Channel Attention Networks for Robust MR Fingerprint Matching

Refik Soyak [ORCID], Ebru Navruz, Eda Ozgu Ersoy, Gastao Cruz [ORCID], Claudia Prieto [ORCID], Andrew P. King [ORCID], Devrim Unay [ORCID], *Member, IEEE*, and Ilkay Oksuz [ORCID]

Abstract—*Objective:* **Magnetic Resonance Fingerprinting (MRF) enables simultaneous mapping of multiple tissue parameters such as T1 and T2 relaxation times. The working principle of MRF relies on varying acquisition parameters pseudo-randomly, so that each tissue generates its unique signal evolution during scanning. Even though MRF provides faster scanning, it has disadvantages such as erroneous and slow generation of the corresponding parametric maps, which needs to be improved. Moreover, there is a need for explainable architectures for understanding the guiding signals to generate accurate parametric maps.** *Methods:* **In this paper, we addressed both of these shortcomings by proposing a novel neural network architecture (CONV-ICA) consisting of a channel-wise attention module and a fully convolutional network. Another contribution of this study is a new channel selection method: attention-based channel selection. Furthermore, the effect of patch size and temporal frames of MRF signal on channel reduction are analyzed by employing a channel-wise attention.** *Results:* **The proposed approach, evaluated over 3 simulated MRF signals, reduces error in the reconstruction of tissue parameters by 8.88% for T1 and 75.44% for T2 with respect to state-of-the-art methods.** *Conclusion:* **It is demonstrated that channel attention mechanism helps to focus on informative channels and fully convolutional network extracts spatial information achieve the best reconstruction performance.** *Significance:* **As a consequence of improvement in fast and accurate manner, presented work can contribute to make MRF appropriate for clinical use.**

Index Terms—**Channel attention, deep learning, MR fingerprinting, reconstruction.**

Ebru Navruz and Eda Ozgu Ersoy are with the Electrical and Electronics Engineering Department, Izmir University of Economics, Turkey.

Refik Soyak is with the Electrical and Electronics Engineering Department, Izmir University of Economics.

Devrim Unay is with the Electrical and Electronics Engineering, Department of Izmir Democracy University, Turkey.

Gastao Cruz, Claudia Prieto, and Andrew P. King are with the School of Biomedical Engineering and Imaging Sciences, King's College London, U.K.

Ilkay Oksuz is with the School of Biomedical Engineering and Imaging Sciences, King's College London, London WC2R 2LS, U.K., and also with the Computer Engineering Department, Istanbul Technical University, Istanbul 34467, Turkey (e-mail: ilkay.oksuz@kcl.ac.uk).

Digital Object Identifier 10.1109/TBME.2021.3116877

I. Introduction

MRI is an essential technique to visualize organs and structures inside the body with applications in basic and clinical sciences. While MRI is an adaptable and powerful tool for imaging, its time inefficiency limits its clinical use for quantitative imaging. In 2013, Ma *et al.* [1] introduced Magnetic Resonance Fingerprinting (MRF) as an alternative of quantitative MRI where signal equations for different types of MR acquisitions are used to estimate conventional MRI.

The key benefit of MRF is that it provides the opportunity to acquire and quantify different tissue parameters such as the longitudinal relaxation time (T1) and the transverse relaxation time (T2) within a single acquisition, and thus eliminating the need for multiple acquisitions. Acquisition parameters are varied pseudo-randomly so that each tissue generates a unique signal evolution or fingerprint. Mapping generation is conventionally carried out by applying a dictionary (template) matching algorithm where the acquired signal is matched with the dictionary signal. However, this technique has some shortcomings that need to be addressed such as the computational time required for the dictionary matching algorithm. In dictionary matching, each acquired signal needs to be compared with the simulated signals resulting in high computational complexity. As the number of combinations in the dictionary increases, reconstruction becomes more expensive with regard to time and storage [2]. In addition and perhaps more importantly, MRF needs to make a trade-off between accuracy and scanning time, and therefore relies on high under-sampling in k-space. As a consequence, dictionary matching may lead to erroneous quantification in the generated maps [3].

II. Related Works

In this section we provide an overview of the previous studies on dictionary matching in MRF, and the neural network techniques that aim to accelerate this process.

A. Dictionary Matching

In the original proposal of MRF [1], a dictionary matching algorithm is used to reconstruct the corresponding tissue parameters. Pattern recognition was employed to map the T1 and T2 tissue parameters by matching the fingerprints with a pre-defined dictionary of predicted signal evolutions. Later, different dictionary matching algorithms were proposed to obtain

faster and more accurate results. McGivney *et al.* [4] compressed the dictionary by using singular value decomposition (SVD), enabling faster computation due to the smaller dictionary size. Gomez *et al.* [5] analyzed the spatio-temporal dictionary and compared it with the temporal MRF dictionary. Besides, some studies suggested to use iterative techniques. For instance, Cline *et al.* [6] proposed a technique called AIR MRF that combines dictionary compression and regularization for accelerated matching. Additionally, Zhao *et al.* [7] proposed a statistical approach that uses maximum likelihood estimation to predict T1 and T2 tissue parameters.

B. Neural Network based Techniques

Successful applications of neural networks for computer vision motivated medical imaging community to use them for accelerating image reconstruction and mapping. Architectures such as Fully-Connected Neural Networks, Convolutional Neural Networks (CNN), and Long Short-Term Memory (LSTM) have demonstrated state-of-the-art performance for dictionary matching. Cohen *et al.* [9] proposed a 4-layer fully connected neural network model with two hidden layers by using magnitude images as input. Chen *et al.* [10] presented a fully connected model with 4 layers that performs dimension reduction with Principal Component Analysis (PCA) at its first layer. Oksuz *et al.* [11] used Recurrent Neural Network (RNN) to extract temporal frames for the prediction of T1 and T2 values relying on the time series nature of the MRF fingerprint. Likewise, Hoppe *et al.* [12] compared CNN and RNN architectures with magnitude and complex-valued inputs, and suggested to use the RNN model with complex-valued MRF signal.

