Deep Learning Strategies for Reconstructing k-t Undersampled Resting State fMRI

Anubhav Goel
Department of Electrical Engineering
Indian Institute of Technology Bombay
Mumbai, India
anubhav.goel@iitb.ac.in

***Abstract*—This report has been prepared after extensive literature review and summarizes the motivation for the need of undersampling in k-space and time during the acquisition of fMRI signals. Further, the report highlights numerous classical techniques used for the reconstruction and concludes with the recent developments in Deep Learning and the applications of these methods to the above mentioned reconstruction.**

***Index Terms*—fMRI, k-t undersampling, bayesian deep learning**

I. Introduction

Resting State fMRI (R-fMRI) measures low frequency fluctuations in the Blood Oxygen Level Dependent (BOLD) signals to obtain information about the functional architecture of the brain [1]. This technique enables the identification of Resting State Networks (discernable functional communities in the brain [2]). The applications of Resting State fMRI in presurgical planning for brain tumour and epilepsy patients has shown potential along with possible applications in the diagnosis of neurological and psychiatric diseases.

In addition to the above, undersampled reconstruction in R-fMRI will lead to improved spatiotemporal resolution and improved brain connectivity maps [3]. Faster acquisition also enables the study of dynamic networks and reduces the risk of data corruption due to patient motion. Current methods to speed up scan rely on numerous techniques such fully-sampled parallel imaging [4], undersampled reconstruction using signal priors [5] and Bayesian graphical framework relying on learning data-adaptive prior models through dictionaries designed to be robust to large physiological fluctuations typical in R-fMRI signals.

With the recent developments in deep learning and the introduction of convolutional architectures which have led to efficient and accurate outcomes in several image processing tasks, there has been an interest in the application of these methods to the reconstruction task as discussed above. Deep Learning-based accelerated fMRI reconstruction is an active topic of research, with constant focus on the characterisation of these architectures and investigating when these networks may fail to reconstruct [6].

In this report, we start with a brief background of R-fMRI, and then a discussion of the classical techniques mentioned above and discuss the various shortcomings associated with them, which motivate the use of deep learning networks to improve performance. A further discussion is presented on the existing deep learning framework, its scope for improvement and the related future work.

II. Background of RfMRI

fMRI using task based or stimulus driven paradigms has been vital in the current understanding of the various functional networks in the brain. This technique focuses on the measurement of the changes in the BOLD signal in various areas of the brain during the performance of a cognitive, language or motor task. In recent times, there has been a piqued interest in the application of this technique at rest, enabling us to study functional networks in the brain called the Resting State Networks, the most fundamental of these being the Default Mode Network (DMN), first identified from PET data by Raichle et al. [7]. This can be instrumental in the study of brain disorders, neurodevelopment in children, or neurodegeneration in aging.

Studies have hypothesized that there are two large opposing systems in the brain, one including the DMN (”task-negative” or ”extrinsic”) and one involving task-based or attentional systems (”task-positive” or ”intrinsic”) [8]. In addition to the DMN, several other RSNs have been identified such as the somatosensory network and the language network, therefore, furthering the interest in this field.

Typically, fMRI signals have component frequencies less than 0.8Hz. However, conventional fMRI acquisition used higher temporal sampling rates to overcome noise and large physiological fluctuations common to these type of signals but this came at the cost of spatial resolution. Faster scans were eventually developed which did not affect the Signal-to-Noise ratio (SNR) and also enabled more reliable studies.

III. Classical Techniques

Having established the idea behind R-fMRI, we now move on the numerous classical techniques which have been used for analyzing the BOLD signals. GRAPPA is analyzed briefly before a more complete description of dictionary-learning and k-t FASTER.

A. Generalized autocalibrating partially parallel acquisitions (GRAPPA)

GRAPPA is a novel partially parallel acquisition method used to accelerate image acquisition using a radio frequency

(RF) field coil for spatial encoding. A detailed, highly accurate RF field map is not needed prior to reconstruction in GRAPPA. This information is obtained from several k-space lines which are acquired in addition to the normal image acquisition. This signal is called the auto-calibration signal (ACS).

IV. Robust, Subject-Invariant, Spatially-Regularized Dictionary Prior

A. Introduction

A Bayesian graphical reconstruction framework is employed that relies on learning data-adaptive prior models through dictionaries that are designed to be robust to large physiological fluctuations that are typical in R-fMRI signals. The robustness comes from modelling the prior to be heavy-tailed non-Gaussian distributions on signals.

