Global k-Space Interpolation for Dynamic MRI Reconstruction Using Masked Image Modeling

Jiazhen Pan[1]([✉]), Suprosanna Shit[1], Özgün Turgut[1], Wenqi Huang[1],
Hongwei Bran Li[1,2], Nil Stolt-Ansó[3], Thomas Küstner[4],
Kerstin Hammernik[3,5], and Daniel Rueckert[1,3,5]

[1] School of Medicine, Klinikum Rechts der Isar, Technical University of Munich,
Munich, Germany
`jiazhen.pan@tum.de`, `suprosanna.shit@tum.de`
[2] Department of Quantitative Biomedicine, University of Zurich, Zürich, Switzerland
[3] School of Computation, Information and Technology, Technical University of
Munich, Munich, Germany
[4] Medical Image and Data Analysis, University Hospital of Tübingen, Tübingen,
Germany
[5] Department of Computing, Imperial College London, London, UK

Abstract. In dynamic Magnetic Resonance Imaging (MRI), k-space
is typically undersampled due to limited scan time, resulting in alias-
ing artifacts in the image domain. Hence, dynamic MR reconstruction
requires not only modeling spatial frequency components in the x and
y directions of k-space but also considering temporal redundancy. Most
previous works rely on image-domain regularizers (priors) to conduct MR
reconstruction. In contrast, we focus on interpolating the undersampled
k-space before obtaining images with Fourier transform. In this work,
we connect masked image modeling with k-space interpolation and pro-
pose a novel Transformer-based k-space Global Interpolation Network,
termed k-GIN. Our k-GIN learns global dependencies among low- and
high-frequency components of 2D+t k-space and uses it to interpolate
unsampled data. Further, we propose a novel k-space Iterative Refine-
ment Module (k-IRM) to enhance the high-frequency components learn-
ing. We evaluate our approach on 92 in-house 2D+t cardiac MR sub-
jects and compare it to MR reconstruction methods with image-domain
regularizers. Experiments show that our proposed k-space interpolation
method quantitatively and qualitatively outperforms baseline methods.
Importantly, the proposed approach achieves substantially higher robust-
ness and generalizability in cases of highly-undersampled MR data.

Keywords: Cardiac MR Imaging Reconstruction · k-space
Interpolation · Masked Image Modeling · Masked Autoencoders ·
Transformers

1   Introduction

CINE Cardiac Magnetic Resonance (CMR) imaging is widely recognized as the gold standard for evaluating cardiac morphology and function [19]. Raw data for CMR is acquired in the frequency domain (k-space). MR reconstruction from k-space data with high spatio-temporal resolutions throughout the cardiac cycle is an essential step for CMR. Short scan times, ideally within a single breath-hold, are preferable to minimize patient discomfort and reduce potential image artifacts caused by patient motion. Typically, due to the restricted scanning times, only a limited amount of k-space data can be obtained for each temporal frame. Note that while some k-space data are unsampled, the sampled ones are reliable sources of information. However, the Fourier transform of undersampled k-space corrupts a broad region of pixels in the image domain with aliasing artifacts because of violating the Nyquist-Shannon sampling theorem. Previous works [9,15,21,26,30] have attempted to remove the image artifacts primarily by regularizing on the image domain using conventional/learning-based image priors. However, after the Fourier transform the artifacts may completely distort and/or obscure tissues of interest before the image-domain regularizers kick in, making these methods challenging to recover true tissue structures.

On a different principle, k-space interpolation methods first attempt to estimate the full k-space leveraging redundancy in sampled frequency components before Fourier transform. Image domain methods rely on artifacts-specific image priors to denoise the corrupted pixels, making them susceptible to variability in artifact types arising from different undersampling factors. Unlike image domain methods, k-space-based methods have a consistent task of interpolating missing data from reliable sampled ones, even though the undersampling factor may vary. This makes k-space interpolation methods simple, robust and generic over multiple undersampling factors.