Balsiger *et al.* [13] analyzed both spatial and temporal frames to reconstruct tissue parameters using a CNN architecture. Cao *et al.* [14] applied a multi-layer perceptron with 4 hidden layers and optimized it to prevent over-fitting. To further increase the speed of training and testing, architectures with pre-processing and pre-training for feature extraction and dimension reduction have been proposed. Since the high dimensionality of the MRF signal requires more computational power to process and creates more redundancy, one of the main focuses of the literature is to reduce signal dimensionality. Thus, a fully connected neural network based feature extraction (prior to U-Net) is proposed by Fang *et al.* [15] to extract important information while reducing the number of channels. Also, Fang *et al.* [16] proposed a U-Net like architecture with residual channel attention blocks to be able to extract more informative features. PCA followed by a fully connected convolutional architecture is used by Chen *et al.* [10]. In order to learn the non-linear relationship between the spatio-temporal MRF image data and multiple quantitative maps, Pirkl *et al.* [17] proposed to use a CNN architecture combined with relaxation and diffusion-sensitized MRF sequence.

In this study, we propose a channel attention-based CNN architecture to weight the important channels before feeding into the CNN architecture. As our proposed method is attention-based, it uses all channels in a weighted fashion instead of

reducing the channels, and therefore eliminates loss of temporal frames due to the channel reduction. Furthermore, the use of the channel attention mechanism allows us to examine the relative importance of the channels.

Accordingly, there are two major contributions of this work:
- To the authors' knowledge, this is the first paper that provides a thorough analysis of attention based methods for MRF.
- We propose novel CONV-ICA model that consists of Input Channel Attention (ICA) and a CNN architecture.
- An in-depth analysis of attention-based channel selections for understanding the significance of each signal in parametric map.

III. METHOD

In this section, we will provide the details of our proposed architecture to estimate the MRF parameters. The proposed method consists of a CNN-based architecture with the channel-wise attention module to empower MRF reconstruction capability. MRF consists of sequentially acquired signals over time that is 2D images + time, and we refer to each temporal frames as a channel, and feed into model accordingly. In channel attention networks, channels are weighted based on their significance so that the model can understand which channels of the MRF signal are more informative to generate the parametric maps.

A. Channel-Wise Attention Module

The suggested channel attention module consists of two pooling layers in parallel, shared fully connected layers, and a sigmoid activation function. The fully connected layer applies the sum over multiplication between each input and each weights to produce output. The output is then passed through the activation function and the attention scores are produced.

P_{max}^c and P_{avg}^c denote max-pooled and average-pooled features per channel C, respectively. These two pooled layers are fed into shared fully connected layers as shown in Fig. 1, so that input (I) becomes $\mathbb{R}^{1 \times 1 \times C}$. In the first fully connected layer, channel size is reduced to $\mathbb{R}^{1 \times 1 \times C/r}$. The *reduction parameter* r, empirically fixed at 50, is a hyper-parameter that controls the reduction. In the next fully connected layer, filter size is up-scaled to the original size $\mathbb{R}^{1 \times 1 \times C}$. The outputs of the shared fully connected layers are summed up and fed into a sigmoid activation function. After the channel attention scores - also known as attention maps - are produced, element-wise multiplication is performed to weight the input I. The formulation of the channel attention is as follows:

$$\alpha = \sigma(fullyconnected(P_{max}^c(I)) + fullyconnected(P_{avg}^c(I))) \qquad (1)$$

where α denotes attention scores and σ refers to the sigmoid activation function. The fully connected layers are followed by a Rectified Linear Unit (ReLu) activation function.
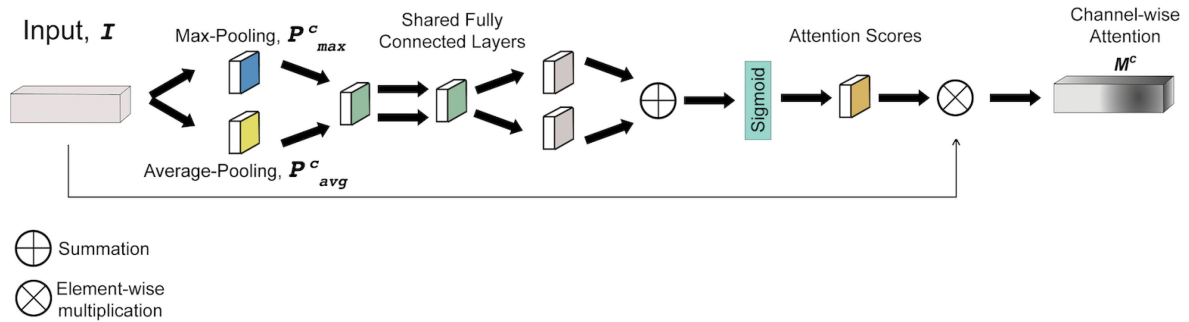
Fig. 1. Channel Attention Module Architecture [8] where weighted input data is produced with attention scores. P^c_{max} denotes max-pooled features, P^c_{avg} denotes average-pooled features and M^c denotes weighted input data.
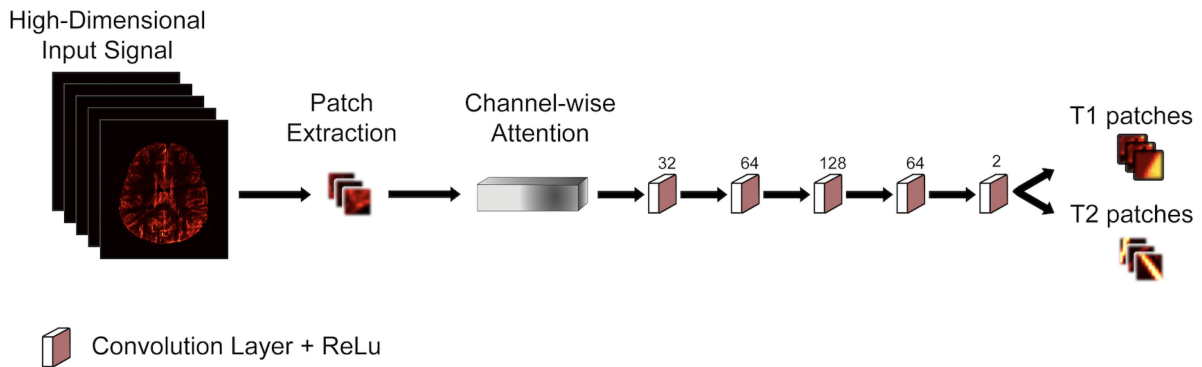