B. Mathematical Framework

A Markov Random Field is used to model the complex valued 4D spatiotemporal R-fMRI image having V voxels and T time points. The BOLD-signal time series is modelled using a dictionary D with J atoms represented as a matrix where column $j \in [1, J]$ is the atom $d_j \in \mathbb{R}^T$.
A few constraints are required in the above setup, namely $\|d_j\|_2 = 1$ and $d_0 := 1$ to model shifts of the average value with the time series at voxel v. Each $\overline{x_v}$ is modeled (upto the time-series's average value) as a linear combination Da_v where $a_v \in \mathbb{R}^J$, of a *few* atoms in the dictionary.
In general, Pearson's correlation, which is used to characterize the key functional networks, is consistent across subjects who exhibit the same network. However, the possible variations in the underlying time-series are modeled using a similarity transform. This transform has been incorporated as:

- a constant atom d_0 for shifting
- coefficients a_v for scaling
- rotation is modeled using an orthonormal $T \times T$ matrix R for all *non-constant* atoms $d_j (j \geq 1)$ that model one subject

Therefore, a new dictionary is obtained which is used to model the magnitude time series corresponding to another subject as $RDa'_v + d_0 a'_{v0}$.
The R-fMRI signal is very weak and prone to a number of physiological fluctuations which are not captured by the diminishing tails which are characteristic of Gaussian distributions. Hence, a heavier-tailed distribution is realised by penalizing the p-th norm of the residual where $p \in (0, 2)$. Another penalization which is included is the q-the power of the q quasi-norm of each coefficient vector a_v where $q \in (0, 1)$. This enforces sparsity on a_v and prevents overfitting.
In addition to the above, spatial correlations present in R-fMRI series are also enforced by spatial depedency in a_v across the voxels.

C. Dictionary Learning Formulation

We first model the time series x_v at voxel v by an independent generalized multivariate complex Gaussian. Define A to be the image of the dictionary coefficient vectors $[a_{v0}, a_v]^T$. For a voxel v, let Φ_v be the transform incorporating phase. The expression for the multivariate Gaussian probability thus becomes:

$$P(x|D, A, R, \Phi) := \eta exp(-\sum_{v=1}^{V} \|x_v - \Phi_v(RDa_v + d_0 a_{v0}\|_2^p) \tag{1}$$

A maximum a posteriori (MAP) estimate is learned over N fully-sampled images to output the dictionary to be used for reconstruction. The estimation problem is solved using gradient ascent with adaptive step size. The atoms are initialized using k-means clustering with J clusters and the free parameters are tuned using cross validation.

V. k-t FASTER: Acceleration of functional MRI data acquisition using low rank constraints

A. Introduction

The paper explores the concept of utilizing the intrinsic spatiotemporal structure of MRI data to accelerate the process of acquisition as opposed to accelerating on a volume-by-volume basis. This technique exploits the low rank structure of fMRI data.

B. Mathematical Framework

The k-t FASTER technique relies on low-rank matrix completion strategies. In reality, the k-t matrix is only approximately low rank due to the presence of thermal and physiological noise.
The matrix completion (MC) problem can be directly compared to compressed sensing (CS). The underlying idea behind CS is that the data can be represented in a known basis using a limited number of coefficients which are non-zero. This can be thought of as enforcing sparsity to limit the possibilities of reconstruction. For the MC problem, instead of the idea of sparsity in some known basis, the idea of imposing a low rank on the data is employed.
The major difference between CS and MC is that CS requires a prespecified basis and computes the non-zero coefficients but MC has no such requirement. From the above formulation, we can now write:

$$X = U\Sigma V^T + N \tag{2}$$

where U is an $m \times r$ matrix of normalized spatial components, Σ is an $r \times r$ diagonal matrix consisting of weights and V is an $n \times r$ matrix of normalized temporal components. N is an $m \times n$ matrix which represents additive noise.
The first term is an outer product which can be thought of as the Singular Value Decomposition (SVD) of a matrix with m voxels and n time points. The SVD decomposition of a matrix is directly linked to its Principal Component Analysis (PCA), therefore, the above analysis establishes the dimesionality reduction aspect of the problem.
The reconstruction approach relies on the Iterative Hard Thresholding algorithm, which solves the constrained optimization problem

$$min\|Y - \Phi(X)\|_2 \tag{3}$$

under the constraints that

$$rank(X) = r \tag{4}$$

The above equations are stated for a more general setting, however in the current setup, X can be thought of the reconstructed image that is being estimated by the algorithm, Y can be thought of as the quantity that is being measure or acquired, which is the undersampled image in this case and Φ is the mapping between the two, which is the Fourier transform followed by the undersampling operator.
The problem is solved by the following recursive relation:

$$X^{n+1} = S_r(X^n + \mu(\Omega(X) - \Omega(X^n))) \tag{5}$$

In the above relation, X^n is the n^{th} estimate, μ is a step size parameter, Ω is a matrix subsampling operator and S_r is the hard thresholding operator that can be expressed as follows:

