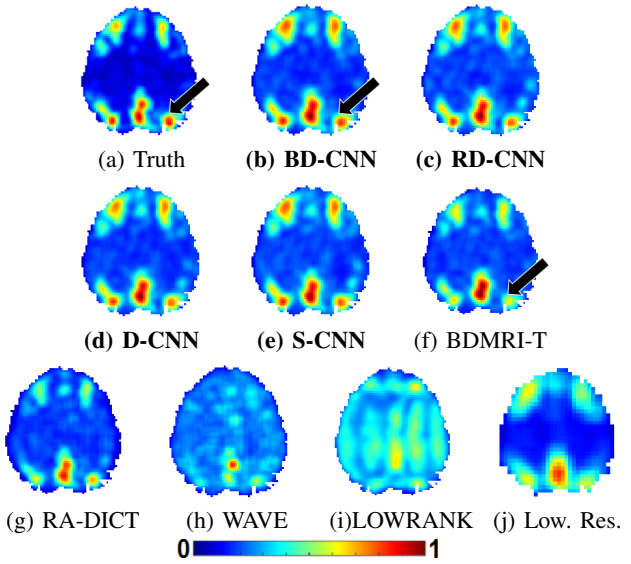

Fig. 3: **Qualitative Results on Brain R-fMRI: DAN.** DAN estimated from **(a)** original data; from *fitted models* using **(b) BD-CNN: mSSIM 0.93**, **(c) RD-CNN: mSSIM 0.92**, **(d) D-CNN: mSSIM 0.93**, **(e) S-CNN: mSSIM 0.92**, **(f)** BDMRI-T: mSSIM 0.92, **(g)** RA-DICT: mSSIM 0.91, **(h)** WAVE: mSSIM 0.85, **(i)** LOWRANK: mSSIM 0.74; and from **(j)** 8× lower spatial resolution of (a): mSSIM 0.82.

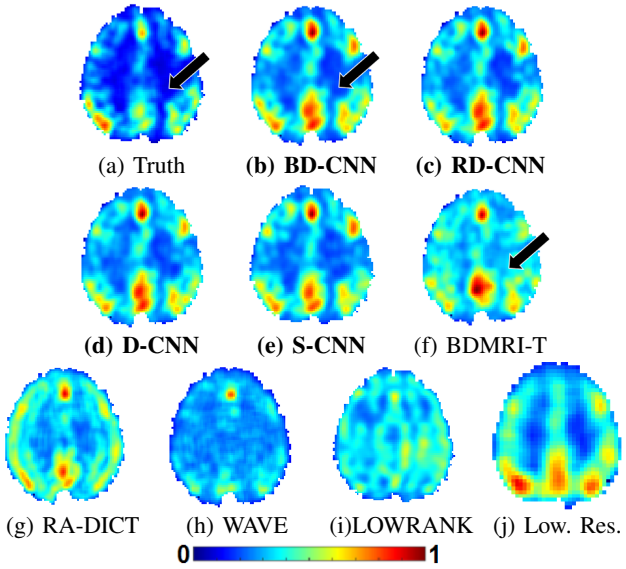

Fig. 4: **Qualitative Results on Brain R-fMRI: ECN.** ECN estimated from **(a)** original data; from *fitted models* using **(b) BD-CNN: mSSIM 0.91**, **(c) RD-CNN: mSSIM 0.90**, **(d) D-CNN: mSSIM 0.90**, **(e) S-CNN: mSSIM 0.91**, **(f)** BDMRI-T: mSSIM 0.85, **(g)** RA-DICT: mSSIM 0.85, **(h)** WAVE: mSSIM 0.87, **(i)** LOWRANK: mSSIM 0.77; and from **(j)** 8× lower spatial resolution of (a): mSSIM 0.76.