Fig. 2. Our proposed CONV-ICA model. The model accepts and produces 4×4 overlapped patches. The convolution layers that follows the channel attention module have filter sizes of $32, 64, 128, 64$ and kernel size of 3. Output convolutional layer has kernel size of 1 with filter size 2, i.e. the number of tissue parameters. All layers are followed by ReLu activation function.

$$M^c = I \otimes \alpha \qquad (2)$$

where M^c is the output of attention weighted input I. The input channels are weighted by element-wise multiplication ($\otimes$) of input I and attention scores α.

B. Proposed Model: $CONV - ICA$

In order to calculate the T1 and T2 maps we propose to use a convolutional network model on ICA. The current literature for reconstruction of medical images in MRF relies on reducing the number of temporal frames which are treated as channels in the network right after the input layer by using a fully connected layer or convolution layers. However, this reduction causes loss of temporal frames. To overcome this problem, this study suggests the use of a channel attention module before reducing channel size (*also known as filter*), as channel attention helps to extract important temporal frames by weighting the channels and prevents the loss of that information. As illustrated in Fig. 2, our proposed model consists of a channel attention module and four convolution layers with filter sizes of 32, 64, 128, and 64, respectively. The output layer is a 2D convolutional layer with a kernel size of 1×1 and a filter size of 2 as it predicts tissue parameters T1 and T2. All layers are followed by the ReLu activation function. The proposed model predicts overlapped patches as shown in Fig. 2. The use of overlapped patches is suggested by

Gomez *et al.* [18] as overlapped patch-wise approach allows to remove the under-sampling noise and incoherent artefacts by averaging over patches.

C. Channel Analysis & Attention-based Channel Selection

In MRF, the data consists of many channels, which is 2000 in this study. For a faster map reconstruction, less informative channels could be removed. Also, if some channels are redundant, channel reduction might improve map reconstruction performance. Accordingly, analysis of these channels is required to investigate which channels are most informative and important for map reconstruction and to eliminate the redundant ones. For this purpose, Balsiger *et al.* [13] zero-filled the channels one-by-one and compared the results to understand the importance of the zero-filled channel. However, such analysis may not be the best way to find the importance of channels as the zero-filled channel could be learned by the network as informative for reconstruction. To eliminate this issue, we propose to use an attention mechanism that will attempt to intrinsically assign attention and score the features of a given feature map by importance. In order to investigate the importance of each channels by benefiting from the advantage of the attention mechanism as explained, we used the channel attention module of Woo *et al.* [8] as shown in
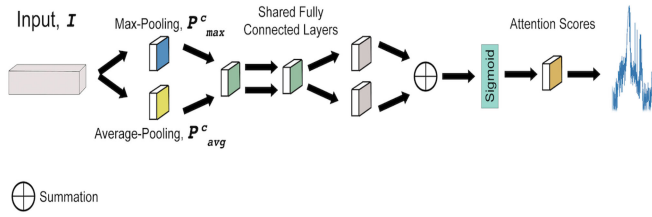
Fig. 3. Architecture to produce the attention scores for each channel. Pooling sizes for Max Pooling and Average Pooling are determined as 4×4 which is input patch size. The units of shared fully connected layers are controlled by reduction parameter r and it is altered to find optimum channel size. After that, in the next fully connected layer, channel size is up-warded to the input channel length, and attention scores are produced through sigmoid function.
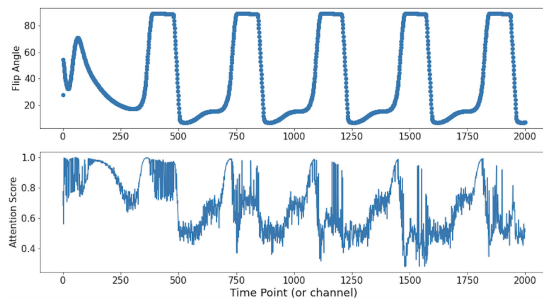

Fig. 4. The first plot shows the change of flip angle, where Echo Time (TE) and Repetition Time (TR) were fixed during acquisition, over the 2000 channels. The second plot is the distribution of corresponding attention scores over the 2000 channels obtained from Channel Attention Module, which allows to observe the importance of channels respectively.

Fig. 1. The feature maps are squeezed through pooling layers and each channel gets scored *(or weighted)* as the most informative channels get higher scores (closer to 1) to reconstruct parametric maps.

Attention scores are produced after the sigmoid activation function as shown in Fig. 3, and discussed in Section II-A. When attention scores are obtained for each patch, scores of all patches are averaged within each individual channel in order to obtain their distribution over the whole 2000 channels independently. For attention-based channel selection, *n channels* with the highest attention scores are selected. Following the channel selection procedure, the model is retrained.