$$\sigma_i = \begin{cases} \sigma_i & \text{if } i \leq r \\ 0 & \text{otherwise} \end{cases}$$

where σ_i represents the i^{th} singular value of the matrix and r represents the fixed rank cutoff. An alternative relaxed approach which uses the shrinkage operator (this approach involves the substitution of the low rank constraint which is non-convex by a convex nuclear norm constraint [12]) can be stated as follows:

$$\sigma_i = \begin{cases} \sigma_i - \tau & \text{if } \sigma_i \geq \tau \\ 0 & \text{otherwise} \end{cases}$$

which τ, the shrinkage parameter shrinks the non-zero singular values by a fixed amount.
The convexity involved in the second method offers certain advantages, however, to also address the fact that the data has an intrinsic low rank structure, we combine both the approaches and obtain

$$\sigma_i = \begin{cases} \sigma_i - \tau & \text{if } i \leq r \\ 0 & \text{otherwise} \end{cases}$$

Notably, this reduces the original hard thresholding problem if $\tau = 0$.

C. Evaluation of k-t FASTER method

The k-t FASTER method was evaluated using two different types of experiments, namely the use of retrospectively undersampled data and the use of prospectively undersampled acquisition.
In the first method, the acquired images are taken to be the ground truth even though they may still have thermal and physiological noise. This data is Fourier transformed along the spatial dimensions and subsequently undersampled to create training data. Undersampling in the time dimension is also done. Lower frequencies are retained in the k-space as they contain more information and a sampling of about 23.8% is done.
In prospectively undersampled data, the reconstruction was performed on a coil-by-coil basis and then the sum of squares was combined. A rank constraint of 72 was used, as opposed to 26 in the retrospectively undersampled case. However, all other sampling and reconstruction parameters were same as that in the restroctively undersampled case.

D. Discussion and Conclusion

The error metric used for evaluating the reconstruction quantitatively is as follows:

$$err_F(\hat{X}) = 100 \times \frac{\|\hat{X} - x\|_F}{\|X\|_F} \tag{6}$$

where error is defined with respect to the Frobenius norm. A rank truncated version of the ground truth using PCA (Rank-128) was also used to compute the error against which this particular error was tested.
Over a list of 6 subjects, the two errors can be summarised as in the table below: As can be seen from the above table,

Subject	Rank-128	k-t FASTER
1	3.5	4.4
2	3.3	3.7
3	2.8	3.4
4	3.4	3.8
5	2.7	3.3
6	3.8	4.2

the values are reasonably close and validate the low rank approximation.
However Frobenius norm is not an effective method to capture the spatial regularities which are present in images, therefore correlation maps are also used as metric to evaluate the results which also confirms the impressive performance of this method.
It can be concluded that even though the success of CS-based acceleration strategies has motivated several strategies in the field of MRI imaging to exploit the low rank structure inherent to this data, the approach has still not been applied to functional neuroimaging which before the introduction of this paper, relied heavily on partial parallel image acquisition. To this effect, k-t FASTER presents a technique making use of the spatiotemporal structure of the data and uses the method of hard thresholding to obtain reconstruction which is found to be robust across the fMRI data of various subjects. However, it is an iterative method and can require several hours of computation time for large k-t matrices. The paper aims to resolve in the future, a better selection criteria for the hard thresholding of ranks as well as explore parallel computational techniques and GPU-based acceleration to improve computational efficiency.

VI. Accelerating Magnetic Resonance Imaging via Deep Learning

A. Introduction

This paper proposes one of the earliest applications of Deep Learning to the field of MRI [11]. A simple convolutional neural net structure is designed to map MRI images obtained from the zero filled reconstruction to the images obtained from the fully-sampled k-space data. The network is built in

a way such that it is also compatible with online constrained reconstruction methods.

B. Mathematical Framework

The raw k-space data which is acquired during the undersampled acquisition can be written as:

$$f = P\mathcal{F}u \tag{7}$$

Here P represents the undersampling transform and $\mathcal{F}$ denotes the normalized Fourier transform matrix (therefore $\mathcal{F}u$ represents the fully-sampled k-space data). Therefore, the zero-filled reconstruction can be written as:

$$z = F^H PFu \tag{8}$$

where H represents the Hermitian transpose operator and since F is the normalized Fourier transform matrix, $F^H F = I$, therefore, F^H acts as the inverse Fourier transform.
The paper now establishes a loss function over which to optimize the parameters of the neural network. This can be stated as follows:

$$L(\Theta) = \frac{1}{2T}\sum_{t=1}^{T} \|C(z_t;\Theta) - u_t\|_2^2 \tag{9}$$

where Θ is the parametrization of the neural network over the set of weights and C represents the mapping that is being learned by the neural network from the undersampled images to the fully-sampled images. A minor refinement is made to this objective function has dividing the image pairs into overlapping subimage pairs which leads to the loss function now being:

$$L(\Theta) = \frac{1}{2TN}\sum_{t=1}^{T}\sum_{n=1}^{N} \|C(z_{t,n};\Theta) - u_{t,n}\|_2^2 \tag{10}$$

This increases the robustness of the model by increasing the size of the training set.

