

MRI RECONSTRUCTION VIA CASCADED CHANNEL-WISE ATTENTION NETWORK

Qiaoying Huang, Dong Yang, Pengxiang Wu, Hui Qu, Jingru Yi, Dimitris Metaxas

Department of Computer Science, Rutgers University, NJ, USA

ABSTRACT

We consider an MRI reconstruction problem with input of k-space data at a very low undersampled rate. This can practically benefit patient due to reduced time of MRI scan, but it is also challenging since quality of reconstruction may be compromised. Currently, deep learning based methods dominate MRI reconstruction over traditional approaches such as Compressed Sensing, but they rarely show satisfactory performance in the case of low undersampled k-space data. One explanation is that these methods treat channel-wise features equally, which results in degraded representation ability of the neural network. To solve this problem, we propose a new model called MRI Cascaded Channel-wise Attention Network (MICCAN), highlighted by three components: (i) a variant of U-net with Channel-wise Attention (UCA) module, (ii) a long skip connection and (iii) a combined loss. Our model is able to attend to salient information by filtering irrelevant features and also concentrate on high-frequency information by enforcing low-frequency information bypassed to the final output. We conduct both quantitative evaluation and qualitative analysis of our method on a cardiac dataset. The experiment shows that our method achieves very promising results in terms of three common metrics on the MRI reconstruction with low undersampled k-space data. Code is public available¹.

Index Terms— reconstruction, attention, skip connection

1. INTRODUCTION

Magnetic resonance imaging (MRI) is widely used due to its high resolution and low radiation, but fully-sampled MRI scan requires lots of time for patients. Therefore, MRI data is usually undersampled in the Fourier domain to improve the scan efficiency. According to Compressed Sensing (CS) theory that assumes signal is sparse in a transform domain, reconstruction is possible even if signal is undersampling below the Nyquist rate [1, 2]. Naturally, undersampling in the transform domain introduces artifacts and significant coefficients are required to infer a de-noised image. Such interference can be formalized as a nonlinear optimization problem, recovering the sparse coefficients and the image itself effectively.

Lustig *et al.* proposed to minimize the ℓ_1 norm of a transformed image, subject to data consistent constraints [3]. Another variant is to replace ℓ_1 with total variation (TV) penalty [4]. RecPF [5] employed the alternating direction method to minimize a linear combination of ℓ_1 and TV norm regularization. FCSA [6] further improved the quality of reconstructed images by weighting solutions from ℓ_1 and TV norm sub-problems in an iterative framework. These works concentrate on designing new objective functions and optimization algorithms, but they omit the fact that reconstruction procedure is computationally inefficient and requires thousands of iterations until convergence. Moreover, objective function largely depends on prior information of data, which is usually unknown in practical problem.

Recently, deep learning based MRI reconstruction methods are widely investigated to overcome such drawbacks due to its strong ability of feature representation. Sun *et al.* [7] first proposed to use neural network in MRI reconstruction problem, where the network was derived from the iterative procedures in Alternating Direction Method of Multipliers (ADMM) algorithm [8] to optimize a general CS-based MRI model. Schlemper *et al.* proposed a deep cascaded convolution neural network with a data consistency unit (DC-CNN) which improves the performance of MRI reconstruction [9]. Recent work proposed a deep cascading of the U-net structure, which further decreases the reconstruction error [10].

One common limitation of existing deep learning methods is that they barely show promising performance on k-space data at a very low undersampled rate. One reason is that these methods treat channel-wise features equally, which degrades representation ability of neural network. To resolve the problem, we propose a new model called MRI Cascaded Channel-wise Attention Network (MICCAN). It is featured by three components: (i) a variant of U-net with Channel-wise Attention (UCA) module, (ii) a long skip connection and (iii) a combined loss. The UCA module exploits channel-wise attention mechanism [11, 12, 13] and aims to attend to important features related to the final goal by filtering irrelevant and noisy features. The long skip connection allows abundant low-frequency information bypassed to the final output which enforces the model to focus on learning high-frequency information. By adaptively reweighing features according to their inter-dependencies among feature channels, representation ability of deep network is further improved. To the best

¹<https://github.com/charwing10/isbi2019miccan>

of our knowledge, this is the first work to employ the channel-wise attention mechanism in MRI reconstruction problem.