The informative channels for tissue parameter estimation could be selected through the channel analysis proposed here ($attention - based~selection$), in contrast to the conventional approaches like PCA as used in [10] and random selection. We hypothesize that selection of the most important *n channels* by analyzing the distribution of attention scores over all channels (Fig. 4) is a better way for channel reduction and elimination of redundant channels.

As a result, in order to evaluate the performance of the proposed channel selection method empirically, we experimented with two conventional channel selection schemes, PCA-based and random, and carried out a performance comparison with our proposed method.

IV. EXPERIMENTAL RESULTS

We provide details of our experimental setup and results of our proposed model in comparison to state-of-the-art methods in this section. Methods of comparisons are trained and tested by using 3-fold cross-validation. As all methods are evaluated over 3 subjects, cross-validation is applied by selecting 1 subject for testing, and the other 2 for training, and continued that process until each subject is selected for testing. Mean absolute error (MAE) in percentage (Eq. 3) is used as the evaluation metric:

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^{N} \frac{\left| P_i - \hat{P}_i \right|}{max(P_i)} \times 100 \qquad (3)$$

where N represents the total number of pixels, $\hat{P}_i$ refers to the estimated value and P_i represents the actual value.

The proposed model is fed by patch-wise input with $4 \times 4 \times \tilde{}channels$ sized patches which were extracted from the MRF signals. During the training, 85498 patches were split as 30% for validation corresponding to 25649 patches while the remaining 70% or 59849 patches for training. ADAM optimization with a learning rate of 15×10^{-4} was used while training by minimizing the mean squared error loss function. Batch size was experimentally selected as 512 and the model was trained for 100 epochs. We stopped training early if no improvement of validation loss is observed for 15 consecutive epochs. The Keras deep learning library with TensorFlow back-end was used in the Google Colaboratory environment with 25 GB RAM of GPU.

A. Data

The model is trained and tested on a synthetically generated complex-valued MRF signal which consist of one slice of 2D images with 2000 temporal frames. Fingerprints were simulated based on previously acquired T1, T2, M0 (ground truths) and coil sensitivity maps. The acquisition used Bloch simulations to synthesize all the MRF contrasts -for every TR- and included multi-coil and radial undersampling. The simulated data was reconstructed with a Low Rank Inversion [19]. Extended Phase Graph (EPG) [20] is used to generate the MRF dictionary for a range of T1 = [0: 2: 500] [500: 5: 1000] [1000: 10: 2000] [2000: 50 : 4000] ms, and T2 = [0: 1: 100] [100: 2: 500] ms. Relevant scan parameters, namely balanced steady-state free precession radial sequence, TE, fixed TR, FOV, in-plane resolution, slice thickness, bandwidth are selected as described in [21]. Only 1 radial spoke was acquired at each time point resulting in an acceleration factor of 251 with respect to a fully sampled radial acquisition. A total of 2000 time points were acquired in approximately 10s.

B. Model Comparison

In this paper, mainly, state-of-the-art methods in tissue parameter reconstruction from MRF signals are selected according to the assumptions they make on the data as explained in Section II-B.

Table I shows that proposed approach achieves the best performance for both T1 and T2 parameters as compared to the

TABLE I
COMPARISONS OF PROPOSED MODEL WITH A VARIETY OF AVAILABLE
TECHNIQUES WHERE ERRORS ARE CALCULATED BETWEEN GROUND TRUTH
IMAGES AND BACKGROUND-MASKED RECONSTRUCTED IMAGES AS SHOWN
IN FIG. 5 BY USING MAE IN PERCENTAGE. *INDICATES THAT SVD IS USED
FOR CHANNEL REDUCTION FROM 2000 TO 100 IN PRE-PROCESSING FOR
DICTIONARY MATCHING. CONV2D ONLY REPRESENTS A MODEL
WITHOUT INPUT CHANNEL ATTENTION

Models	T1 - MAE(%)	T2 - MAE(%)
Dictionary Matching	4	1.01
Dictionary Matching*	4.18	1.01
Hoppe et al. [12] (RNN)	5.61	1.79
Chen et al. [17]	3.38	1.3
Hoppe et al. [23] (CNN)	2.07	1.03
CONV2D only	2.03	0.8
Balsiger et al. [13]	1.82	1
Fang et al. [15]	1.72	1.14
CONV-ICA	**1.58**	**0.57**

TABLE II
COMPARISON OF NUMBER OF TRAINABLE PARAMETERS (PARAMS) IN
MILLION, TRAINING TIME IN MINUTES AND TEST TIME IN SECONDS
BETWEEN THE PROPOSED MODEL AND VARIOUS AVAILABLE TECHNIQUES
AS WELL AS THE DICTIONARY MATCHING METHOD. *INDICATES THAT SVD
IS USED FOR CHANNEL REDUCTION FROM 2000 TO 100 IN
PRE-PROCESSING FOR DICTIONARY MATCHING

Model	Params (mn)	Training Time (min)	Test Time (sec)
Dictionary Matching	-	-	16392
Dictionary Matching*	-	-	6320
Hoppe et al. [12] (RNN)	16.17	15	43.61
Chen et al. [17]	0.41	1.95	7.44
Hoppe et al. [23] (CNN)	0.22	4	1.45
Balsiger et al. [13]	9.74	61.66	145.32
Fang et al. [15]	1.97	58.33	18.58
CONV-ICA	2.74	30	21.81