(ii) the standard deviation of the time series. The GM-tSNR of the HCP images averaged across 50 evaluation subjects is 76.1. We do *not* simulate any additional physiological noise in the

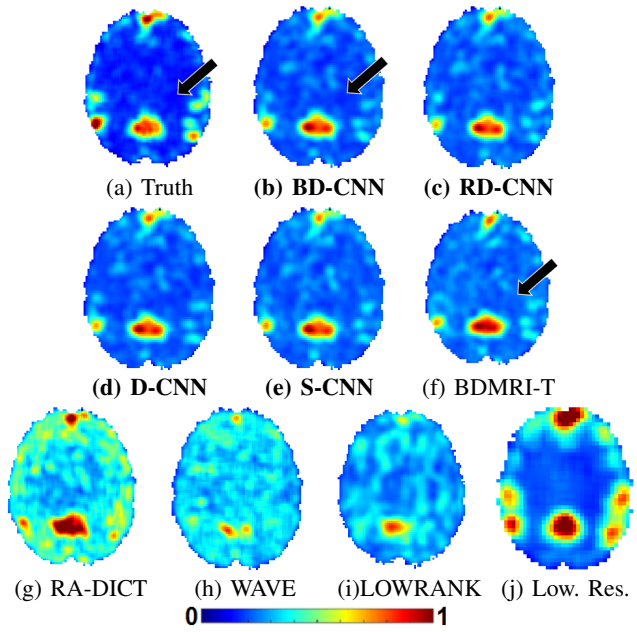

Fig. 5: **Qualitative Results on Brain R-fMRI: DMN.** DMN estimated from **(a)** original data; from *fitted models* using **(b) BD-CNN: mSSIM 0.94**, **(c) RD-CNN: mSSIM 0.94**, **(d) D-CNN: mSSIM 0.94**, **(e) S-CNN: mSSIM 0.93**, **(f)** BDMRI-T: mSSIM 0.92, **(g)** RA-DICT: mSSIM 0.85, **(h)** WAVE: mSSIM 0.86, **(i)** LOWRANK: mSSIM 0.83; and from **(j)** 8× lower spatial resolution of (a): mSSIM 0.83.

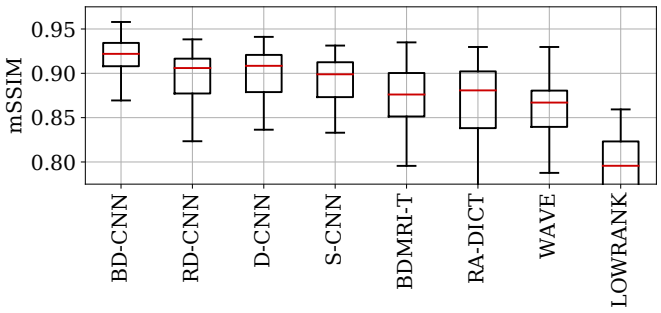

Fig. 6: **Quantitative Results on Brain R-fMRI.** Comparison of all methods, through mSSIM boxplots over 50 evaluation-set subjects and all functional-networks.

data. To simulate acquisition noise in k-space, we add complex Gaussian zero-mean independent and identically distributed (i.i.d.) noise of standard-deviation σ_{noise}, such that the inverse-Fourier-transformed R-fMRI image in the space-time domain has a reduced GM-tSNR of 73.4.

Undersampling Strategies for Methods. Let R_k and R_t be the undersampling factors in k-space and time, respectively, such that the overall undersampling factor $R = R_k * R_t$. We evaluate all methods in their ability to enable an undersampling of $R = 8$. We consider combinations of (R_k, R_t) as $(R_k = 8, R_t = 1)$, $(R_k = 4, R_t = 2)$, $(R_k = 8/3, R_t = 3)$, and $(R_k = 2, R_t = 4)$. We avoid the case of $(R_k = 1, R_t = 8)$