2. METHODOLOGY

In CS-based MRI reconstruction problem, the goal is to find a reconstructed image \tilde{x} such that it minimizes the reconstruction error between original k-space data y and Fourier transform of x :

$$\tilde{x} = \arg \min_x \|F_u(x) - y\|_2^2 + R(x). \quad (1)$$

Here F_u is an operator that transforms x into Fourier domain with undersampling. $R(\cdot)$ is a regularization term that depends on the prior information of data and general choices are ℓ_1 or ℓ_0 .

2.1. Preliminary

Traditionally, the objective function of Equation 1 is solved in an iterative manner that requires thousands of iteration steps until convergence. A more efficient way is to approximate the optimal reconstructed image from undersampled k-space data via deep neural network. One issue is that as the network goes deeper, the original information may be degraded. Schlemper *et al.* [9] proposed a Data Consistency (DC) layer to avoid loss of information. Basically, DC layer takes as input the reconstructed image x_n from the n -th reconstruction block and outputs an updated ‘‘reconstructed’’ image x_n^{dc} . Formally, DC layer is defined as

$$x_n^{dc} = \text{DC}(x_n) = F^H(\tau(F(x_n))). \quad (2)$$

Here $F(x)$ is Fourier transform that takes image x as input and outputs \hat{x} in Fourier domain and $F^H(\hat{x})$ is inverse Fourier transform. $\tau(\hat{x})$ is the data fidelity operation whose output has the same dimension as \hat{x} :

$$\tau(\hat{x})[i, j] = \begin{cases} \hat{x}[i, j], & (i, j) \notin \Omega \\ \frac{\hat{x}[i, j] + v y[i, j]}{1+v}, & (i, j) \in \Omega \end{cases} \quad (3)$$

where $[i, j]$ is the matrix indexing operation, Ω is the set of sampled positions of k-space data and $v \in [0, \infty)$ is the noise level. In the noiseless case ($v \rightarrow \infty$), we have $\hat{x}[i, j] = y[i, j]$ if $(i, j) \in \Omega$, *i.e.* filling \hat{x} with the original values of k-space data y at position (i, j) .

2.2. MRI Cascaded Channel-wise Attention Network

In this section, we propose a new network called *MRI Cascaded Channel-wise Attention Network* (MICCAN). As shown in Figure 1, MICCAN mainly consists of two parts: *U-net with Channel-wise Attention* (UCA) module and DC layer. These components are cascadedly coupled together and repeat for N times. Formally, denote the n -th UCA module

and the n -th DC layer respectively by UCA_n and DC_n . The starting point of MICCAN is undersampled k-space data y , which is later converted into a zero-filling image $x_0 = F^H(y)$ through inverse Fourier transform F^H and fed to a UCA module. Our cascaded model can be simply formalized as

$$\begin{cases} x_n = \text{UCA}_n(x_n^{dc}) + x_{n-1}^{dc} \\ x_n^{dc} = \text{DC}_n(x_n) \end{cases} \quad (n = 1, \dots, N) \quad (4)$$

where x_0^{dc} is initialized as x_0 . The final reconstructed image of MICCAN, namely x_N^{dc} , is produced by the last DC layer (DC_N).

2.2.1. U-net with Channel-wise Attention

In the previous work on reconstruction problem, deep learning based methods have two major issues. First, they treat each channel-wise feature equally, but contributions to the reconstruction task vary from different feature maps. Second, receptive field in convolutional layer may cause to lose contextual information from original images, especially high-frequency components that contain valuable detailed information such as edges and texture. Therefore, inspired by [13], we develop the UCA module by introducing an attention mechanism that filters the useless features and enhance the informative ones. The attention technique is only applied in the decoder part. The intuition is that features of the decoder are extracted from coarse to fine feature-maps of multiple scales via skip connection. The attention module filters salient and prunes irrelevant and noisy features such that allows model parameters in shallower layers to be updated mostly that are relevant to a given task. To the best of our knowledge, this is the first work to employ channel-wise attention to MRI reconstruction problem.