C. Reconstruction Formulation

The parameters Θ can be learned by minimizing over the objective function as derived above. Now, the reconstruction problem can be established as an optimization problem which is stated as follows:

$$argmin_u \|C(\mathcal{F}^H f;\Theta) - u\|_2^2 + \lambda\|f - \mathcal{F}u\|_2^2 \tag{11}$$

This is a least squares problem which admits an analytical solution which can be finally expressed as follows:

$$\mathcal{F}u(k_x,k_y) = \begin{cases} S(k_x,k_y) & \text{if } (k_x,k_y) \notin \Omega \\ \frac{S(k_x,k_y)+\lambda S_0(k_x,k_y)}{1+\lambda} & \text{if } (k_x,k_y) \in \Omega \end{cases}$$

where Ω is the set where the samples are available.

D. Architectural Details

The network consists of three layers of convolution. The filter weights in each layer are initialized as random values from a Gaussian distribution with zero mean and a standard deviation of 0.001. The biases are initialized to 0.

E. Discussion and Conclusion

The paper is one of the earliest applications of deep learning methods to MRI images and the reconstruction task with respect to them. The neural network being considered in the paper is a convolutional model consisting of 3 layers. A brief theoretical framework is presented which is concluded using an analytical solution to obtain the reconstructed image from the undersampled one.
The results show that the proposed network performs well and restores the details and fine structures that are lost in the zero-filled image.

VII. Deep learning for undersampled MRI reconstruction

A. Introduction

This paper [13] explores the application of a convolutional network, specifically UNET, which is a popular architecture for image processing tasks. Uniform subsampling is used in the phase encoding direction (which consumes maximum time during acquisition). This allows the capture of high-resolution information, but still permits the image-folding problem dictated by the Poisson summation formula.
However, to deal with the localization uncertainty which arises as a result of image folding, a small percentage of low frequency k-space data is added.

B. Mathematical Framework

Undersampled MRI problem can be equivalently stated as an optimal reconstruction function $f : x \rightarrow y$, a mapping between undersampled k-space data (x) and the MRI image (y) which corresponds to the fully-sampled k-space data. Compressed sensing MRI, which makes of the prior on the data that enforces sparsity, formulates the problem as follows.

$$y = argmin_y \|x - \mathcal{S}\circ\mathcal{F}(y)\|_{l_2}^2 + \lambda\|\mathcal{T}(y)\|_{l_1} \tag{12}$$

where $\mathcal{F}$ denotes the normalized Fourier transform operator. $\mathcal{S}$ denotes the subsampling operator and $\mathcal{T}$ represents a transform that enforces sparsity prior on the reconstruction. is the regularization parameter controlling the trade-off between the residual norm and regularity. A popularity used example of a sparsity enforcing prior is the L1-norm of the gradient of the reconstruction.
The deep learning paradigm can be seen as opposite to the least squares minimization approach. Here, the network aims to learn a function $f : x \rightarrow y$ where it uses a significant amount of training data to learn the function as follows:

$$f = argmin_{f\in\mathbb{U}_{net}} \frac{1}{N}\sum_{i=1}^{N} \|f(x^i) - y^i\|^2 \tag{13}$$

A UNET can provide low dimensional latent space representation of an image and the higher resolution features can be recovered by concatenating feature maps from the earlier layers to the further layers. The reconstruction function can be viewed as an inverse mapping of the forward model which is subject to the constraint of MR images, which are assumed

to exist in a low dimension manifold.
Let y be an MRI image which is to be reconstructed, consisting on N^2 pixels. Consider a framework wherein this image has a fully-sampled k-space data x_{full}. The reconstruction of y from x_{full} can be written as:

$$y(n,m) = \sum_{a=1-N/2}^{N/2} \sum_{b=1-N/2}^{N/2} x_{full}(a,b) e^{2i\pi(an+bm)/N} \quad (14)$$

