Region-Guided Channel-Wise Attention Network for Accelerated MRI reconstruction

Region-Guided Channel-Wise Attention Network for Accelerated MRI Reconstruction

Anonymous

Anonymous Organization

Abstract. Magnetic resonance imaging (MRI) has been widely used in clinical practice for medical diagnosis of diseases. However, the long acquisition time hinders its development in time-critical applications. In recent years, deep learning-based methods leverage the powerful representations of neural networks to recover high-quality MR images from undersampled measurements, which shortens the acquisition process and enables accelerated MRI scanning. Despite the achieved inspiring success, it is still challenging to provide high-fidelity reconstructions under high acceleration factors. As an important mechanism in deep neural networks, attention modules have been used to improve the reconstruction quality. Due to the computational costs, many attention modules are not suitable for applying to high-resolution features or to capture spatial information, which potentially limits the capacity of neural networks. To address this issue, we propose a novel channel-wise attention which is implemented under the guidance of implicitly learned spatial semantics. We incorporate the proposed attention module in a deep network cascade for fast MRI reconstruction. In experiments, we demonstrate that the proposed framework produces superior reconstructions with appealing local visual details, compared to other deep learning-based models, validated qualitatively and quantitatively on the FastMRI knee dataset.

Keywords: MRI reconstruction · Deep Learning · Region-guided channel-wise attention.

1 Introduction

Magnetic resonance imaging (MRI) provides a powerful and non-invasive tool for medical diagnosis. The acquisition process is notoriously time-consuming due to physiological and hardware constraints. Undersampling k-space data is a common practice to accelerate the process, which however inevitably causes aliasing artifacts in image domain. The ill-posed problems can be modeled as,

$$\min_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|^2 + \lambda R(\mathbf{x}), \tag{1}$$

where $\mathbf{A}$ denotes the encoding operation, $\mathbf{y}$ is the k-space measurement, and $R(\mathbf{x})$ is a regularization on the reconstruction $\mathbf{x}$. Compressed sensing (CS) methods assume the sparsity of signals in image domain [6] or in some transformed space [11, 18, 22], and solve the optimization problem using iterative model-based

algorithms. Nevertheless, it is challenging to hold the sparsity assumption in real scenarios and remove the aliasing artifacts via conventional methods [28], which restrains the growth of CS methods in modern MRI.

Recently, deep neural networks have been shown to perform favorably in imaging tasks [15, 7, 26]. Incorporating the representations of neural networks in MRI reconstruction shows superior performance in many works [29, 10]. The method in [16] removes aliasing artifacts in MR images using dual magnitude and phase networks. The method in [32] introduces a primal-dual network to solve the traditional CS-MRI problem. However, the models trained with pixel-wise losses, e.g. MSE, can fail to produce sharp and realistic local details and lead to smoothed reconstructions. Generative adversarial networks (GAN) [8] exploit the adversarial game between the generator and discriminator to implicitly model the data distribution, and have been used to enhance the image quality of MRI reconstructions [28, 30, 19]. However, GAN-based models can potentially produce undesired details and fail to preserve faithful diagnostic features, by lacking of k-space data consistency constraints. Concurrently, attention mechanism plays an important role in vision tasks [33], which learns to capture feature dependencies to enhance the model representation capacity. Many methods leverage attention mechanism to improve MRI reconstruction quality [30]. Nevertheless, the significantly increased computational overheads limit its implementations to high-resolution features which are more closely associated with dense predictions, e.g. reconstruction tasks. Alternatively, channel-wise attention models the inter-plays between feature channels efficiently and can potentially provide better MRI reconstructions [13, 19, 17], whereas the spatial information is not considered, limiting the representation capacity of attention over channels.

In this paper, we introduce a novel region-guided channel-wise attention network for fast MRI reconstruction to exploit the channel-wise attention and improve the reconstruction quality. It has been incorporated with the implicitly learned spatial semantics to increase the attention diversities and gain performance boost. To provide more accurate restoration, we leverage the k-space consistency information in a densely connected network cascade and train the model in an adversarial diet. The main contributions of our work can be summarized as follows: 1) a novel region-guided channel-wise attention network for MRI reconstruction, which introduces spatial information into channel attention mechanism; 2) deriving the region-based semantic information to guide attention over channels, which increases the attention diversity and achieves performance gains; 3) by experiments, we demonstrate that the proposed method outperforms other deep learning-based approaches qualitatively and quantitatively.