because this would make the temporal sampling frequency smaller than the Nyquist sampling frequency of around 0.2 Hz. We pick optimal combination for each method based on the validation set. We find that the strong spatial regularization in WAVE makes it quite insensitive to the combination of R_k and R_t, unless we limit undersampling to purely in k-space with $R_k = 8$ that would lead to poor-quality initializations using zero-filled inverse-Fourier-transforms. Thus, we choose $(R_k = 2, R_t = 4)$ for WAVE. We use $(R_k = 2, R_t = 4)$ for RA-DICT because its performance deteriorates for higher R_k possibly due to the weak and indirect spatial regularization on the dictionary coefficients. For BDMRI, a high R_k leads to poor-quality inputs that are zero-filled inverse-Fourier-transforms, making the DNN-learning more difficult. Thus, we extend BDMRI to **BDMRI-T** where we employ an optimal combination of R_k and R_t, and then perform zero-filled inverse-Fourier-transforms for each acquired timepoint followed by temporal interpolation to get the input to the DNN. We use $(R_k = 2, R_t = 4)$ for BDMRI-T as it leads to optimal performance. For our CNN-based methods, i.e., S-CNN, D-CNN, RD-CNN, and BD-CNN, we use $(R_k = 2, R_t = 4)$ as if the data were sampled at 0.35 Hz, with R-fMRI frequencies ≤ 0.10 Hz still preserved, as it leads to optimal performance. We use $(R_k = 8, R_t = 1)$ for LOWRANK because its performance deteriorates for $R_t > 1$ possibly due to the absence of measurements in entire rows of the k-space$\times$time data matrix.

Training, Validation, and Evaluation Datasets. We learn our CNN models and BDMRI-T using a training set of 5 subjects of fully-sampled *corrupted* (i.e., with introduced noise) HCP data. We learn the dictionary model within RA-DICT using the *uncorrupted* (without noise introduced) R-fMRI of the same training subjects. We tune all free parameters underlying all methods on a separate validation set of 5 subjects (different from the training set) to maximize the mSSIM averaged over subjects and networks. To evaluate the sensitivity of our results to the choice of training and validation sets (as shown in Section III-B later), we repeat this training 5 times, each time using a completely new set of 5 subjects for training and 5 other subjects for validation. We evaluate the performance of all CNN-based and baseline methods on a separate evaluation set of 50 subjects, which has *no* overlap with any training set or validation set. We estimate all resting-state functional networks using seed-based normalized time-series cross-correlations in the reconstructed image.

Qualitative and Quantitative Evaluation. Functional network estimates from BD-CNN reconstructions are qualitatively closer to the ground-truth functional networks than all other methods. Across the three connectivity networks overall (Figures 3–5), BD-CNN clearly performs better than all the baselines. Compared to BD-CNN (Figures 3–5(b)), functional networks estimated from RD-CNN (robust loss; without Bayesian modeling; Figures 3–5(c)), D-CNN (without robust loss; without Bayesian modeling; Figures 3–5(d)) and S-CNN (less depth; without robust loss; without Bayesian modeling; Figures 3–5(e)) show slightly inferior results qualitatively (with lower mSSIM values). Functional-network estimates from RA-DICT (Figures 3–5(g)), and WAVE (Figures 3–5(h)) show more undesirable artifacts and those from LOWRANK (Figures 3–5(i)) show larger distortions in the shapes, compared to our CNN-based models. For DAN, BDMRI-T reduces the intensity of the indicated network regions (Figure 3(f)) compared to the ground truth (Figure 3(a)), whereas our BD-CNN maintains it (Figure 3(b)). ECN from BDMRI-T reconstruction (Figure 4(f)) has large distortions in the contrast and shape of the network compared to the ground truth (Figure 4(a)), unlike our BD-CNN (Figure 4(b)). BDMRI-T leads to reduced contrast in DMN (Figure 5(f)) compared to the ground truth (Figure 5(a)) and our BD-CNN (Figure 5(b)). Our CNN-based methods show good denoising ability along with faithful reconstructions, as can be specially seen from the DAN and ECN estimates from our CNN-based reconstructions (Figures 3(b)–(e), 4(b)–(e)) that show better contrast than even the ground-truth DAN and ECN (Figures 3(a), 4(a)). Our four CNN-based methods show significantly higher mSSIM values compared to all the baselines i.e. BDMRI-T, RA-DICT, WAVE, and LOWRANK, over all three functional connectivity networks (Figure 6). Comparing our different CNN-based methods, we find that BD-CNN improves over all our other CNN-based methods. Our CNN framework can be extended for multicoil R-fMRI by reconstructing each channel separately, and then solving for the underlying signal using methods like [33], [34].