In this work, we are interested in learning an entirely k-space-based interpolation for the Cartesian undersampled dynamic MR data. An accurate learnable k-space interpolator can be achieved via (1) *a rich representation of the sampled k-space data*, which can facilitate the exploitation of the limited available samples, and (2) *global dependency modeling of k-space* to interpolate unsampled data from the learned representation. Modeling global dependencies are beneficial because a local structure in the image domain is represented by a wide range of frequency components in k-space. Furthermore, in the context of dynamic MR, the interpolator also has to exploit temporal redundancies.

In the recent past, masked image modeling [11,34] has emerged as a promising method for learning rich generalizable representation by reconstructing the whole image from a masked (undersampled) input. Masked Autoencoders (MAE) [11] are one such model that leverages the *global dependencies* of the undersampled input using Transformers and learns masked-based *rich feature representation.* Despite sharing the same reconstruction principle, MAE has not been explored in k-space interpolation of Cartesian undersampled data. In this work, we cast 2D+t k-space interpolation as a masked signal reconstruction problem and propose a novel Transformer-based method entirely in k-space. Further, we intro-

duce a refinement module on k-space to boost the accuracy of high-frequency interpolation. **Our contributions can be summarized as follows**:
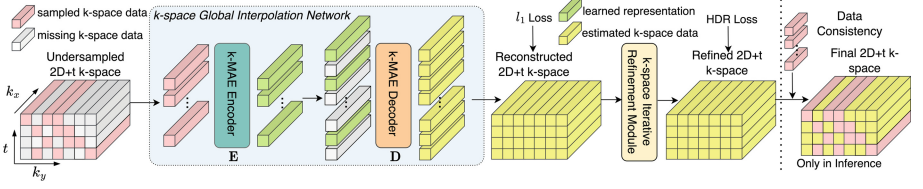
1. We propose a novel k-space Global Interpolation Network, termed k-GIN, leveraging masked image modeling for the first time in k-space. To the best of our knowledge, our work enables the first Transformer-based k-space interpolation for 2D+t MR reconstruction.
2. Next, we propose k-space Iterative Refinement Module, termed k-IRM that refines k-GIN interpolation by efficiently gathering spatio-temporal redundancy of the MR data. Crucially, k-IRM specializes in learning high-frequency details with the aid of customized High-Dynamic-Range (HDR) loss.
3. We evaluate our approach on 92 in-house CMR subjects and compare it to model-based reconstruction baselines using image priors. Our experiments show that the proposed k-space interpolator outperforms baseline methods with superior qualitative and quantitative results. Importantly, our method demonstrates improved robustness and generalizability regarding varying undersampling factors than the model-based counterparts.

2   Related Work

Reconstruction with image priors is broadly used with either image-only denoising [17,31,36] or with model-based approaches to incorporate the k-space consistency by solving an inverse problem. For the latter, the physics-based model can be formulated as a low-rank and a sparse matrix decomposition in CMR [14,26], or a motion-compensated MR reconstruction problem [3,27,28], or data consistency terms with convolution-based [9,30] or Transformers-based [12,18] image regularizers.

k-space-Domain Interpolation. Methods include works [8,22], which have introduced auto-calibration signals (ACS) in the multi-coil k-space center of Cartesian sampled data. RAKI [2,16] uses convolutional networks for optimizing imaging and scanner-specific protocols. Nevertheless, these methods have limited flexibility during the scanning process since they all require a fixed set of ACS data in k-space. [20,32] introduced k-space interpolation methods which do not require calibration signals. [10] proposed a k-space U-Net under residual learning setting. However, these methods heavily rely on local operators such as convolution and may overlook non-local redundancies. [7] uses Transformers applicable only on radial sampled k-space data. However, using masked image modeling with Transformers in k-space for dynamic MR imaging e.g. 2D+t CMR data has not been studied yet.

Hybrid Approaches. Combine information from both k-space and image-domain. KIKI-Net [6] employs an alternating optimization between the image domain and k-space. [25,33] use parallel architectures for k-space and image-domain simultaneously. However, their ablation shows limited contribution coming from the k-space compared to the image domain, implying an under-exploitation of the k-space. Concurrently, [37] use Transformers in k-space but their performance is heavily dependent on the image domain fine-tuning at the final stage.