2 Methods

The proposed region-guided channel-wise attention network for MRI reconstruction endows channel-wise attention with spatial diversities to enhance the reconstruction performance. We elaborate on the details as follows.

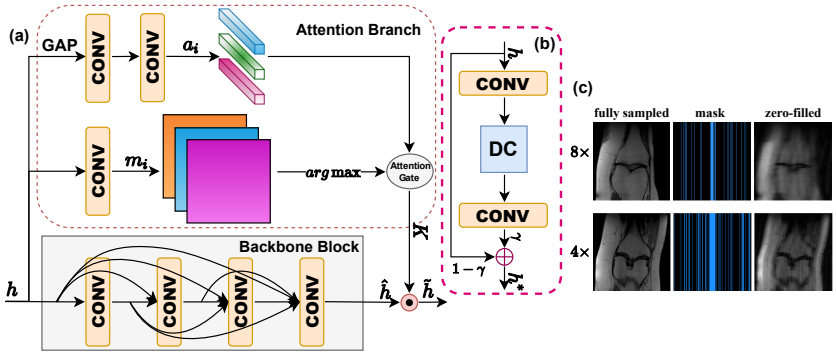

Fig. 1: Illustration of (a) RG-CAM, (b) RDCB, and (c) undersampling.

2.1 Region-Guided Channel-Wise Attention Module (RG-CAM)

A novel region-guided channel-wise attention module (RG-CAM) is proposed here to implicitly learn the spatial semantic information to guide the attention mechanism over feature channels. As displayed in Fig. 1 (a), the input features $h \in R^{C\times H\times W}$, where C and $H \times W$ denote the channel and spatial sizes, respectively attend to the backbone and attention branches. The output features $\hat{h}$ from the backbone are refined using the output K from the attention branch. The densely connected layers are used as the backbone for their effective representations [12], which can be replaced by arbitrary structures. Owing to its flexible design, RG-CAM can be easily combined with other network structures, e.g. the spatial attention modules [33, 27, 17], to retrieve further improvements. The details of RG-CAM are presented in the following.

Channel-Wise Attention Kernel Bank: The input h is squeezed via the global average pooling (GAP) and pass it to a non-linear mapping comprising two linear layers with a GELU activation in between, as shown in Fig. 1 (a). The output is resized to be $M \times C \times 1 \times 1$, representing M channel-wise attention kernels. Each kernel, dubbed a_i, is mapped to (0, 1) via a Sigmoid activation, where 1 means full attention to this channel and 0 denotes no attention. The kernels in bank $\{a_i\}_{i=1}^{M}$ are then incorporated together in a region-guided manner.

Spatial Guiding Mask: M spatial guiding masks $m_i \in R^{H\times W}$ are generated from h, as presented in Fig. 1 (a), to guide the implementation of channel-wise attention. For each spatial location x, the guiding masks are normalized as below, to represent pixels from similar regions which will share an attention pattern,

$$m_j(x) = \begin{cases} 1, & j = arg\max_i m_i(x) \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

Region-Guided Attention Gate: The guiding masks are amalgamated with the attention kernels to construct the region-guided attention gate $K \in R^{C\times H\times W}$,

by stacking the kernels a_i via the criteria below,

$$K(x) \leftarrow a_i, \text{ if } m_i(x) = 1. \tag{3}$$

Due to the spatial invariance of convolutional networks, regions with similar semantics potentially have the same values in m_i and share the attention patterns. The final output $\tilde{h}$ of RG-CAM is given as follows,

$$\tilde{h} = K \odot \hat{h}, \tag{4}$$

where $\odot$ is the element-wise multiplication. RG-CAM endows the spatial diversity to attention over channels in a flexible, efficient, and light-weight manner. Compared to the backbone branch, the parameters and computational costs of the attention branch are negligible due to the GAP operation. It requires neither labeled data nor extra supervision, and is end-to-end trained. We use the Softmax trick [14] to enable the gradient propagation w.r.t the guiding masks and select $M = 8$ for all RG-CAMs.