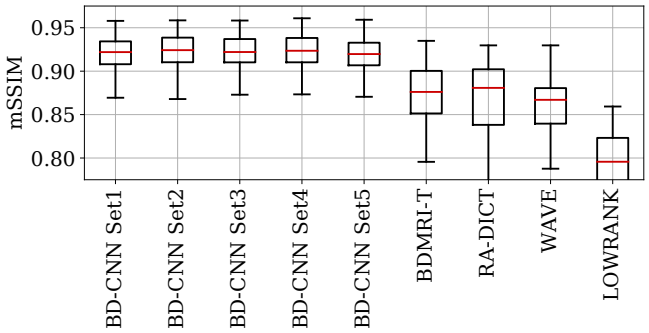

Fig. 7: **Insensitivity of BD-CNN to Choice of Training and Validation Sets.** mSSIM Boxplots, over 50 evaluation-set subjects and all functional networks, for BD-CNN learned from 5 different training and validation sets (i.e., Set1, Set2, Set3, Set4, Set5). The boxplots for baselines are as in Figure 6.

B. Ablation Studies

We perform ablation studies on our proposed BD-CNN to further analyze its use for R-fMRI reconstruction.

Sensitivity to Choice of Training and Validation Sets. Results in III-A show that (i) our CNN-based methods are preferable over the baselines, and (ii) within our CNN-based methods, BD-CNN is preferable over others. We now evaluate the sensitivity of BD-CNN to the choice of training sets and validation sets. We train our BD-CNN model over 5 different training and validation sets (all distinct from the evaluation set) from the HCP data and show that BD-CNN is quite insensitive to the choice of a specific training and validation set, with

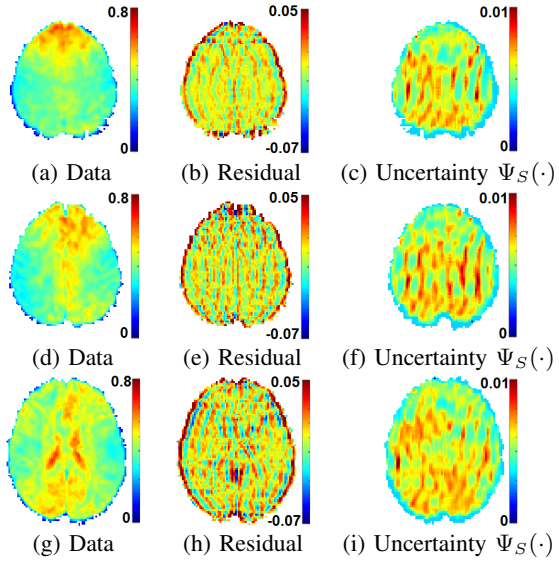

Fig. 8: **BD-CNN's Uncertainty Visualization in Reconstructed Cerebral BOLD Signals within Connectivity-Network Regions.** Original data, residual between original data and zero-filled reconstruction, and per-voxel uncertainty map (standard-deviation map $\Psi_S(\cdot)$) respectively in the **(a),(b),(c)** DAN region, **(d),(e),(f)** ECN region, and **(g),(h),(i)** DMN region.

the distributions of mSSIM values, over the evaluation set, remaining virtually unchanged (Figure 7).