Fig. 1. The proposed k-space-based dynamic MR reconstruction framework consists of k-space Global Interpolation Network (k-GIN) (see 3.1) and k-space Iterative Refinement Module (k-IRM) for refining the k-GIN interpolation (see 3.2). In the final stage for the inference, we replace k-space estimation at the sampled position with ground-truth k-space values, ensuring the data consistency.

3 Method

Fully sampled complex 2D+t dynamic MR k-space data can be expressed as $\mathbf{y} \in \mathbb{C}^{XYT}$, X and Y are the height (k_x) and the width (k_y) of the k-space matrix, and T is the number of frames along time. In this work, we express k-space as 2 channels (real and imaginary) data $\mathbf{y} \in \mathbb{R}^{2XYT}$. For the MR acquisition, a binary Cartesian sampling mask $M \in \mathbb{Z}^{YT}|M_{ij} \in \{0,1\}$ is applied in the k_y-t plane, i.e. all k-space values along the k_x (readout direction) are sampled if the mask is 1, and remains unsampled if the mask is 0. Figure 1 shows a pictorial undersampled k-space data. Let us denote the collection of sampled k-space lines as $\mathbf{y}_s$ and unsampled lines as $\mathbf{y}_u$. The dynamic MR reconstruction task is to estimate $\mathbf{y}_u$ and reconstruct $\mathbf{y}$ using $\mathbf{y}_s$ only. In this work, we propose a novel Transformer-based reconstruction framework consisting of 1) k-GIN to learn global representation and 2) k-IRM to achieve refined k-space interpolation with a focus on high-frequency components.

3.1 k-space Global Interpolation Network (k-GIN)

In our proposed approach, we work on the k_y-t plane and consider k_x as the channel dimension. Further, we propose each point in the k_y-t plane to be an individual *token*. In total, we have YT number of tokens, out of which YT/R are sampled tokens for an undersampling factor of R. Our objective is to contextualize global dependencies among every sampled token. For that, we use a ViT/MAE [5,11] encoder $\mathbf{E}$ consisting of alternating blocks of multi-head self-attention and multi-layer-perceptrons. The encoder takes advantage of each token's position embedding to correctly attribute its location. Following ViT, we use LayerNorm and GELU activation. We obtain rich feature representation $\mathbf{f}_E = \mathbf{E}(\mathbf{y}_s)$ of the sampled k-space from the encoder.

Next, we want a preliminary estimate of the undersampled k-space data from the learned feature representation. To this end, we employ a decoder $\mathbf{D}$ of similar architecture as the encoder. We initialize all the unsampled tokens $\mathbf{y}_u$ with a single learnable token shared among them. Subsequently, we add their corresponding position embedding to these unsampled tokens. During the decoding
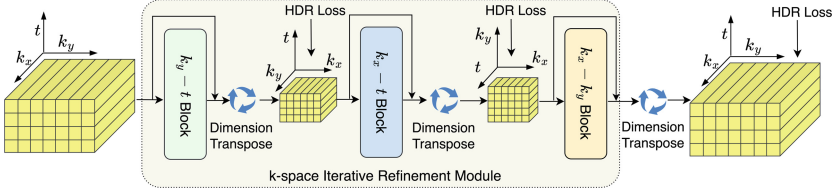
Fig. 2. k-space Iterative Refinement Module (k-IRM) refines high-frequency components of the k-space data (see 3.1). Its refinement Transformer blocks extract the spatio-temporal redundancy by operating the on k_y-t, k_x-t and k_x-k_y plane.

process, the unsampled tokens attend to the well-contextualized features $\mathbf{f}_E$ and produce an estimate of the whole k-space $\hat{\mathbf{y}}_r = \mathbf{D}\left(\left[\mathbf{f}_E, \mathbf{y}_u\right]\right)$. Since our masking pattern includes more sampled data in low-frequency than high-frequency components, we observe k-GIN gradually learn from low-frequency to high-frequency. Note that the imbalance of magnitude in k-space results in more emphasis on low-frequency when ℓ_1 loss is applied between estimation and ground-truth. We leverage this property into the learning behavior of k-GIN and deliberately use ℓ_1 loss between $\hat{\mathbf{y}}_r$ and $\mathbf{y}$, read as $\mathcal{L}_{\ell_1} = \|\hat{\mathbf{y}}_r - \mathbf{y}\|_1$.