The convention is taken such that the frequency encoding is along the a-axis and the phase encoding is along the b-axis in the k-space data.
In the undersampled case, there is violation of the Nyquist criterion and sub-Nyquist sampling is used to speed up the time-consuming phase-encoding.
According to Poisson summation formula, the discrete Fourier transform of uniformly subsampled image which a subsampling factor of in the phase encoding will give us the following two-folded image:

$$y_{2-fold}(n,m) = y(n,m) + y(n,m+N/2) \quad (15)$$

The deep leaning needs to find an unfolding map $y_{2-fold} \rightarrow y$, so that the data can be acquired at a higher rate and the neural net reconstructs the data from the undersampled acquisition.
However, it is not possible to obtain the above mentioned unfolding map for two reasons. There is a possible construction of a case (presented in Figure 2 in the paper) wherein we consider two similar images y_1 and y_2 with small anomalies at different locations. It is clear that $\mathcal{F}(y_1)$ and $\mathcal{F}(y_2)$ are not the same. However, for uniformly subsampled data with a subsampling factor of 2, it is observed that:

$$\mathcal{P} \circ \mathcal{S} \circ \mathcal{F}(y_1) = \mathcal{P} \circ \mathcal{S} \circ \mathcal{F}(y_2) \quad (16)$$

In the above equation, $\mathcal{S}$ is the subsampling operator and P is the zero padding operator. Therefore, the acquired data for both the images is completely identical and the unfolding map will give the same output for both the inputs. Therefore, the location of the anomaly can not be determined by this formulation.
The above problem is what the paper refers to as the problem of separability. This problem can be addressed by adding a few low frequency data, which helps in capturing numerous anomalies that form a part of MRI scans and help in achieving separability in conditions like the above example where the anomalies differ only in their position. Therefore, the subsampling strategy is decided as follows: there is an undersampling by a factor of 4 (25% k-space data, 64 out of 256 lines) but an extra 4% data at lower frequencies (about 12 out of 256 lines).

C. Image Reconstruction Function

From the above reconstruction strategy, around 70% of the data has not been acquired and the first step in the reconstruction process is to fill in zeros in place of the unacquired data. One solution can be the minimum norm solution which can be stated as follows:

$$y_b = argmin_{ys.t.\mathcal{S}\circ\mathcal{F}(y)=x} \|y\|_{l^2} \quad (17)$$

This solution is equal to the zero filled solution (inverse Fourier Transform of the zero-filled k-space data), however, practically, this solution is not the correct one and therefore, undesirable in most cases.
The paper makes use of the UNET architecture to fill in the zeroes at the unacquired locations to generate the correct reconstruction. However, during this recovery, the acquired parts of the data may get distorted so they are manually fixed later to generate the correct reconstruction. This process is called as k-space correction and is represented by the function f_{corr}.
To above obtained result, an inverse Fourier transform is applied and the absolute value is given as output of the reconstruction process. This can be summarized by the following equation:

$$f = |\mathcal{F}^{-1}| \circ f_{corr} \circ \mathcal{F} \circ f_d \circ |\mathcal{F}^{-1}| \circ \mathcal{P} \quad (18)$$

where f_d is the trained UNET function which is learnt by standard training strategies.

D. Discussion and Conclusion

The proposed method suppresses the artifacts due to folding and also provides surprisingly sharp and natural looking images. If the UNET is used without the k-space correction, most of the folding artifacts are removed. However, if k-space correction is used, even the remaining folding artifacts are removed. The results have been summarized in the table below, which lists the mean values of MSE and SSIM for the three cases.

Method	**Mean MSE**	**Mean SSIM**
Aliased	0.0043	0.6516
UNET	0.0012	0.8782
UNET with f_{corr}	0.0004	0.9039

The above results have been taken from the paper. This demonstrates the use of the deep learning formulation to this problem and especially the effectiveness of k-space correction. The paper starts with the mathematical framework of the problem and proceeds to highlight the problem of separability. This is solved by using a specific subsampling strategy. The UNET architecture is used to learn the mapping from undersampled images to the reconstruction.
The experiments show that the learned function appears to have highly expressive representation capturing analytical geometry as well as small anomalies. The flexibility of the proposed model is tested by applying the learned model to the reconstruction of other similar images (CT images) that were never used in the training data. The competitive results on the untrained datasets suggested that the proposed model is fairly flexible.

VIII. Bayesian Deep Learning for Accelerated MR Image Reconstruction

A. Introduction

This paper explores the application of Bayesian deep learning techniques to model the uncertainty associated with deep learning-based reconstructions. MC-dropout and heteroscedastic loss is applied to model *epistemic* and *aleatoric* uncertainty. The main objective is to analyze the kind of errors that are produced by the deep learning networks. This kind of analysis is crucial for the deployment of these networks in clinical settings.

Taking inspiration from their application in MRI image quality transfer/super-resolution tasks, MC-dropout and heteroscedastic loss are used to capture model and data uncertainty respectively. These two techniques are primarily applied to two architectures, UNET [9] and a deep cascade of CNNs (DC-CNN) [10].

The paper shows that Bayesian Deep Learning methods are able to approximately characterize the confidence associated with the generated reconstructions. However, it has been pointed out that the proposed formulation seems overly simplistic and might need to be refined before being employed in a more practical setting.