Performance for Different Values of Free-Parameter λ. We compare the performance of our D-CNN at different values of λ, the balancing hyper-parameter in Equation 1. We find that performance of our D-CNN deteriorates statistically significantly as $\lambda \rightarrow 1$, compared to scenarios where λ is sufficiently less than 1. The average mSSIM (and standard deviation for all functional networks and evaluation subjects) for $\lambda \in [0, 0.75]$ is 0.90 (0.03), and for $\lambda = 1$ is 0.88 (0.03); this demonstrates the utility of the third stage of our architecture. Thus, we propose to set the value of λ significantly less than 1. The hypothesis that the mSSIM values from $\lambda \in [0, 0.75]$ and from $\lambda = 1$ come from the same distribution leads to a very low p-value (0.02 in a two-sample t-test; 0.005 in the two-sample K-S test). Hence, $\lambda \in [0, 0.75]$ can give good results. We propose to set $\lambda = 0.5$, because it leads to reduced training time in practice. We find that our other CNN-based methods follow similar trends, with respect to λ, in performance and training time.

Effect of Head Motion. We simulate head motion for each subject during the 15-minute scan as described by the model in [35] that rotates the head about the spine every minute. We choose the rotation angle θ to generate realistic head motion as suggested in [35], and add noise as described in Section III-A. We reconstruct the R-fMRI from the k-t undersampled (2× in k-space, 4× in time) and motion-corrupted data. The average mSSIM (and standard deviation) over all functional networks and evaluation subjects are (i) BD-CNN: 0.90 (0.04), (ii) BDMRI-T: 0.87 (0.05), (iii) RA-DICT: 0.86 (0.06), (iv) WAVE: 0.86 (0.02), and (v) LOWRANK: are 0.79 (0.04). This shows that our BD-CNN performs better than the baselines even for motion-corrupted acquisitions.

C. Uncertainty of Reconstruction in Cerebral BOLD Signals

Consistent with the general finding, in the literature [36], of improved performance of Bayesian DNNs, compared to their non-Bayesian counterparts, we find that our BD-CNN leads to an increase in mSSIM values, averaging around 2–3%, over RD-CNN, D-CNN, and S-CNN. Furthermore, our BD-CNN outputs a PDF over the reconstructed images, parameterized by a per-voxel mean and a per-voxel standard deviation. While BD-CNN treats the mean values as estimates of the reconstructed intensities, we can treat the standard-deviation values as estimates of the relative uncertainty, between voxels, in the reconstructed intensities. The artifacts introduced due to k-space undersampling of the original data in the transaxial acquisition plane, are clearly seen in the residuals (Figures 8(b),(e),(h)) between the original data (Figures 8(a),(d),(g)) and the zero-filled reconstruction. The corresponding per-voxel standard-deviation maps show higher values, and, thereby, higher uncertainty, in the reconstructed cerebral BOLD time-series forming spatial patterns (Figures 8(c),(f),(i)) that are similar to those resulting from zero-filled reconstructions of undersampled k-space data.

D. Enabling Higher Spatial Resolution in R-fMRI

Consider an R-fMRI image with lower spatial resolution, e.g., 4×4×4 mm^3 voxels at the same temporal sampling rate of 1.4 Hz and for the same length of time as the HCP images used in this manuscript. Consider that the scan acquires data without any undersampling in k-space and time. Such data will clearly lead to functional network maps of lower spatial resolution (Figures 3–5(j)) even though the contrast is improved over the ground truth (Figures 3–5(a)). Our BD-CNN method can enable acquisition / reconstruction of R-fMRI images with higher spatial resolution, e.g., with 2×2×2 mm^3 voxels, at the same temporal sampling rate and without increasing the scan time over the low-resolution scan. Our BD-CNN results in Section III-A utilize the same scan time as the fully-sampled (in k-space and time both) low-spatial-resolution scan, by freeing up time through 8× k-t undersampling (2× in k-space and 4× in time) and using the freed-up time to acquire data at higher spatial resolution. We then use our learned BD-CNN model to reconstruct the required high-spatiotemporal-resolution volume at each timepoint. Our CNN-based functional networks (Figures 3–5(b)) have much higher spatial resolution and mSSIM. Unlike the network maps from low-spatial-resolution data, which have improved contrast at the cost of spatial resolution, our BD-CNN reconstructions maintain high spatial resolution while improving the contrast.