3.2 k-space Iterative Refinement Module (k-IRM)

Using ℓ_1 loss in k-GIN makes it focus more on the low-frequency components learning but the high-frequency estimation is still sub-optimal. Inspired by the iterative refinement strategy [38] which is widely used to improve estimation performance, we propose to augment k-GIN's expressive power, especially in high-frequency components, with k-space Iterative Refinement Module. This consists of three Transformer blocks that operate on three orthogonal planes. All three blocks are identical in architecture. The first block operates on k_y-t plane and treats k_x as channel dimension. The second block operates on k_x-t plane and considers k_y as channels, while the final block operates on k_x-k_y plane with t as channel dimension. Note that for the final refinement block uses 4×4 size token while the previous two blocks consider each point as a single token. These configurations enable scalable exploration of the spatio-temporal redundancy present in the output of the k-GIN. We denote $\hat{\mathbf{y}}_1$, $\hat{\mathbf{y}}_2$ and $\hat{\mathbf{y}}_3$ as the estimation after each block in the k-IRM. We apply the skip connection between each refinement block and iteratively minimize the residual error at each stage.

Inspired by [13,24], we applied an approximated logarithm loss function called High-Dynamic Range (HDR) loss for all the stages of k-IRM. HDR loss handles the large magnitude difference in the k-space data and makes the network pay more attention to high-frequency learning. The HDR loss function is defined as: $\mathcal{L}_{\mathrm{HDR}} = \sum_{i=1}^{3} \left\| \frac{\hat{\mathbf{y}}_i - \mathbf{y}}{s(\hat{\mathbf{y}}) + \epsilon} \right\|_2^2$ where $s(\cdot)$ is the stop-gradient operator preventing the network back-propagation of estimation in the denominator and ϵ controls the operational range of the logarithmic approximation.

3.3   Inference

The inference is identical to the training till obtaining the refined output from the k-IRM. Then we replace k-space estimation at the sampled position with ground-truth k-space values, ensuring the data-consistency. Note that this step is not done during the training as it deteriorates learning k-space representation. Once full k-space has been estimated, we use Fourier transform to obtain the image reconstruction during the inference. Note that image reconstruction is not needed during the training since our framework is entirely based in k-space.

4   Data and Experiments

Dataset. The training was performed on 81 subjects (a mix of patients and healthy subjects) of in-house acquired short-axis 2D CINE CMR, whereas testing was carried out on 11 subjects. Data were acquired with 30/34 multiple receiver coils and 2D balanced steady-state free precession sequence on a 1.5T MR (Siemens Aera with TE=1.06 ms, TR=2.12 ms, resolution=1.9×1.9mm^2 with 8mm slice thickness, 8 breath-holds of 15 s duration). The MR data were acquired with a matrix size of 192×156 with 25 temporal cardiac phases (40ms temporal resolution). Afterwards, these data were converted to single-coil MR imaging and k-space data using coil sensitivity maps, simulating a fully sampled single-coil acquisition. A stack of 12 slices along the long axis was collected, resulting in 415/86 image sequence (2D+t) for training/test.

Implementation Details. We use an NVIDIA A6000 GPU to train our framework. The batch size was set to 1 with a one-cycle learning-rate scheduler (max. learning rate 0.0001). We use 8 layers, 8 heads and 512 embedding dimensions for all of our Transformer blocks. We train our network with joint ℓ_1 and HDR-loss with ϵ tuned to 0.5. Training and inference were carried out on retrospectively undersampled images with masks randomly generated by VISTA [1]. We train the network with $R = 4$ undersampled data while we test our method on an undersampled factor $R = 4$, 6 and 8 during the inference. We can use this inference strategy to test our model's generalizability and robustness to different undersampling factors in comparison to the following baseline methods.