Specifically, we use global average pooling to extract the channel-wise global spatial information to vector $z \in \mathbb{R}^C$, whose c dimension is defined as

$$z_c = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W f_c[i, j] \quad (c = 1, \dots, C) \quad (5)$$

where $f_c \in \mathbb{R}^{W \times H}$ is the feature map in the c -th channel. Such operation squeezes the spatial information of the whole image into a vector length of C . To further extract feature related to the final task, we introduce another gating mechanism as follows:

$$\hat{x}_c = \sigma(\delta(z * \mathbf{W}_1) * \mathbf{W}_2) \odot f_c, \quad (6)$$

where ‘‘*’’ is convolution operator and $\delta(\cdot)$ is ReLU activation function to encode the channel-wise dependencies. \mathbf{W}_1 is a kernel in the first convolutional layer that reduces the C -dimensional feature vector into C/r . On the contrary, kernel \mathbf{W}_2 increases feature size back to C . Sigmoid function $\sigma(\cdot)$ is used to compute weighted attention map, which is later

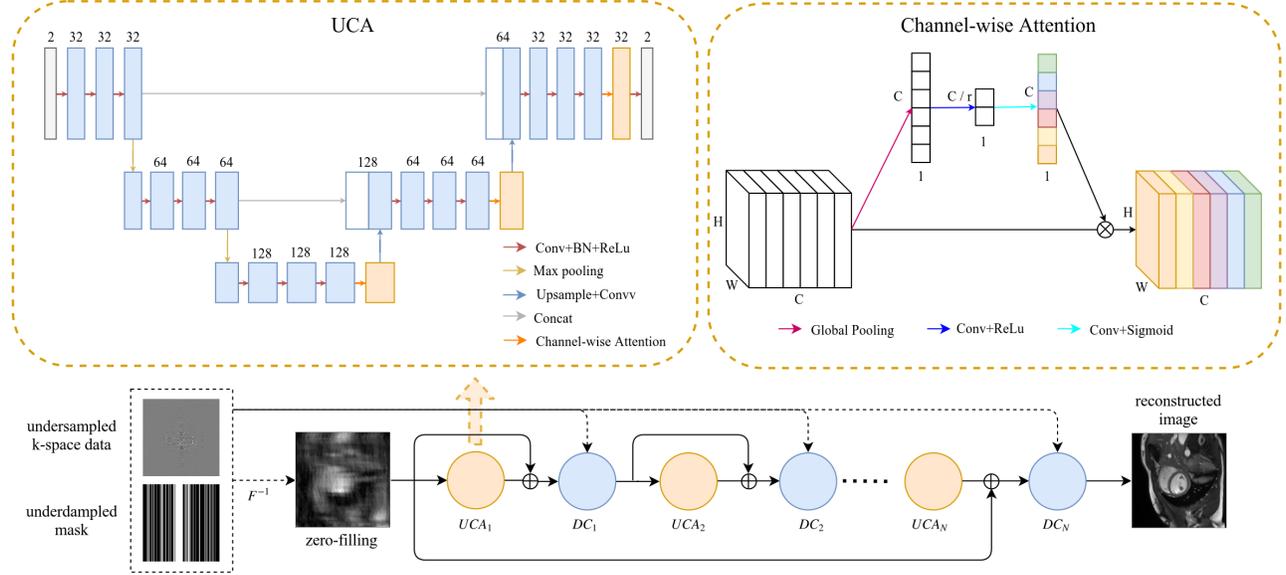


Fig. 1. The framework of the proposed MICCAN network. The end-to-end network consists of UCA blocks, Data Consistency blocks and Channel-wise Attention units. It takes as input the undersampled k-space data and the undersampled mask (left-most) and outputs the reconstructed MRI (right-most image). The zero-filled reconstruction (second-left image) works as the start point for reconstruction. The Data Consistency unit employs the original k-space data for further refinement.

applied to rescaling the input feature f_c . Based on this attention mechanism in the UCA module, our model MICCAN achieves very promising results and outperforms several state-of-the-art methods.