B. Mathematical Framework

Consider a complex valued MRI image which is fully sampled. Denote this image by $x \in \mathbb{C}^n$. Correspondingly, consider $y \in \mathbb{C}^m$ which is given as follows:

$$y = \mathcal{F}_u x + \epsilon \tag{19}$$

where $\mathcal{F}_u$ is the undersampling Fourier operator and $\epsilon\ \mathcal{N}(0, \sigma^2 I)$.

The network aims to learn the inversion map $p(x|y)$ or $p(x|x_u)$ ($x_u = \mathcal{F}_u^H y$ is the zero filled reconstruction). A typical approach to solving this is to approach the problem using the concept of maximum a posteriori (MAP) estimate which can be written as:

$$argmax_x p(x|y) = armin_x - logp(y|x) - logp(x) \tag{20}$$

The above can now be solved as a minimisation problem, and the likelihood and prior terms correspond to data fidelity and regularization terms respectively. In this light, deep learning frameworks are viewed as an approximation to learning a MAP inference.

The deep learning framework basically learns an inversion function $f^w(x_u) \approx x$ where w is a parametrization over the weights of the neural network. The network parameters are learnt over the set of input dataset $\mathcal{D}$.

The major shortcoming of the above setup is that it provides a point estimate. For deep learning, there exists no theoretical framework that establishes a clear relationship between the number of measurements, validity of assumed prior and the reconstruction error. Therefore, it remains unknown when the network fails in the reconstruction task.

In this respect, it might be more useful to model the distribution $p(x|y)$ instead to obtain the confidence associated with the reconstruction (can be modeled with respect to the variance associated with the output). This motivates the Bayesian deep learning setup which is specified as follows.

For an undersampled image x_u and dataset $\mathcal{D}$, a predictive distribution over the reconstructed image is given as follows:

$$p(x|x_u, \mathcal{D}) = \int p(x|x_u, w) p(w|\mathcal{D}) dw \tag{21}$$

The important thing to note here is that $p(w|\mathcal{D})$ is in practice indeterminable and is therefore approximated by $q(w)$ (variational inference). The predictive distribution is computed using Monte Carlo integration as an analytical solution does not exist in most cases.

C. Types of Uncertainty

Given the current theoretical framework, two types of uncertainty can be identified. The first kind is the **aleatoric uncertainty**. This refers to the irreducible uncertainty which is observed in the data. In the current setup, this uncertainty can stem from measurement noise. However, there is also an inherently high level of ambiguity whether a pixel value represents an aliasing pattern, some anatomy or a texture. If such an uncertainty is encountered by the network in a specific region, the model should output low confidence in this region of the image.

The second kind of uncertainty is called **epistemic uncertainty**. This results from the fact that given a dataset D, there are more than one distinct sets of parameters w, that will map the undersampled images to the correct reconstructions. This causes a major issue as the only way to deal with this uncertainty is to increase the size of the dataset which, in the medical imaging domain, is often quite difficult. This makes it crucial to account for the variability caused in the network output by this uncertainty.

D. Modeling Uncertainties

Aleatoric uncertainty and epistemic uncertainty are modeled by incorporating *heteroscedastic loss* and *MC-dropout* respectively. The likelihood function is written as:

$$p(x|x_u, w) = \mathcal{N}(x|f^w(x_u), g^w(x_u)) \tag{22}$$

where $f^w(x_u)$ gives the mean prediction and $g^w(x_u)$ models the variance of the prediction, thereby accounting for the uncertainty found in the input. The pixel-wise error is weighted by the predicted inverse pixel variance.

MC-dropout is simply the application of dropout to the network activation maps. During inference, the predictive mean is given as the output.

$$\mathbb{E}[x] \approx \frac{1}{T} \sum_{t=1}^{T} f^{w_t}(x_u) \tag{23}$$

where w_t is the configuration of the network after dropout has been applied. Finally, the variance of each complex-valued pixel is given by the sum of the variance of real component and the variance of imaginary component.

E. Network Architectures

The two popular architectures for image processing tasks are UNET and DC-CNN. These architectures are used to implement the functions ${}^{w}(x_u)$ and ${}^{w}(x_u)$. Each of the functions can be parametrized independently, or using one network with multiple heads.
Two variants of the DC-CNN are considered:

- **DC-CNN1:** This takes the penultimate feature maps from each of the convolutional networks and feeds these into a 5-layer variance network which is used to obtain the function ${}^{w}(x_u)$.
- **DC-CNN2:** This trains a separate 5-layer variance network which takes the undersampled image as the input itself.

UNET uses a common encoder to train the functions $f^{w}(x_u)$ and $g^{w}(x_u)$, however, it used separate decoders to obtain both the functions.