Baseline Methods and Metrics. We compare the proposed framework with three single-coil MR reconstruction methods that apply image priors: TV-norm Optimization (TV-Optim) which is widely used in reconstruction [21,23], L+S [26] and DcCNN [30]. TV-Optim reconstructs the image using TV-norm [29] as the image regularizer. L+S leverages compressed sensing techniques and addresses the reconstruction using low rankness and sparsity of the CMR as the image prior, whilst DcCNN employs 3D convolutional neural networks in the image domain together with data-consistency terms. We use the same training and inference strategy for DcCNN to test its model robustness.

Table 1. Quantitative analysis of reconstruction for accelerated CINE CMR (R=4, 6 and 8) using TV-Optim, L+S [26], DcCNN [30] and the proposed method. PSNR (sequence based), SSIM and NMSE are used to evaluate the reconstruction performance. The mean value with the standard deviations are shown. The best results are marked in bold.

Acc R	Methods	NMSE ↓	SSIM ↑	PSNR ↑
4	TV-Optim	0.120 ± 0.031	0.922 ± 0.026	36.743 ± 3.233
	L+S [26]	0.097 ± 0.026	0.949 ± 0.021	39.346 ± 2.911
	DcCNN [30]	$\mathbf{0.087 \pm 0.022}$	0.957 ± 0.019	40.293 ± 2.891
	Proposed	0.088 ± 0.022	$\mathbf{0.958 \pm 0.019}$	$\mathbf{40.368 \pm 3.030}$
6	TV-Optim	0.161 ± 0.047	0.887 ± 0.040	34.066 ± 3.605
	L+S	0.154 ± 0.042	0.901 ± 0.036	34.921 ± 3.174
	DcCNN	0.116 ± 0.029	0.932 ± 0.026	37.666 ± 2.768
	Proposed	$\mathbf{0.109 \pm 0.029}$	$\mathbf{0.940 \pm 0.026}$	$\mathbf{38.461 \pm 3.095}$
8	TV-Optim	0.289 ± 0.078	0.808 ± 0.067	29.052 ± 3.817
	L+S	0.245 ± 0.061	0.826 ± 0.047	30.413 ± 2.888
	DcCNN	0.276 ± 0.047	0.821 ± 0.040	29.778 ± 2.822
	Proposed	$\mathbf{0.151 \pm 0.041}$	$\mathbf{0.904 \pm 0.036}$	$\mathbf{35.674 \pm 3.293}$

We Fourier transform our interpolated full k-space to obtain the image reconstruction and utilize Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM) and Normalized Mean Squared Error (NMSE) to evaluate the reconstruction performance with the baseline quantitatively.