2.2.2. Learn Global Residual by Long Skip Connection

To address the problem of vanishing low frequency in deep learning based MRI reconstruction, we utilize a long skip connection from the zero-filling image to the final reconstruction block. Specifically, we replace the residual in the last UCA module, namely $x_N = \text{UCA}_N(x_{N-1}^{dc}) + x_{N-1}^{dc}$, with a long skip connection:

$$x_N = \text{UCA}_N(x_{N-1}^{dc}) + x_0 \quad (7)$$

This simple modification is used to learn the global residual and to stabilize the gradient flow in deep residual network.

2.2.3. Combined Loss Function

Common choice of loss function for reconstruction problem is ℓ_2 , but the resulting reconstructed image is of low quality and lacks high frequency detail. Therefore, we propose to use a combination of loss functions including ℓ_1 loss and perceptual loss ℓ_p [14]. Given target image x_s and reconstructed image $x \doteq x_N^{dc}$ of MICCAN parameterized by θ , the combined loss is defined as

$$\ell^\theta(x, x_s) = \lambda_1 \ell_1^\theta(x, x_s) + \lambda_p \ell_p^\theta(x, x_s) \quad (8)$$

where

$$\ell_1^\theta(x, x_s) = \|x - x_s\|_1 \quad (9)$$

$$\ell_p^\theta(x, x_s) = \sum_{k=1}^K \|\phi_{VGG}^k(x) - \phi_{VGG}^k(x_s)\|_2^2 \quad (10)$$

where λ_1 and λ_p are weighing factors for two losses, $\phi_{VGG}^k(\cdot)$ represents features of the k -th activation layer in VGG network. Note that perceptual loss ℓ_p minimizes the ℓ_2 distance between reconstruction image and target image in K different feature spaces, or equivalently it encourages the predicted image to be perceptually similar to the target image.

3. EXPERIMENTS

To evaluate the effectiveness of our proposed model, we compare it with several state-of-the-art approaches on a simulated cardiac k-space dataset.

Methods We evaluate following methods in the experiment. Two traditional CS-MRI methods include ℓ_1 -wavelet [3] and TV norm [4], which are implemented in the BART toolbox [15]. Two deep learning based models include DC-CNN[9] and MRN5[10]. The only difference between our proposed MICCAN and MRN5 is that MRN5 does not used attention module and long skip connection. We also consider three variants of our methods: MICCAN with ℓ_2 loss (MICCAN-A), MICCAN with combined loss proposed in Section 2.2.3 (MICCAN-B) and MICCAN with both long skip connection and combined loss (MICCAN-C). The same as [11, 13], we set the reduction ratio r that is mentioned in Section 2.2.1 as 8 for all our MICCAN models. For the combined loss, we set λ_1 as 10 and λ_p as 0.5. As the shallow features encode details of images, we compute the perceptual