TABLE III
COMPARISON OF VARIOUS PATCH SIZES AND RECONSTRUCTION MAE
VALUES (EQ. 3) IN PERCENTAGE RESPECTIVELY. BECAUSE OF LACK OF
RESOURCES, CHANNELS WERE DECREASED TO INCREASE PATCH SIZE.
AFTER ATTENTION-BASED CHANNEL SELECTION IS PERFORMED WHERE
CHANNEL SIZE IS REDUCED FROM 2000 TO 40, PATCH SIZES ARE
INCREASED AND RECONSTRUCTION ERRORS ARE OBSERVED

Patch Size	T1 - MAE (%)	T2 - MAE (%)
4×4	2.67	1.05
8×8	**2.46**	**0.99**
12×12	2.63	1.05
16×16	2.88	1.12
24×24	3.02	1.19

TABLE IV
THE COMPARISON OF PIXEL-WISE (1D) AND PATCH-WISE (2D)
APPROACHES FOR PROPOSED MODEL. THE RECONSTRUCTIONS OF
MODELS ARE BACKGROUND-MASKED, AND ERRORS ARE CALCULATED
ACCORDINGLY
USING MAE IN PERCENTAGE

Model Type	T1 - MAE(%)	T2 - MAE(%)
CONV-ICA (1D)	2.34	0.99
CONV-ICA (2D)	**1.58**	**0.57**

TABLE V
COMPARISON OF CROSS-VALIDATED MAE IN PERCENTAGE FOR EACH
CHANNEL SELECTION METHOD, ATTENTION-BASED SELECTION, REDUCTION
BY PCA AND RANDOM CHANNEL SELECTION WHEN n NUMBER OF
CHANNELS ARE SELECTED FOR BOTH T1 AND T2 VALUES

Method / # channels	T1 - MAE(%)	T2 - MAE(%)
Attention / 10	**2.71**	1.02
PCA / 10	4.81	1.58
Random / 10	3.29	**1.01**
Attention / 50	**2.19**	0.92
PCA / 50	8.36	2.64
Random / 50	2.34	**0.78**
Attention / 100	**2.07**	**0.79**
PCA / 100	6.22	1.48
Random / 100	2.26	0.84
Attention / 200	**1.88**	**0.68**
PCA / 200	4.49	1.71
Random / 200	2.08	0.69
Attention / 300	**1.93**	**0.77**
PCA / 300	4.85	1.65
Random / 300	2.08	0.78
Attention / 500	**1.98**	0.81
PCA / 500	4.63	1.49
Random / 500	2.11	**0.77**
Attention / 1000	**2.00**	0.79
PCA / 1000	4.42	1.48
Random / 1000	2.02	**0.78**

state-of-the-art models. The results suggest that, using a channel attention mechanism to produce weighted channels leads to a decrease in the reconstruction error for our data set as weighted channels help to avoid losing the important temporal frames while reducing the number of channels. Because of under-sampling in k-space during MRF acquisition, the dictionary matching algorithm produces erroneous reconstructions.

Table I also shows the effect of the ICA. The proposed CONV-ICA model is compared with the CONV2D model without ICA and the model is represented as *CONV2D only* in Table I.

Also as shown in Table II, when comparing the number of trainable parameters, training time and prediction time according to the models, it can be seen that our study is faster in training time compared to studies of Fang *et al.* [15] and Balsiger *et al.* [13] and faster in prediction time compared to studies of Hoppe *et al.* [12], Fang *et al.* [15] as well as dictionary matching methods.

C. Effect of Patch Size

In this section, various patch sizes are tested in order to examine the effect of patch size on reconstruction capability. Table III shows that the use of 8×8 as the patch size achieves the lowest reconstruction error values for both T1 and T2 tissue parameters. Patch size 12×12 performs the second-best reconstruction for both T1 and T2 value. As a result, it is observed that patch size 8×8 is optimal for the lowest reconstruction error. However, because of insufficient computational resources, patch size was selected as 4×4 in this study. For investigation of the effect of patch size, attention-based channel selection is applied to reduce channel size from 2000 to 40 by selecting the best

40 channels with attention-based channel selection explained in Section III-C. Then, patches are extracted from data with reduced channels, and trained with the proposed model.

In addition, we compare the pixel-wise version (1D) of proposed model with its initial 2D counterpart as shown in Table IV. Because the proposed model is structured for 2D data, it has to be modified such that 1D convolutional layers are used instead of 2D convolutional layers to test 1D input.

D. Effect of Channel Selection

In this section, the aim is to evaluate the proposed attention weights with an additional experiment. Fig. 6 shows the true images, reconstructed images and error images between the ground truth and reconstructed image. Respectively, Table V shows MAE in percentage as seen in Eq. 3 for n number of channels selected by the methods: Attention-based, PCA, and Random. It is observed in Table V that attention-based channel
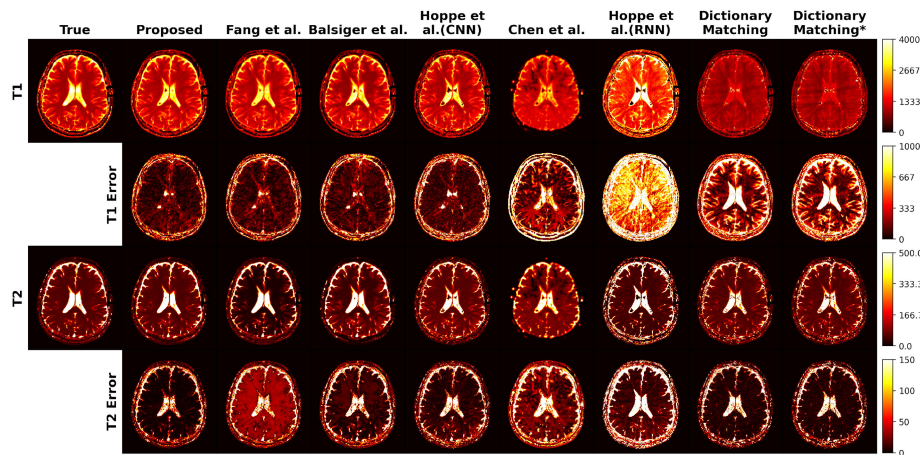
Fig. 5. Proposed model to state-of-the-art methods compared qualitatively by testing on a test subject. T1 and T2 errors are the mean absolute differences as shown in Eq. (3) between reconstructed parameters and true parameters in ms. Backgrounds of reconstructed images are masked to highlight interested region and errors in Table I are calculated accordingly. *Indicates that SVD was used for channel reduction from 2000 to 100 in pre-processing for dictionary matching.