5   Results and Discussion

The quantitative results in Table 1 show consistent superior performance of the proposed method across every single undersampling factor compared to all other baseline methods. Figure 3 shows a qualitative comparison for a typical test sample. It can be seen that the reconstruction methods with image priors can still provide comparable results at $R = 4$, however, suffer from a large performance drop when acceleration rates get higher, especially at $R = 8$. The non-trivial hyper-parameters tuning has to be carried out for L+S and TV-Optim to adapt to the specific image prior at different acceleration factors. It is also noteworthy that the proposed method and DcCNN are both trained only on $R = 4$ undersampled CMR. DcCNN demonstrates inferior reconstruction for $R = 8$ since there is a mismatch in artifact characteristics between $R = 4$ and $R = 8$. On the contrary, the task of interpolating k-space for $R = 4$ and $R = 8$ remains the same, i.e., to estimate missing data from sampled data. We efficiently leverage rich contextualized representation of k-GIN to interpolate full k-space even when a lesser number of sampled k-space data are given as input than seen during training. The observation confirms the superior robustness and generalizability of our proposed framework.
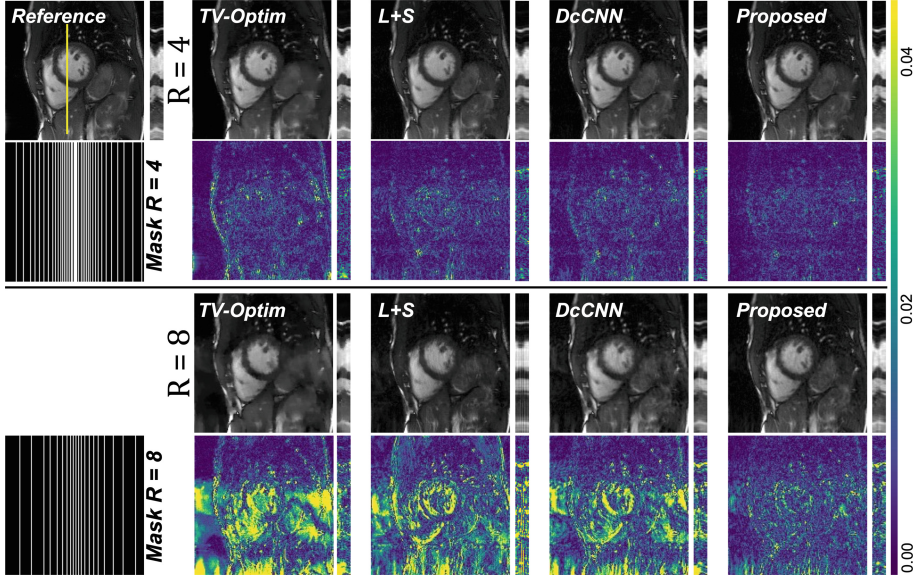
Fig. 3. Qualitative comparison of the proposed method with TV-Optim, L+S and DcCNN in the $R = 4$ and 8 undersampled data. Reference images, undersampling masks, reconstructed (x-y and y-t plane) images and their corresponding error maps are showcased. The selected y-axis is marked with a yellow line in the reference image.

Next, we conduct an ablation study to validate our architectural design. We carry out experiments to investigate the impact of applying k-IRM. We conduct the interpolation using 1) only k-GIN, 2) k-GIN + k_y-t plane refinement, 3) k-GIN + k_x-t plane refinement, 4) k-GIN + k_x-k_y plane refinement and 5) k-GIN with all three refinement blocks. Table 2 in supplementary presents quantitative comparisons amongst five configurations as above. We observe k-IRM offers the best performance when all 3 refinement blocks are used together. In the second ablation, Table 3 in supplementary shows the usefulness of applying ℓ_1 in k-GIN and HDR in k-IRM. HDR makes k-GIN's learning inefficient since HDR deviates from its learning principle of "first low-frequency then high-frequency". On the other hand, $\ell_1 + \ell_1$ combination hinders high-frequency learning.

Outlook. Previous works [6,25] have speculated limited usefulness coming from k-space in a hybrid setting. However, our work presents strong evidence of k-space representation power which can be leveraged in future work with hybrid reconstruction setup. Furthermore, one can utilize our work as a pre-training task since the image reconstruction itself is an "intermediate step" for downstream tasks e.g. cardiac segmentation and disease classification. In the future, one can reuse the learned encoder representation of k-GIN to directly solve downstream tasks without requiring image reconstruction.

Limitation. We also acknowledge some limitations of the work. First, we have not evaluated our method on prospectively collected data, which would be our focus in future work. Second, the current study only investigates the single coil setup due to hardware memory limitations. In the future, we will address the multi-coil scenario by applying more memory-efficient Transformers backbones e.g. [4,35].

6   Conclusion

In this work, we proposed a novel Transformer-based method with mask image modeling to solve the dynamic CMR reconstruction by only interpolating the k-space without any image-domain priors. Our framework leverages Transformers' global dependencies to exploit redundancies in all three k_x-, k_y- and t-domain. Additionally, we proposed a novel refinement module (k-IRM) to boost high-frequency learning in k-space. Together, k-GIN and k-IRM not only produce high-quality k-space interpolation and superior CMR reconstruction but also generalize significantly better than baselines for higher undersampling factors.