We can extend the argument to a scenario where the voxel size is reduced to 1×1×1 mm^3. The 8× smaller voxels will lead to an 8× reduction in signal strength per voxel per timepoint, which is akin to an 8× increase in noise level. BD-CNN

continues to perform well and better than the baselines even at $8\times$ higher acquisition noise ($8\sigma_{\text{noise}}$), such that average GM-tSNR over all evaluation subjects reduces to 42.8. The average mSSIM (and standard deviations) of all functional networks over the evaluation set for $8\times$ higher noise level are (i) BD-CNN: $0.92(0.03)$, (ii) BDMRI-T: $0.89(0.04)$, (iii) RA-DICT: $0.80(0.07)$, (iv) WAVE: $0.86(0.05)$, and (V) LOWRANK: $0.79(0.04)$. This shows the potential for our CNN-based framework to enable R-fMRI scans with higher spatial resolution inspite of an $8\times$ increase in noise level.

IV. Conclusion

We proposed a novel approach for reconstruction of k-t undersampled R-fMRI, based on a novel multi-stage CNN framework with end-to-end learning. We proposed four methods within our framework including a Bayesian CNN that estimates the uncertainty of reconstructions. The computational advantages of our framework include an efficient CNN architecture leading to low memory needs for GPU-based learning, and fast inference of the order of a minute. Our CNN framework improves the estimation of functional networks on brain R-fMRI, and also provides some insights into reconstruction quality through uncertainty maps. Results show that our CNN framework can potentially enable ($8\times$) higher spatial resolution without compromising the temporal resolution and without increasing the scan time. Future research directions include exploring (i) schemes to enable the use of very large / complex CNN models for R-fMRI reconstruction, because the space complexity increases with model size, and (ii) feasible extensions of the current framework that can leverage information from spatially neighboring slices during reconstruction, thereby enabling undersampling in the third spatial direction as well.

References

[1] B Biswal, F Yetkin, V Haughton, and J Hyde, "Functional connectivity in the motor cortex of resting human brain using echo-planar MRI," *Magn. Res. Med.*, vol. 34, pp. 537–41, 1995.

[2] S Smith, P Fox, K Miller, D Glahn, P Fox, C Mackay, N Filippini, K Watkins, R Toro, A Laird, and C Beckmann, "Correspondence of the brain's functional architecture during activation and rest," *Proc. Nat. Acad. Sci.*, vol. 106, no. 31, pp. 106–31, 2009.

[3] W Liu, S P Awate, J Anderson, and P T Fletcher, "A functional networks estimation method of resting-state fMRI using a hierarchical Markov random field," *NIMG*, vol. 100, pp. 520–34, 2014.

[4] M Lee, C Smyser, and J Shimony, "Resting-state fMRI: a review of methods and clinical applications," *Am. J. Neuroradiol.*, vol. 34, no. 10, pp. 1866–72, 2013.

[5] T White, R Muetzel, M Schmidt, S Langeslag, V Jaddoe, A Hofman, V Calhoun, F Verhulst, and H Tiemeier, "Time of acquisition and network stability in pediatric resting-state functional magnetic resonance imaging," *Brain Connectivity*, vol. 4, no. 6, pp. 417–27, 2014.

[6] Z Fang, N Le, M Choy, and J Lee, "High spatial resolution compressed sensing (HSPARSE) functional MRI," *Magn. Res. Med.*, vol. 76, pp. 440–55, 2016.