F. Discussion and Conclusion

The Bayesian models perform competitively when the testing data is away from the training distribution. The uncertainty maps also showed a strong correlation with the error maps.
However, numerous shortcomings of the above model must also be addressed. The characterization of aleatoric uncertainty is heavily dependent on whether or not f and g are modeled to be correlated or not, a choice that is heavily influenced by architectural decisions. This decision must be made by the user based on the task.
Another shortcoming of the above model is are noticeable differences between the error maps and the uncertainty maps. This could possibly be attributed to the fact that MC-dropout is quite a simple technique and may not be able to represent the complex model uncertainty in the reconstruction task. However, Bayesian deep learning shows promising potential and can be improvised to a stage where it is suitable for clinical applications.

IX. Implementation and Analysis Work

The implementation work centered around the further application of deep learning techniques especially, convolutional neural networks to the problem of undersampled reconstruction to MRI images [14]. The network can be described in terms of a three stage CNN architecture:

- The first stage consists of a CNN which is used to reconstruct the k-space data which has not been acquired using the data present in spatial and temporal neighbourhood (exploits the spatiotemporal structure of the data)
- The second stage takes the data from the k-space to the image domain by the use of an inverse Fourier transform
- The thirst stage uses a CNN learned from quality enhancement in the spatiotemporal domain and refines the output features and details

The use of Bayesian Deep Learning is employed for uncertainty estimation and a loss function which is robust to large physiological fluctuations in these type of signals is used.

A. Architectural Details

The architecture as specified above has been divided into three stages:

- **Stage 1:** This consists of a CNN which takes as input the zero-filled k-space acquired data for all the timepoints for which an acquisition has been made. For these timepoints, the CNN fills in the values of the unacquired data points. The k-space data is divided into two halves. A convolutional neural net is applied individually to each of these halves with separate weights.
- **Stage 2:** The framework now takes the fully-sampled k-space images which are still undersampled in the time domain and takes the inverse Fourier transform of these images to obtain the images in the spatial domain. Now the images are temporally upsampled and this is done using linear interpolation so that it is computationally efficient.
- **Stage 3:** The third stage takes the complete time series and models a probability density function over this which is paramterized by a univariate Gaussian with real valued means and positive real-valued standard deviations. The mean is output as the reconstructed image and the standard deviation is used to generate the uncertainty map.

B. Loss Metrics

Three loss functions were proposed to evaluate the performance of the model. They have been summarized below.

- **Mean-Squared Error Loss:** This is the typical loss used in the optimization problem setup. However, in addition to computing the loss of the final output of the third stage against the ground truth images that we had for training samples, we also compute the loss between the inverse Fourier transform of the output of the first stage of the neural net for the time points for which we have samples available. A convex combination of these two terms was taken.
- **Robust Loss:** The mean-squared error loss arises from the assumption of a Gaussian distribution on the data. However, in the case of MRI images, the physiological noise leads to a more heavy-tailed distribution than the Gaussian, therefore, making use of the mean-squared error loss does not make the model robust to physiological noise. Therefore, the robust loss is used which penalizes the p^{th} power of the Frobenius norm where $p \leq 2$ is a free parameter that can be tuned.
- **Loss Based on Bayesian Modeling:** Assuming that the ground truth X is drawn from a Gaussian distribution with mean Ψ^{M} and standard deviation Psi^{S} (both of which are outputs of the neural network), the posterior probability density can be stated as:

$$P(X|Y) = \prod_{v=1}^{V} \prod_{t=1}^{T} G(X_{v,t}; \Psi^{M}_{v,t}, \Psi^{S}_{v,t}) \qquad (24)$$

The log likelihood of the above posterior probability can be maximized as an objective which will be equivalent to minimizing the Bayesian loss.

C. Results

The results were measured in the form of mSSIM (mean Structural Similarity), the graph for which has been presented below. The high values of mSSIM show that the architecture

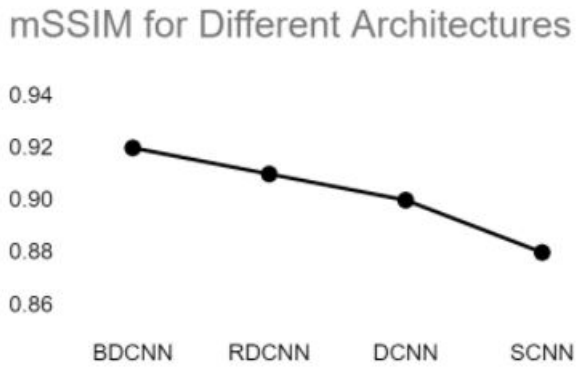

Fig. 1. Results for Different Loss Functions

works fairly well. Bayesian loss in particular performs significantly well and robust loss also performs competitively as they both capture different aspects of the task.