References

1. Ahmad, R., Xue, H., Giri, S., Ding, Y., Craft, J., Simonetti, O.P.: Variable density incoherent spatiotemporal acquisition (VISTA) for highly accelerated cardiac MRI. Magn. Reson. Med. **74**(5), 1266–1278 (2015)
2. Akçakaya, M., Moeller, S., Weingärtner, S., Uğurbil, K.: Scan-specific robust artificial-neural-networks for k-space interpolation (RAKI) reconstruction: database-free deep learning for fast imaging. Magn. Reson. Med. **81**(1), 439–453 (2019)
3. Batchelor, P., Atkinson, D., Irarrazaval, P., Hill, D., et al.: Matrix description of general motion correction applied to multishot images. Magn. Reson. Med. **54**, 1273–1280 (2005)
4. Dao, T., Fu, D., Ermon, S., Rudra, A., Ré, C.: FlashAttention: fast and memory-efficient exact attention with IO-awareness. In: Advances in Neural Information Processing Systems, vol. 35, pp. 16344–16359 (2022)
5. Dosovitskiy, A., et al.: An image is worth 16x16 words: transformers for image recognition at scale. arXiv preprint: arXiv:2010.11929 (2020)
6. Eo, T., Jun, Y., Kim, T., Jang, J., Lee, H.J., Hwang, D.: KIKI-Net: cross-domain convolutional neural networks for reconstructing undersampled magnetic resonance images. Magn. Reson. Med. **80**(5), 2188–2201 (2018)
7. Gao, C., Shih, S.F., Finn, J.P., Zhong, X.: A projection-based K-space transformer network for undersampled radial MRI reconstruction with limited training subjects. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) MICCAI 2022. Lecture Notes in Computer Science, vol. 13436. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16446-0_69

8. Griswold, M.A., et al.: Generalized autocalibrating partially parallel acquisitions (GRAPPA). Magn. Reson. Med. **47**(6), 1202–10 (2002)

9. Hammernik, K., Klatzer, T., Kobler, E., et al.: Learning a variational network for reconstruction of accelerated MRI data. Magn. Reson. Med. **79**(6), 3055–3071 (2018)

10. Han, Y., Sunwoo, L., Ye, J.C.: k-space deep learning for accelerated MRI. IEEE TMI **39**, 377–386 (2019)

11. He, K., Chen, X., Xie, S., Li, Y., Dollár, P., Girshick, R.: Masked autoencoders are scalable vision learners (2021). arXiv:2111.06377

12. Huang, J., Fang, Y., Wu, Y., Wu, H., et al.: Swin transformer for fast MRI. Neurocomputing **493**, 281–304 (2022)

13. Huang, W., Li, H.B., Pan, J., Cruz, G., Rueckert, D., Hammernik, K.: Neural implicit k-space for binning-free non-cartesian cardiac MR imaging. In: Frangi, A., de Bruijne, M., Wassermann, D., Navab, N. (eds.) IPMI 2023. Lecture Notes in Computer Science, vol. 13939. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-34048-2_42

14. Huang, W., Ke, Z., Cui, Z.X., et al.: Deep low-rank plus sparse network for dynamic MR imaging. Med. Image Anal. **73**, 102190 (2021)

15. Jin, K.H., McCann, M.T., Froustey, E., Unser, M.: Deep convolutional neural network for inverse problems in imaging. IEEE TIP **26**(9), 4509–4522 (2017)

16. Kim, T., Garg, P., Haldar, J.: LORAKI: autocalibrated recurrent neural networks for autoregressive MRI reconstruction in k-space. arXiv preprint: arXiv:1904.09390 (2019)

17. Kofler, A., Dewey, M., Schaeffter, T., Wald, C., Kolbitsch, C.: Spatio-temporal deep learning-based undersampling artefact reduction for 2D radial cine MRI with limited training data. IEEE TMI **39**, 703–717 (2019)