2.2 Residual Data Consistency Block (RDCB)

GAN-based models have been proven to generate photo-realistic images for MRI reconstruction [28, 9, 4]. However, the lack of k-space constraints can lead to irrelevant details [3] and degradation in the reconstruction quality. To encourage more consistent reconstructions with the measurements $\mathbf{y}$, we propose a residual data consistency block (RDCB) to leverage the k-space information to "correct" the intermediate predictions. DC methods conventionally reduce the feature channels to handle complex-valued signals, i.e. using 2 channels, which can be detrimental to the model performance due to the bottleneck design. For example, the input and output channel sizes of the first convolution in (5) and Fig. 1 (b) are 16 and 2, representing the feature maps and complex signals, and vice versa for the second convolution. Instead, we take advantage of residual learning in the feature space to facilitate feature propagation in DC blocks. As presented in Fig. 1 (b), the resultant RDCB can be formulated as below,

$$h^* \leftarrow (1 - \gamma) \times h + \gamma \times \text{conv}(DC(\text{conv}(h), \mathbf{y}; \mathbf{A})), \tag{5}$$

where h and h^* denote the input and output features, γ is a trainable parameter, and DC refers to the data consistency operation [21, 25] which is given by,

$$DC(\mathbf{x}, \mathbf{y}; \mathbf{A}) = \mathbf{x} - \mathbf{A}^H(\mathbf{A}\mathbf{x} - \mathbf{y}). \tag{6}$$

In Sect. 3.3, we demonstrate that the slight modification delivers performance improvements, showing the efficacy of the proposed residual design.

2.3 Densely Connected Reconstruction Cascade

Deep cascaded networks are shown to yield higher performance in MRI reconstruction [16, 5, 24, 1], by virtue of the representation power of deep structures.

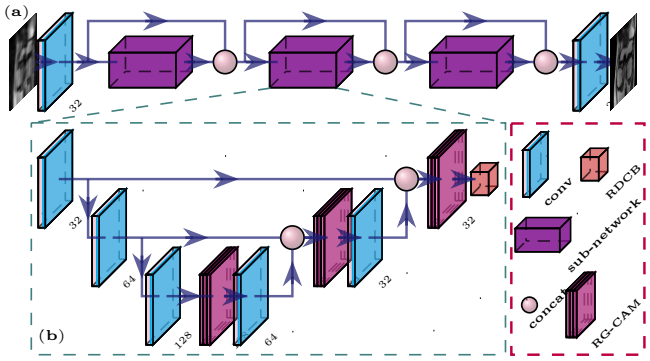

Fig. 2: Illustration of (a) densely connected reconstruction cascade and (b) U-shaped sub-network.

Inspired by dense connections in [12], we propose a densely connected reconstruction cascade to facilitate feature reuse and transmission. As shown in Fig. 2 (a), the current predictions are collected with the outputs from the preceding sub-networks via concatenation and passed as input to the following model structure. The collection of output features from all sub-networks are fused to give the final reconstruction. The framework takes the zero-filled as input and adopts five U-shaped sub-networks as illustrated in Fig. 2 (b).

2.4 Objective Function

We adopt the L_1 metric and structural similarity index (SSIM) L^{SSIM} to measure the reconstruction errors. We also train the model with an adversarial loss L_{adv} [20] to encourage sharp details. The total objective is given as below, with practically selecting $\alpha = 0.4$ and $\lambda_{adv} = 0.01$ in our experiments,

$$L = (1-\alpha)L_1(G,s) + \alpha L^{SSIM}(G,s) + \lambda_{adv}L_{adv}(G,s), \tag{7}$$

where G and s refer to the reconstruction and reference.