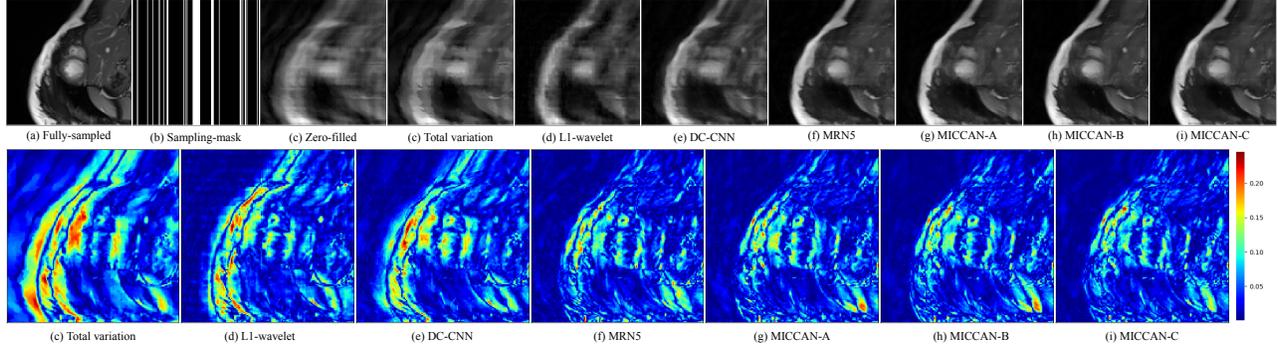


Fig. 2. Visualization of reconstructed images and reconstruction errors with k-space data at undersampled rate 12.5%

Table 1. Reconstruction results of the proposed MICCAN models and other methods.

Methods	NRMSE	PSNR	SSIM
TV	0.1087 \pm 0.0346	19.7388 \pm 2.8942	0.5923 \pm 0.0898
ℓ_1 -wavelet	0.0753 \pm 0.0185	22.7054 \pm 2.0255	0.6333 \pm 0.0553
DC-CNN	0.0587 \pm 0.0118	24.7993 \pm 1.7612	0.6612 \pm 0.0582
MRN5	0.0427 \pm 0.0076	27.5373 \pm 1.5971	0.7851 \pm 0.0297
MICCAN-A	0.0402 \pm 0.0072	28.0664 \pm 1.6231	0.8005 \pm 0.0301
MICCAN-B	0.0391 \pm 0.0074	28.3283 \pm 1.7080	0.8214\pm0.0327
MICCAN-C	0.0385\pm0.0073	28.4489\pm1.6953	0.8198 \pm 0.0302

loss of layer *relu1-2*, *relu2-1*, *relu2-2* and *relu3-1* of the VGG-16 network. All deep learning models are implemented using PyTorch and trained on NVIDIA K80. Learning rate is initialized as 10^{-4} with decreasing rate of 0.5 for every 15 epochs. The training batch is 8 and the maximum number of epochs is 50. For fair comparison, we follow [9, 10] by setting the number of reconstruction blocks N as 5.

Dataset A cardiac MRI dataset with 15 subjects is adopted in our experiments. We randomly choose 10 subjects as training set, 2 subjects as validation set and the rest 3 subjects as test set. We follow the k-space data simulation method proposed in [16]. It assumes the sampled mask follows a zero-mean Gaussian distribution and the Cartesian undersampling method is adopted, also keeps the eight lowest spatial frequencies. Our model is evaluated on data with undersampled rate at 12.5% (acceleration rate $8\times$).

3.1. Quantitative Evaluation

In this experiment, we quantitatively evaluate all models with three widely used measurements: normalized root square mean error (NRMSE), peak signal-to-noise ration (PSNR) and the structural similarity index measure (SSIM). The results are shown in Table 1 and we basically observe following three trends. First, all our MICCAN variants outperform other baseline models. This mainly attributes to the attention mechanism proposed in Section 2.2.1. Even trained with ℓ_2 loss, MICCAN-A achieves NRMSE of 0.0402, PSNR of

28.0664 and SSIM of 0.8005 that beats MRN5 model. Second, MICCAN-B model that is trained with combined loss gains better result than MICCAN-A. For example, MICCAN-B has 2.7% decrease in NRMSE compared with MICCAN-A and also achieves the best SSIM value of 0.8214. This indicates that our combined loss is better than the ℓ_2 loss, making image less blur and keeping the perceptual details. Third, with long skip connection, MICCAN-C further improves reconstruction performance with the lowest NRMSE of 0.0385 and the highest PSNR value of 28.4489. Overall, all these results demonstrate the effectiveness of the channel-wise attention modules and the proposed combined loss.