TABLE VI
COMPARISON OF TEST TIME AND TRAINING TIME PER EPOCH IN SECOND
FOR DIFFERENT CHANNEL SIZES N. IT IS OBSERVED THAT TESTING AND
TRAINING TAKE MORE TIME AS N INCREASES

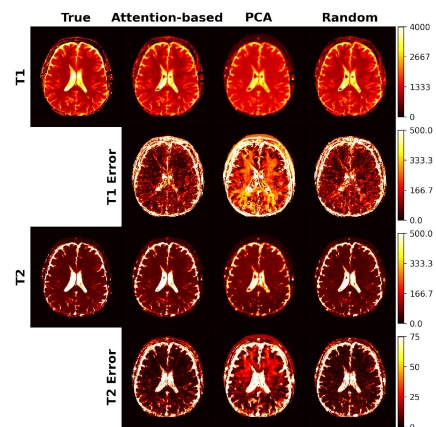| Channel size (n)	Test time (sec)	Training Time/epoch (sec)
10	5.51	3.28
50	5.93	3.29
100	8.64	5.43
200	10.27	5.44
300	10.87	5.46
500	13.44	6.55
1000	20.69	15.12
1500	24.98	26.22

Fig. 6. Channel selection methods are compared with true tissue parameters, T1 and T2 respectively, when channel size is reduced from 2000 to 100. The rows marked as T1 and T2 (1st and 3rd rows, respectively), show the reconstructed tissue parameters by the models. T1 error and T2 error images (2nd and 4th rows, respectively) show the differences between the reconstructed and true tissue parameters.

reduction results in a decrease in the reconstruction error while PCA-based and random selection schemes cause to increase the reconstruction error. The important advantage of attention-based channel selection is reducing the need of resources and run time for algorithms.

Table VI shows that test time and train training time increases significantly as channel size n increases. For instance, as shown in Table V, while selecting 200 channels with attention-based channel selection, T1 reconstruction error in MAE percentage decreases 6%, test time decreases 58.89% and training time decreases 79.25% with respect to selecting 2000 channels.

E. Segmentation Comparison

Table VII quantitatively, Figs. 7–8 qualitatively express comparison of the proposed model and state-of-the-art models for different brain tissues, namely Skull Stripped, Gray Matter (GM), White Matter (WM), and Cerebrospinal Fluid (CSF) that are extracted by automated segmentation. For segmentation, the Statistical Parametric Mapping (SPM12) [23] is used and errors due to automated segmentation were ignored. Results demonstrate that the proposed method achieved the best reconstruction performance for both parametric maps in all the brain tissues.

The qualitative results presented in Fig. 5 visually demonstrate the reconstruction performances of the proposed model

and the models in comparison for the T1 and T2 parameters. The rows marked as T1 and T2 (1st and 3rd rows, respectively), show the reconstructed tissue parameters by the models. T1 error and T2 error images (2nd and 4th rows, respectively) show the differences between the reconstructed and the true tissue parameters. These qualitative results further support that the proposed model achieves the best performance visually for the reconstruction of T1 and T2 tissue parameters.

V. DISCUSSION

We proposed a new deep learning architecture to address the drawbacks of the dictionary matching algorithm whilst accelerating and improving reconstruction of T1 and T2 values. The proposed method takes advantage of the channel attention mechanism which focuses on important temporal frames. Therefore, channel attention mechanism at the beginning of the model

TABLE VII

PROPOSED MODEL AND STATE-OF-THE-ART METHODS COMPARED QUANTITATIVELY FOR SEGMENTED BRAIN TISSUES, SUCH AS SKULL STRIPPED, GRAY MATTER (GM), WHITE MATTER (WM), AND CEREBROSPINAL FLUID (CSF). MAE IN PERCENTAGE (EQ. 3) ARE CALCULATED BETWEEN SEGMENTATION OF GROUND TRUTH AND RECONSTRUCTED IMAGE FOR BOTH T1 AND T2 TISSUE PARAMETERS. *INDICATES THAT SVD-BASED CHANNEL REDUCTION FROM 2000 TO 100