[7] S Moeller, E Yacoub, C Olman, E Auerbach, J Strupp, N Harel, and K Ugurbil, "Multiband multislice GE-EPI at 7 tesla, with 16-fold acceleration using partial parallel imaging with application to high spatial and temporal whole-brain fMRI," *Magn. Res. Med.*, vol. 63, no. 5, pp. 1144–53, 2010.

[8] L Chen, A Vu, J Xu, S Moeller, K Ugurbil, E Yacoub, and D Feinberg, "Evaluation of highly accelerated simultaneous multi-slice EPI for fMRI," *NeuroImage*, vol. 104, pp. 452–9, 2015.

[9] P Kulkarni, K Gupta, S Merchant, and S Awate, "R-fMRI reconstruction from k-t undersampled simultaneous-multislice (SMS) MRI with controlled aliasing: Towards higher spatial resolution," in *IEEE Int. Symp. Biomed. Imag.*, 2020.

[10] M Chiew, S Smith, P Koopmans, N Graedel, T Blumensath, and K Miller, "k-t FASTER: Acceleration of functional MRI data acquisition using low rank constraints," *Magn. Res. Med.*, vol. 74, pp. 353–64, 2014.

[11] M Lustig, J Santos, D Donoho, and J Pauly, "k-t SPARSE: High frame rate dynamic MRI exploiting spatio-temporal sparsity," in *ISMRM*, 2006.

[12] L Chaari, P Ciuciu, S Meriaux, and J Pesquet, "Spatio-temporal wavelet regularization for parallel MRI reconstruction: application to functional MRI," *Magn. Reson. Mater. Phy.*, vol. 27, pp. 509–529, 2014.

[13] M Chiew, N Graedel, J McNab, N Graedel, S Smith, and K Miller, "Accelerating functional MRI using fixed-rank approximations and radial-cartesian sampling," *Magn. Res. Med.*, vol. 76, pp. 1825–1836, 2016.

[14] N Graedel, J McNab, M Chiew, and K Miller, "Motion correction for functional MRI with three-dimensional hybrid radial-cartesian EPI," *Magn. Res. Med.*, vol. 78, pp. 527–540, 2017.

[15] P Kulkarni, S Merchant, and S Awate, "Bayesian reconstruction of R-fMRI from k-t undersampled data using a robust, subject-invariant, spatially-regularized dictionary prior," in *IEEE Int. Symp. Biomed. Imag.*, 2018, pp. 302–306.

[16] P Kulkarni, S Merchant, and S Awate, "R-fMRI reconstruction from k-t undersampled data using a subject-invariant dictionary model and VB-EM with nested minorization," *Med. Imag. Anal.*, vol. 65, 2020.

[17] Y Yang, J Sun, H Li, and Z Xu, "Deep ADMM-Net for compressive sensing MRI," in *Adv. Neural Info. Proc. Systems*, 2016, pp. 10–8.

[18] S Wang, Z Su, L Ying, X Peng, S Zhu, F Liang, D Feng, and D Liang, "Accelerating magnetic resonance imaging via deep learning," in *IEEE Int. Symp. Biomed. Imag.*, 2016, pp. 514–7.

[19] J Schlemper, J Caballero, J Hajnal, A Price, and D Rueckert, "A deep cascade of convolutional neural networks for dynamic MR image reconstruction," *IEEE Trans. Med. Imag.*, vol. 37, no. 2, pp. 491–503, 2018.

[20] C Hyun, H Kim, S Lee, S Lee, and J Seo, "Deep learning for undersampled MRI reconstruction," *Physics in Medicine & Biology*, vol. 63, no. 13, pp. 135007, 2018.

[21] B Zhu, J Liu, S Cauley, B Rosen, and M Rosen, "Image reconstruction by domain transform manifold learning," *Nature*, 2018.