3.2. Qualitative Analysis

In this experiment, we visualize the reconstructed image and reconstruction errors on a test sample and qualitatively analyze the results of all methods. As shown in Figure 2, most of the baseline methods cannot completely recover detail of the image and suffer from severe blurring artifacts. In contrast, our three MICCAN methods eliminate most blurring artifacts and recover more details from low undersampled k-space data. Furthermore, for the transitional methods such as ℓ_1 -wavelet and TV, the reconstructed images are similar to the zero-filling image and suffer from heavy blurring artifacts. We highlight that our MICCAN achieves best results and with much fewer reconstruction errors.

4. CONCLUSION

In this paper, we propose a robust and effective model called MICCAN for MRI reconstruction from very low undersampled k-space data. We devise a variant of U-net with channel-wise attention as reconstruction block together with long skip connection technique and a combined loss function. We evaluate our method and current state-of-the-art approaches on a cardiac benchmark. The experimental results show that our method outperforms all other compared methods and achieves very promising results under the condition of low undersampled rate.

5. REFERENCES

- [1] David L Donoho, "Compressed sensing," *IEEE Transactions on information theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [2] Emmanuel J Candès et al., "Compressive sampling," in *Proceedings of the international congress of mathematicians*. Madrid, Spain, 2006, vol. 3, pp. 1433–1452.
- [3] Michael Lustig, David Donoho, and John M Pauly, "Sparse mri: The application of compressed sensing for rapid mr imaging," *Magnetic resonance in medicine*, vol. 58, no. 6, pp. 1182–1195, 2007.
- [4] Michael Lustig, David L Donoho, Juan M Santos, and John M Pauly, "Compressed sensing mri," *IEEE signal processing magazine*, vol. 25, no. 2, pp. 72–82, 2008.
- [5] Junfeng Yang, Yin Zhang, and Wotao Yin, "A fast alternating direction method for tvl1-l2 signal reconstruction from partial fourier data," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 288–297, 2010.
- [6] Junzhou Huang, Shaoting Zhang, and Dimitris Metaxas, "Efficient mr image reconstruction for compressed mr imaging," *Medical Image Analysis*, vol. 15, no. 5, pp. 670–679, 2011.
- [7] Jian Sun, Huibin Li, Zongben Xu, et al., "Deep admm-net for compressive sensing mri," in *Advances in Neural Information Processing Systems*, 2016, pp. 10–18.
- [8] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, Jonathan Eckstein, et al., "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine learning*, vol. 3, no. 1, pp. 1–122, 2011.
- [9] Jo Schlemper, Jose Caballero, Joseph V Hajnal, Anthony Price, and Daniel Rueckert, "A deep cascade of convolutional neural networks for mr image reconstruction," in *International Conference on Information Processing in Medical Imaging*. Springer, 2017, pp. 647–658.
- [10] Liyan Sun, Zhiwen Fan, Yue Huang, Xinghao Ding, and John Paisley, "Joint cs-mri reconstruction and segmentation with a unified deep network," *arXiv preprint arXiv:1805.02165*, 2018.
- [11] Jie Hu, Li Shen, and Gang Sun, "Squeeze-and-excitation networks," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2018.
- [12] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Binen Zhong, and Yun Fu, "Image super-resolution using very deep residual channel attention networks," in *European Conference on Computer Vision*. Springer, 2018.
- [13] Abhijit Guha Roy, Nassir Navab, and Christian Wachinger, "Concurrent spatial and channel squeeze & excitation in fully convolutional networks," in *Medical Image Computing and Computer Assisted Intervention*. Springer, 2018.
- [14] Justin Johnson, Alexandre Alahi, and Li Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *European Conference on Computer Vision*. Springer, 2016, pp. 694–711.
- [15] Jonathan I Tamir, Frank Ong, Joseph Y Cheng, Martin Uecker, and Michael Lustig, "Generalized magnetic resonance image reconstruction using the berkeley advanced reconstruction toolbox," in *ISMRM Workshop on Data Sampling & Image Reconstruction, Sedona, AZ*, 2016.
- [16] Hong Jung, Jong Chul Ye, and Eung Yeop Kim, "Improved k-t blast and k-t sense using focuss," *Physics in Medicine & Biology*, vol. 52, no. 11, pp. 3201, 2007.