| Models	SKULL STRIPPED		GM		WM		CSF	
	T1 - MAE(%)	T2 - MAE(%)	T1 - MAE(%)	T2 - MAE(%)	T1 - MAE(%)	T2 - MAE(%)	T1 - MAE(%)	T2 - MAE(%)
Dictionary Matching	2.80	0.68	0.99	0.06	0.34	0.03	2.20	0.57
Dictionary Matching*	2.82	0.74	0.98	0.06	0.35	0.03	2.23	0.59
Hoppe et al. [12] (RNN)	3.89	1.51	2.03	0.34	1.17	0.18	2.96	1.16
Chen et al. [17]	1.57	0.94	0.62	0.18	0.33	0.08	1.19	0.75
Hoppe et al. [23] (CNN)	0.98	0.74	0.38	0.08	0.22	0.04	0.77	0.59
Balsiger et al [13]	0.83	0.73	0.40	0.21	0.21	0.15	0.66	0.54
Fang et al. [15]	0.76	0.84	0.35	0.38	0.19	0.23	0.59	0.62
CONV-ICA	**0.68**	**0.37**	**0.31**	**0.04**	**0.15**	**0.01**	**0.54**	**0.29**
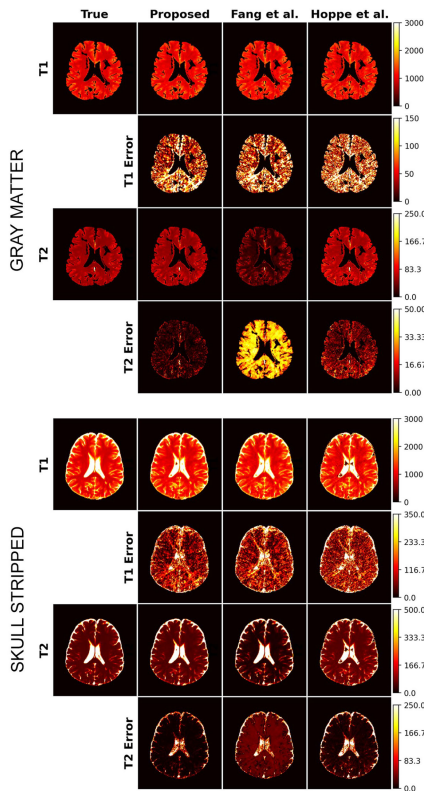
Fig. 7. Example results for ground truth (true) image, and reconstructions of proposed, Fang *et al.* [15] and Hoppe *et al.* [22] architectures. Two blocks show the reconstruction capabilities for Gray Matter and Skull Stripped respectively. Both T1 and T2 tissue parameters and error maps are visualized for each architecture.
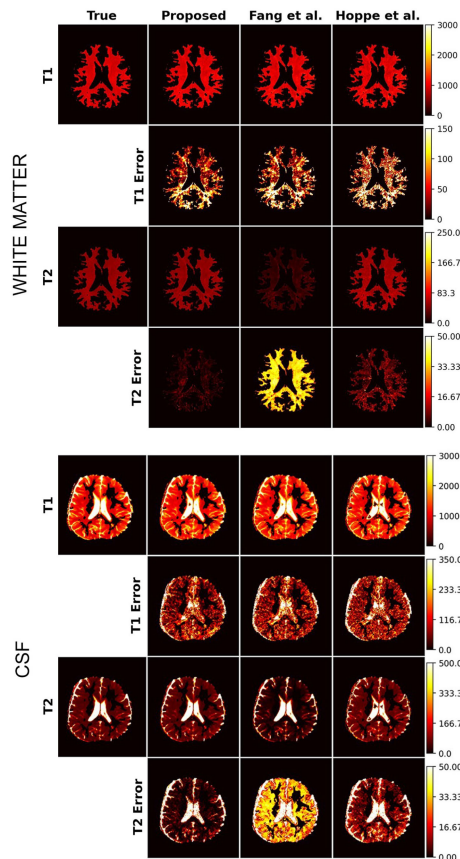

Fig. 8. Example results for ground truth (true) image, and reconstructions of proposed, Fang *et al.* [15] and Hoppe *et al.* [22] architectures. Two blocks show the reconstruction capabilities for White Matter and Cerebrospinal Fluid (CSF) respectively.

helps to eliminate the loss of temporal frames, while fully convolutional network extracts the spatial information from patchwise input to reconstruct from MRF signal more accurately than the previously proposed methods. Qualitative and quantitative results demonstrate that the use of the channel attention mechanism at the beginning of the model, where temporal frames is mostly lost due to direct channel reduction (e.g. from 2000 to 40), increases reconstruction capability by 8.88% for T1 value and 75.44% for T2 value.

In a separate setup we suggest a new channel selection method: attention-based selection in Section IV-D. As MRF data consists of many channels and thus requires high computational resources to process, selection of the most informative channels is needed to decrease the demand of resources such as RAM.

In Table V, for bSSFP-MRF sequence used in this study, it has been observed that approximately 200 channels are sufficient to significantly reduce the need for resources, while maintaining the reconstruction capability. Besides, it helps to accelerate runtime and computational time as data becomes smaller. Additionally and more importantly, during the acquisition of the MRF signal, the parameters are continuously varied to get unique signal evaluations, which may create redundant temporal frames [11]. Therefore, channel selection in MRF data is crucial to keep the most informative and the least redundant temporal frames. Results shows that attention-based selection helps for the best reconstruction among other channel selection methods

such as PCA and random selection for the T1 and T2 tissue parameters. As attention-based channel selection method is deep learning-based, increasing the amount of data will lead to better generalization of channel selection in a more reliable manner.

We studied and proposed new deep learning-based techniques, which needs larger dataset to generalize the solution. One limitation of this study is the lack of in-vivo data, and its validation. Accordingly, we plan to work on bigger and more heterogeneous datasets with different pathologies for different tissues including in-vivo data. In future, the proposed input channel attention mechanism could be adapted to the U-Net architecture [24] which has shown promising reconstruction capabilities recently [25] and other powerful architectures for MRF tissue reconstruction. MRF signal is originally a complex-valued data comprising real and imaginary parts. While the magnitude of the MRF signal is exploited solely in this study (similar to most of the state-of-the-art), in the future it is worth exploring the combined use of real and imaginary parts of the signal as some studies suggest [12], [26], [27].

In conclusion, this study shows that employing channel reduction at the beginning of the model causes loss of temporal frames which is important for accurate reconstruction of T1 and T2 parameters. To overcome this problem, we proposed deep learning-based input channel attention that can be easily applied to any model. Additionally, we analyzed the effect of patch size on reconstruction performance, and furthermore demonstrated quantitatively and qualitatively that the proposed attention-based channel selection achieves the best reconstruction performance.