X. Future Work

The model can still be improved in terms of accuracy by further changes to the architecture (using deeper networks or making changes to activation or pooling functions using more intelligent alternatives).

Another strategy we have been looking at is the expectation maximization framework. Typically used in an unsupervised setting, it is used for the clustering task and reduces to the famous K-means clustering algorithm under special circumstances. However, it has also been shown to be used for reconstruction task in the medical image domain with success [15]. This will potentially form the basis of our work as we go ahead.

Acknowledgments

I would like to thank Prof. Suyash Awate, who has guided me throughout this research experience. From defining the problem statement, helping with the literature survey and finally, the implementation of a successful architecture, he has been an extremely helpful guide every step of the way.

I would also like to thank Prachi Kulkarni, who has assisted me with literature survey numerous times, helping me understand the context and explaining any conceptual details I struggled with. Finally, I would also like to thank Prof. Shabbir N. Merchant, who guided me about how to formally present research and also helped me every time I needed it.

References

[1] Megan H. Lee, Christopher D. Smyser, Joshua S. Shimony, *Resting state fMRI: A review of methods and clinical applications*, NIHMSID: NIHMS561036

[2] Martijn P. van den Heuvel and Olaf Sporns, *An Anatomical Substrate for Integration among Functional Networks in Human Cortex* , Journal of Neuroscience 4 September 2013

[3] Prachi H. Kulkarni, S. N. Merchant, Suyash P. Awate, *Bayesian Reconstruction of R-fMRI from k-t Undersampled Data using a Robust, Subject-Invariant, Spatially-Regularized Dictionary Prior*, 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)

[4] Mark A. Griswold, Peter M. Jakob, Robin M. Heidemann, Mathias Nittka, Vladimir Jellus, Jianmin Wang, Berthold Kiefer, Axel Haase, *Generalized autocalibrating partially parallel acquisitions (GRAPPA)*, Magnetic Resonance in Medicine, Volume47, Issue6, June 2002, Pages 1202-1210

[5] Mark Chiew, Stephen M. Smith, Peter J. Koopmans, Nadine N. Graedel, Thomas Blumensath, Karla L. Miller, *k-t FASTER: Acceleration of Functional MRI Data Acquisition Using Low Rank Constraints*, Magnetic Resonance in Medicine 74:353–364 (2014)

[6] Jo Schlemper, Daniel C. Castro, Wenjia Bai, Chen Qin, Ozan Oktay, Jinming Duan, Anthony N. Price, Jo Hajnal, Daniel Rueckert, *Bayesian Deep Learning for Accelerated MR Image Reconstruction*

[7] Raichle ME, MacLeod AM, Snyder AZ, Powers WJ, Gusnard DA, Shulman GL, *A default mode of brain function*, Proc Natl Acad Sci U S A. 2001 Jan 16; 98(2):676-82

[8] Power JD, Cohen AL, Nelson SM, Wig GS, Barnes KA, Church JA, Vogel AC, Laumann TO, Miezin FM, Schlaggar BL, Petersen SE, *Functional network organization of the human brain*, Neuron. 2011 Nov 17; 72(4):665-78

[9] Olaf Ronneberger, Philipp Fischer, Thomas Brox, *U-Net: Convolutional Networks for Biomedical Image Segmentation*, arXiv:1505.04597 [cs.CV]

[10] Schlemper, Caballero, Hajnal, Price, Rueckert, *A deep cascade of convolutional neural networks for dynamic MR image reconstruction*, IEEE Trans. Med. Imaging 37 (2017)

[11] Shanshan Wang, Zhenghang Su, Leslie Ying, Xi Peng, Shun Zhu, Feng Liang, Dagan Feng, Dong Liang, *Accelerating Magnetic Resonance Imaging via Deep Learning*, NIHMSID: NIHMS1018932

[12] Recht B, Fazel M, Parrilo PA.,*Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization*, Siam Rev. 2010;52:471–501

[13] Chang Min Hyun, Hwa Pyung Kim, Sung Min Lee, Sungchul Lee, Jin Keun Seo, *Deep learning for undersampled MRI reconstruction*, arXiv:1709.02576 [stat.ML]

[14] Karan Taneja, Prachi H. Kulkarni, S. N. Merchant, Suyash P. Awate, *A Bayesian Deep CNN Framework for Reconstructing k-t-Undersampled Resting-fMRI*, Unpublished Work

[15] J. M. Ollinger, *Maximum-likelihood reconstruction of transmission images in emission computed tomography via the EM algorithm*, in IEEE Transactions on Medical Imaging, vol. 13, no. 1, pp. 89-101, March 1994, doi: 10.1109/42.276147.