3 Experiment and Results

3.1 Data and Implementation Details

We use the FastMRI single-coil knee cases [31] to conduct the experiments. Two random sampling masks are used with acceleration factors of 8× and 4×, as shown in Fig. 1 (c). We use two channels to represent complex-valued signals. The model is trained for 35 epochs with a batch size of 5, using an Adam optimizer with β_1=0.5, β_2=0.999, and a learning rate of 2×10^{-4}. The method was implemented in PyTorch on a NVIDIA RTX 3090 GPU. More results, including 4× accelerated reconstructions, are presented in Supplementary Material.

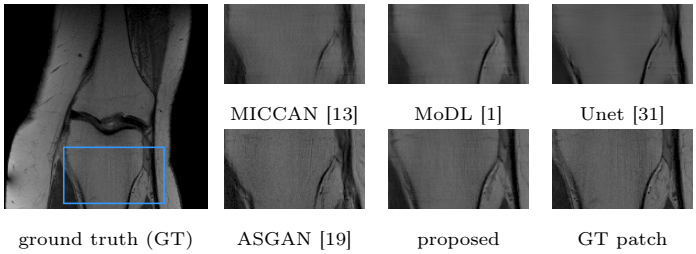

Fig. 3: Comparison results of 8× accelerated MRI reconstruction.

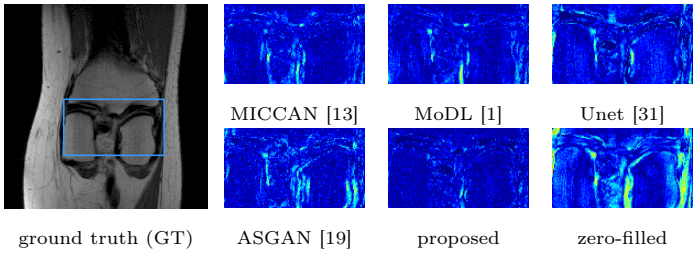

Fig. 4: Error maps (2× amplified) of 8× accelerated MRI reconstruction.

3.2 Comparison Experiments

We present the comparison results with other state-of-the-art methods: MICCAN [13], MoDL [1], FastMRI Unet [31], and ASGAN [19]. MICCAN and MoDL both adopt deep network cascades for reconstruction, where a channel-wise attention module is used in MICCAN. ASGAN adopts a GAN-based framework and performs attention selection for feature channels. We present the reconstructions using 8× and 4× acceleration factors in Fig. 3. It can be observed that the proposed method produces more faithful results with rich detailed structures, compared to other results. From the residual maps in Fig. 4, it is also shown that the proposed method restores the undersampled images with a higher accuracy, particularly at a high acceleration rate. Quantitatively, we use PSNR and SSIM as reconstruction error measurement and adopt FID and KID [2] for the visual evaluation. Table 1 lists the evaluation results, where the proposed method outperforms other competing approaches at both sampling rates (p-value$\ll$0.05). Due to the encoding-decoding structure, our method shows competitive inference speed, potentially enabling real-time reconstruction. Additionally, we replace conventional convolutions in sub-networks with separable (depth-wise+point-wise) convolutions [23] to strike a better accuracy-latency trade-off. The resultant variant "proposed-S" with fewer parameters is adopted to further verify that the superior performance is attributable to the proposed model components and structure, and not simply due to the size of model. We report the ablation results in the following section.

Table 1: Quantitative Evaluation on Accelerated MRI Reconstruction. 4× Accelerated Reconstructions Have the Same Run-Time and Model Size as 8×.

	method	PSNR↑	SSIM↑	FID↓	KID↓	run-time(s)↓	size(MB)↓
8×	proposed	**28.65**	**0.758**	**74.26**	**0.012**	0.049	44.5
	proposed-S	28.12	0.747	80.00	0.014	0.041	22.9
	ASGAN [19]	25.45	0.638	104.34	0.036	0.056	17.0
	Unet [31]	25.82	0.703	160.35	0.121	**0.013**	10.5
	MoDL [1]	27.13	0.620	143.65	0.080	0.091	22.3
	MICCAN [13]	26.61	0.642	180.66	0.146	0.043	**10.1**
	zero-filled	20.54	0.388	423.32	0.533	-	-
4×	proposed	**32.22**	**0.854**	**57.18**	**0.003**	-	-
	proposed-S	32.01	0.850	58.15	0.004	-	-
	ASGAN [19]	27.73	0.711	82.18	0.016	-	-
	Unet [31]	28.35	0.771	118.07	0.061	-	-
	MoDL [1]	30.34	0.745	98.86	0.042	-	-
	MICCAN [13]	30.11	0.711	99.44	0.040	-	-
	zero-filled	23.94	0.486	255.06	0.239	-	-