[22] H K Aggarwal, M P Mani, and M Jacob, "MoDL: Model-based deep learning architecture for inverse problems," *IEEE Trans. Med. Imag.*, vol. 38, no. 2, pp. 394–405, 2019.

[23] F Knoll, K Hammernik, C Zhang, S Moeller, T Pock, D K Sodickson, and M Akçakaya, "Deep-learning methods for parallel magnetic resonance imaging reconstruction: A survey of the current approaches, trends, and issues," *IEEE Signal Process. Mag.*, vol. 37, no. 1, pp. 128–140, 2020.

[24] H K Aggarwal, M P Mani, and M Jacob, "MoDL-MUSSELS: Model-based deep learning for multishot sensitivity-encoded diffusion MRI," *IEEE Trans. Med. Imag.*, vol. 39, no. 4, pp. 1268–1277, 2019.

[25] J Schlemper, D Castro, W Bai, C Qin, O Oktay, J Duan, A Price, J Hajnal, and D Rueckert, "Bayesian deep learning for accelerated MR image reconstruction," in *Machine Learning for Medical Image Reconstruction Workshop at MICCAI*, 2018, pp. 64–71.

[26] C Zhang, S Hosseini, S Weingartner, K Ugurbil, S Moeller, and M Akcakaya, "Optimized fast gpu implementation of robust artificial-neural-networks for k-space interpolation (RAKI) reconstruction," *PLOS ONE*, vol. 14, no. 10, pp. e0223315, 2019.

[27] M Akçakaya, S Moeller, S Weingärtner, and K Uğurbil, "Scan-specific robust artificial-neural-networks for k-space interpolation (RAKI) reconstruction: Database-free deep learning for fast imaging," *Magn. Res. Med.*, vol. 81, no. 1, pp. 439–453, 2019.

[28] A El-Gazzar, M Quaak, L Cerliani, P Bloem, G van Wingen, and R Thomas, "A hybrid 3DCNN and 3DC-LSTM based model for 4D spatio-temporal fMRI data: An ABIDE autism classification study," in *Context-Aware Operating Theaters and Machine Learning in Clinical Neuroimaging*, 2019, pp. 95–102.

[29] M Qureshi, J Oh, and B Lee, "3D-CNN based discrimination of schizophrenia using resting-state fMRI," *Artificial Intelligence in Medicine*, vol. 98, pp. 10–7, 2019.

[30] Y Zhao, X Li, H Huang, W Zhang, S Zhao, M Makkie, M Zhang, Q Li, and T Liu, "4D modeling of fMRI data via spatio-temporal convolutional neural networks (ST-CNN)," *IEEE Trans. Cognitive and Developmental Systems*, 2019.

[31] D Kingma and J Ba, "Adam: A method for stochastic optimization," in *Int. Conf. Learning Representations*, 2014.

[32] M Welvaert and Y Rosseel, "On the definition of Signal-To-Noise Ratio and Contrast-To-Noise Ratio for fMRI data," *PLOS ONE*, vol. 8, no. 11, pp. 1–10, 2013.

[33] K P Pruessmann, M Weiger, M B Scheidegger, and P Boesiger, "SENSE: sensitivity encoding for fast MRI," *MRM*, vol. 42, no. 5, pp. 952–962, 1999.

[34] P Roemer, W Edelstein, C Hayes, S Souza, and O Mueller, "The NMR phased array," *MRM*, vol. 16, no. 2, pp. 192–225, 1990.

[35] I Drobnjak, D Gavaghan, E Süli, J Pitt-Francis, and M Jenkinson, "Development of a functional magnetic resonance imaging simulator for modeling realistic rigid-body motion artifacts," *Magn. Res. Med.*, vol. 56, no. 2, pp. 364–380, 2006.

[36] A Kendall and Y Gal, "What uncertainties do we need in Bayesian deep learning for computer vision?," in *Adv. Neural Info. Proc. Systems*, 2017, pp. 5574–84.