REFERENCES

[1] D. Ma *et al.*, "Magnetic resonance fingerprinting," *Nature*, vol. 495, no. 7440, pp. 187–192, 2013.

[2] E. Hoppe *et al.*, "RinQ fingerprinting: Recurrence-informed quantile networks for magnetic resonance fingerprinting," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Interv.*, 2019, pp. 92–100.

[3] Z. Wang *et al.*, "MRF denoising with compressed sensing and adaptive filtering," in *Proc. IEEE 11th Int. Symp. Biomed. Imag.*, 2014, pp. 870–873.

[4] D. F. McGivney *et al.*, "SVD compression for magnetic resonance fingerprinting in the time domain," *IEEE Trans. Med. Imag.*, vol. 33, no. 12, pp. 2311–2322, Dec. 2014.

[5] P. A. Gómez *et al.*, "Simultaneous parameter mapping, modality synthesis, and anatomical labeling of the brain with MR fingerprinting," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Interv.*, 2016, pp. 579–586, 2016.

[6] C. C. Cline *et al.*, "AIR-MRF: Accelerated iterative reconstruction for magnetic resonance fingerprinting," *Magn. Reson. Imag.*, vol. 41, pp. 29–40, Sep. 2017.

[7] B. Zhao *et al.*, "Maximum likelihood reconstruction for magnetic resonance fingerprinting," in *Proc. IEEE 12th Int. Symp. Biomed. Imag.*, 2015, pp. 905–909.

[8] S. Woo *et al.*, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 3–19.

[9] O. Cohen, B. Zhu, and M. S. Rosen, "MR fingerprinting deep reconstruction network (DRONE)," *Magn. Reson. Med.*, vol. 80, no. 3, pp. 885–894, 2018.

[10] D. Chen *et al.*, "Deep fully convolutional network for mr fingerprinting," in *Int. Conf. Med. Imag. Deep Learn. - Extended Abstract Track*, Jul. 2019, *arXiv:1911.09846*.

[11] I. Oksuz *et al.*, "Magnetic resonance fingerprinting using recurrent neural networks," in *Proc. IEEE 16th Int. Symp. Biomed. Imag.*, 2019, pp. 1537–1540.

[12] E. Hoppe *et al.*, "Magnetic resonance fingerprinting reconstruction using recurrent neural networks," *Stud. Health Technol. Informat.*, vol. 267, pp. 126–133, 2019.

[13] F. Balsiger *et al.*, "On the spatial and temporal influence for the reconstruction of magnetic resonance fingerprinting," in *Proc. Int. Conf. Med. Imag. Deep Learn.*, 2019, pp. 27–38.

[14] P. Cao *et al.*, "Development of fast deep learning quantification for magnetic resonance fingerprinting in VIVO," *Magn. Reson. Imag.*, vol. 70, pp. 81–90, 2020.

[15] Z. Fang *et al.*, "Deep learning for fast and spatially constrained tissue quantification from highly accelerated data in magnetic resonance fingerprinting," *IEEE Trans. Med. Imag.*, vol. 38, no. 10, pp. 2364–2374, Oct. 2019.

[16] Z. Fang *et al.*, "RCA-U-Net: Residual channel attention U-Net for fast tissue quantification in magnetic resonance fingerprinting," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Interv.*, 2019, pp. 101–109.

[17] C. M. Pirk *et al.*, "Deep learning-based parameter mapping for joint relaxation and diffusion tensor MR fingerprinting," *Med. Imag. Deep Learn.*, 2020, vol. 121, pp. 638–654.

[18] P. A. Gómez *et al.*, "Learning a spatiotemporal dictionary for magnetic resonance fingerprinting with compressed sensing," in *Proc. Patch-Based Techn. Med. Imag.*, 2015, pp. 112–119.

[19] J. Assländer *et al.*, "Low rank alternating direction method of multipliers reconstruction for mr fingerprinting," *Magn. Reson. Med.*, vol. 79, pp. 83–96, Jan. 2018.

[20] J. Hennig, M. Weigel, and K. Scheffler, "Calculation of flip angles for echo trains with predefined amplitudes with the extended phase graph (EPG)-algorithm: Principles and applications to hyperecho and traps sequences," *Magn. Reson. Med.*, vol. 51, no. 1, pp. 68–80, 2004.

[21] A. Bustin *et al.*, "High-dimensionality undersampled patch-based reconstruction (HD-PROST) for accelerated multi-contrast MRI," *Magn. Reson. Med.*, vol. 81, no. 6, pp. 3705–3719, 2019.

[22] E. Hoppe *et al.*, "Deep learning for magnetic resonance fingerprinting: A new approach for predicting quantitative parameter values from time series.," *Stud. Health Technol. Inform.*, vol. 243, pp. 202–206, 2017.

[23] J. Ashburner *et al.*, SPM12 Manual. Wellcome Trust Centre Neuroimaging. London, U.K., 2014, Art. no. 2464.

[24] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Interv.*, 2015, pp. 234–241.

[25] Z. Fang *et al.*, "Submillimeter MR fingerprinting using deep learning-based tissue quantification," *Magn. Reson. Med.*, vol. 84, no. 2, pp. 579–591, 2020.

[26] P. Virtue, S. Yu, and M. Lustig, "Better than real: Complex-valued neural nets for MRI fingerprinting," in *Proc. IEEE Int. Conf. Image Process.*, 2017, pp. 3953–3957.

[27] M. Barbieri *et al.*, "Circumventing the curse of dimensionality in magnetic resonance fingerprinting through a deep learning approach," 2018, *arXiv:1811.11477*.