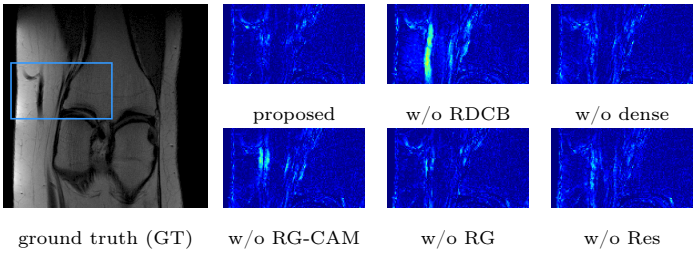

Fig. 5: Ablation residual maps (2× amplified) of 8× accelerated reconstruction.

3.3 Ablation Analysis

We conduct ablation experiments to evaluate the role of model components. We present the ablation results in Table 2, where "w/o RDCB", "w/o dense", and "w/o RG-CAM" respectively mean the proposed method without RDCB, dense network connections, and RG-CAM. For fair comparisons, we use feature repetition to maintain the channel size and model parameters, when removing the dense connections. The backbone in RG-CAM contains significantly more parameters than the attention branch, and it is maintained for a fair comparison. The convolutional layers in RDCB are also kept for the same reason. The results show that the proposed model components consistently improve the reconstruction performance in terms of evaluation metrics (p-value$\ll$0.05).

To further demonstrate the role of RG-CAM, the variant, dubbed "w/o RG", adopts a single kernel to perform channel-wise attention without using the spatial guiding masks, similar to [27, 17, 13]. From Table 2, we found that the incorporation of the region-guided mechanism enhances the model performance, comparing "w/o RG" and "proposed". In contrast, adopting conventional channel-wise attention fails to notably gain performance boost, comparing "w/o RG-CAM" and "w/o RG", which suggests the usefulness of the proposed region-guided method.

Table 2: Ablation Studies on Model Components at 8× Acceleration Factor.

method	PSNR↑	SSIM↑	FID↓	KID↓
proposed	**28.65**	**0.758**	**74.26**	**0.012**
w/o RDCB	27.33	0.731	83.43	0.017
w/o dense	28.03	0.748	81.25	0.015
w/o RG-CAM	28.29	0.753	78.67	0.014
w/o RG	28.26	0.754	77.40	0.013
w/o Res	28.34	0.748	80.13	0.015

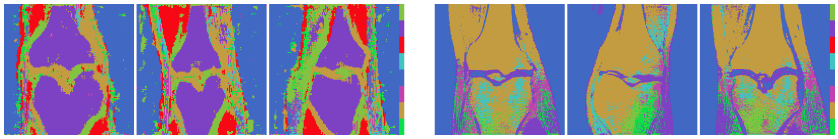

Fig. 6: Visualization of guiding masks from (left) penultimate decoding level of the 4-th sub-network and (right) last decoding level of the 5-th sub-network.

Additionally, the variant "w/o Res" refers to the removal of the residual structure in RDCB. The results in Table 2 confirm its efficacy in delivering performance gains. From the residual maps in Fig. 5, it is shown that the proposed method introduces more accurate reconstructions, compared to other candidates.

3.4 Region-Guided Mask Visualization

To visualize the spatial information learned in RG-CAM, we conflate the guiding masks where non-zero pixels are shaded in different colors. From the heat maps in Fig. 6, we can observe clear "segmentations" of different structures. It indicates that the spatial semantics are implicitly captured by "clustering" pixels from similar regions, which share the same attention patterns. It is worth noting that the region-based guiding information is learned and incorporated in channel recalibration without requiring any annotations for supervision.

4 Conclusions

In this paper, a novel region-guided channel-wise attention network is introduced for accelerated MRI reconstruction, which adopts an efficient and light-weight structure to simultaneously make use of the channel-wise attention and the implicitly learned spatial semantics. Incorporated with network dense connections and data consistency priors, it is demonstrated that the proposed method yields superior reconstruction performance at different acceleration factors, which can considerably shorten the MRI scanning time. For the future works, we plan to apply our method to other anatomical structures, and extend it to dynamic MRI reconstruction.

References

1. Aggarwal, H., Mani, M., Jacob, M.: MoDL: model-based deep learning architecture for inverse problems. IEEE Transactions on Medical Imaging **38**(2), 394–405 (2019). https://doi.org/10.1109/TMI.2018.2865356
2. Bińkowski, M., Sutherland, D., Arbel, M., Gretton, A.: Demystifying MMD GANs. International Conference on Learning Representations (2018)
3. Chen, S., Sun, S., Huang, X., Shen, D., Wang, Q., Liao, S.: Data-consistency in latent space and online update strategy to guide GAN for fast MRI reconstruction. Machine Learning for Medical Image Reconstruction - 3rd International Workshop, MLMIR 2020, Held in Conjunction with MICCAI 2020, Proceedings pp. 82–90 (2020). https://doi.org/10.1007/978-3-030-61598-7_8
4. Deora, P., Vasudeva, B., Bhattacharya, S., P.P., M.: Structure preserving compressive sensing MRI reconstruction using generative adversarial networks. The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops (June 2020)
5. Duan, J., Schlemper, J., Qin, C., Ouyang, C., Bai, W., Biffi, C., Bello, G., Statton, B., Regan, D., Rueckert, D.: VS-Net: variable splitting network for accelerated parallel MRI reconstruction. Springer International Publishing pp. 713–722 (2019)
6. Fair, M., Gatehouse, P., DiBella, E., Firmin, D.: A review of 3D first-pass, whole-heart, myocardial perfusion cardiovascular magnetic resonance. Journal of Cardiovascular Magnetic Resonance (2015)
7. Gatys, L., Ecker, A., Bethge, M.: Image style transfer using convolutional neural networks. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) pp. 2414–2423 (June 2016). https://doi.org/10.1109/CVPR.2016.265
8. Goodfellow, I., Pouget, A., Mirza, M., Xu, B., Warde, F., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial networks. Advances in Neural Information Processing Systems **27**, 2672–2680 (2014)
9. Guo, Y., Wang, C., Zhang, H., Yang, G.: Deep attentive Wasserstein generative adversarial networks for MRI reconstruction with recurrent context-awareness. Medical Image Computing and Computer Assisted Intervention – MICCAI 2020
10. Hammernik, k., Klatzer, T., Kobler, E., Recht, M., Sodickson, D., Pock, T., Knoll, F.: Learning a variational network for reconstruction of accelerated MRI data. Magnetic Resonance in Medicine **79**(6), 3055–3071 (2018). https://doi.org/https://doi.org/10.1002/mrm.26977
11. Hong, M., Yu, Y., Wang, H., Liu, F., Crozier, S.: Compressed sensing MRI with singular value decomposition-based sparsity basis. Physics in Medicine and Biology pp. 6311–6325 (Sep 2021)
12. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) pp. 2261–2269 (2017)
13. Huang, Q., Yang, D., Wu, P., Qu, H., Yi, J., Metaxas, D.: MRI reconstruction via cascaded channel-wise attention network. 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019) pp. 1622–1626 (2019). https://doi.org/10.1109/ISBI.2019.8759423
14. Kenji, I., Kuroki, R., Uchida, S.: Explaining convolutional neural networks using Softmax gradient layer-wise relevance propagation. International Conference on Computer Vision Workshop, ICCVW 2019 pp. 4176–4185 (Oct 2019)
15. Krizhevsky, A., Sutskever, I., Hinton, G.: ImageNet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems **25** (2012)

16. Lee, D., Yoo, J., Tak, S., Ye, J.: Deep residual learning for accelerated MRI using magnitude and phase networks. IEEE Trans. Biomed. Eng **65**(9), 1985–1995 (2018)
17. Li, G., Lv, J., Wang, C.: A modified generative adversarial network using spatial and channel-wise attention for CS-MRI reconstruction. IEEE Access **9**, 83185–83198 (2021)
18. Lingala, S., Jacob, M.: Blind compressive sensing dynamic MRI. IEEE Transactions on Medical Imaging **32**(6), 1132–1145 (2013)
19. Liu, J., Yaghoobi, M.: Fine-grained MRI reconstruction using attentive selection generative adversarial networks. ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) pp. 1155–1159 (2021)
20. Mao, X., Li, Q., Xie, H., Lau, R., Wang, Z., Smolley, S.: Least squares generative adversarial networks. 2017 IEEE International Conference on Computer Vision (ICCV) pp. 2813–2821 (2017)
21. Pezzotti, N., Yousefi, S., Elmahdy, M., van Gemert, J., Schülke, C., Doneva, M., et al: An Adaptive Intelligence Algorithm for Undersampled Knee MRI Reconstruction. arXiv e-prints arXiv:2004.07339 (Apr 2020)
22. Ravishankar, S., Bresler, Y.: MR image reconstruction from highly undersampled k-space data by dictionary learning. IEEE Transactions on Medical Imaging **30**(5), 1028–1041 (2011)
23. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.: MobileNetV2: inverted residuals and linear bottlenecks. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2018)
24. Schlemper, J., Caballero, J., Hajnal, J., Price, A., Rueckert, D.: A deep cascade of convolutional neural networks for dynamic MR image reconstruction. IEEE Transactions on Medical Imaging **37**(2), 491–503 (2018)
25. Sriram, A., Zbontar, J., Murrell, T., Defazio, A., Zitnick, C., Yakubova, N., Knoll, F., Johnson, P.: End-to-end variational networks for accelerated MRI reconstruction. Medical Image Computing and Computer Assisted Intervention - MICCAI **12262**, 64–73 (2020)
26. Wang, Y., Tao, X., Qi, X., Shen, X., Jia, J.: Image inpainting via generative multi-column convolutional neural networks. Advances in Neural Information Processing Systems pp. 331–340 (2018)
27. Woo, S., Park, J., Lee, J., Kweon, I.: CBAM: convolutional block attention module. CoRR **abs/1807.06521** (2018)
28. Yang, G., Yu, S., Dong, H., Slabaugh, G., Dragotti, P.L., Ye, X., Liu, F., Arridge, S., Keegan, J., Guo, Y., Firmin, D.: DAGAN: deep de-aliasing generative adversarial networks for fast compressed sensing MRI reconstruction. IEEE Transactions on Medical Imaging **37**(6), 1310–1321 (2018)
29. Yang, Y., Sun, J., Li, H., Xu, Z.: Deep ADMM-Net for compressive sensing MRI. Advances in Neural Information Processing Systems **29** (2016)
30. Yuan, Z., Jiang, M., Wang, Y., Wei, B., et al: SARA-GAN: self-attention and relative average discriminator based generative adversarial networks for fast compressed sensing MRI reconstruction. Frontiers in Neuroinformatics pp. 1–12 (2020)
31. Zbontar, J., Knoll, F., Sriram, A., Muckley, M., Bruno, M., et al: FastMRI: an open dataset and benchmarks for accelerated MRI. CoRR **abs/1811.08839** (2018)
32. Zhang, C., Liu, Y., Shang, F., Li, Y., Liu, H.: A novel learned primal-dual network for image cmpressive sensing. IEEE Access **9**, 26041–26050 (2021). https://doi.org/10.1109/ACCESS.2021.3057621
33. Zhang, H., Goodfellow, I., Metaxas, D., Odena, A.: Self-attention generative adversarial networks. Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019 **97**, 7354–7363 (2019)