18. Korkmaz, Y., et al.: Unsupervised MRI reconstruction via zero-shot learned adversarial transformers. IEEE TMI **41**(7), 1747–1763 (2022)

19. Lee, D., et al.: The growth and evolution of cardiovascular magnetic resonance: a 20-year history of the society for cardiovascular magnetic resonance (SCMR) annual scientific sessions. J. Cardiovasc. Magn. Reson. **20**(1), 1–11 (2018)

20. Lee, J., Jin, K., Ye, J.: Reference-free single-pass EPI Nyquist ghost correction using annihilating filter-based low rank Hankel matrix (ALOHA). Magn. Reson. Med. **76**(6), 1775–1789 (2016)

21. Lingala, S.G., Hu, Y., DiBella, E., Jacob, M.: Accelerated dynamic MRI exploiting sparsity and low-rank structure: k-t SLR. IEEE TMI **30**(5), 1042–1054 (2011)

22. Lustig, M., Pauly, J.M.: SPIRiT: iterative self-consistent parallel imaging reconstruction from arbitrary k-space. Magn. Reson. Med. **64**(2), 457–471 (2010)

23. Lustig, M., Donoho, D., Pauly, J.M.: Sparse MRI: the application of compressed sensing for rapid MR imaging. Magn. Reson. Med. **58**(6), 1182–1195 (2007)

24. Mildenhall, B., Hedman, P., Martin-Brualla, R., Srinivasan, P.P., Barron, J.T.: NeRF in the dark: high dynamic range view synthesis from noisy raw images. In: CVPR, pp. 16169–16178 (2022)

25. Nitski, O., Nag, S., McIntosh, C., Wang, B.: CDF-Net: cross-domain fusion network for accelerated MRI reconstruction. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12262, pp. 421–430. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59713-9_41

26. Otazo, R., Candès, E., Sodickson, D.K.: Low-rank plus sparse matrix decomposition for accelerated dynamic MRI with separation of background and dynamic components. Magn. Reson. Med. **73**(3), 1125–1136 (2015)

27. Pan, J., Huang, W., Rueckert, D., Küstner, T., Hammernik, K.: Reconstruction-driven motion estimation for motion-compensated MR cine imaging. arXiv preprint: arXiv:2302.02504 (2023)

28. Pan, J., Rueckert, D., Kustner, T., Hammernik, K.: Learning-based and unrolled motion-compensated reconstruction for cardiac MR CINE imaging. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Speidel, S. (eds.) MICCAI 2022. Lecture Notes in Computer Science, vol. 13436. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16446-0_65

29. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D **60**(1), 259–268 (1992)

30. Schlemper, J., et al.: A deep cascade of convolutional neural networks for dynamic MR image reconstruction. IEEE TMI **37**(2), 491–503 (2018)

31. Shi, J., Liu, Q., Wang, C., Zhang, Q., Ying, S., Xu, H.: Super-resolution reconstruction of MR image with a novel residual learning network algorithm. Phys. Med. Biol. **63**(8), 085011 (2018)

32. Shin, P.J.: Calibrationless parallel imaging reconstruction based on structured low-rank matrix completion. Magn. Reson. Med. **72**(4), 959–970 (2014)

33. Singh, N.M., Iglesias, J.E., Adalsteinsson, E., Dalca, A.V., Golland, P.: Joint frequency- and image-space learning for fourier imaging. Mach. Learn. Biomed. Imaging (2022)

34. Xie, Z., et al.: SimMIM: a simple framework for masked image modeling. In: CVPR (2022)

35. Xiong, Y., et al.: Nyströmformer: a nyström-based algorithm for approximating self-attention. In: AAAI, vol. 35, pp. 14138–14148 (2021)

36. Yang, G., Yu, S., Dong, H., Slabaugh, G.G., et al.: DAGAN: deep de-aliasing generative adversarial networks for fast compressed sensing MRI reconstruction. IEEE TMI **37**(6), 1310–1321 (2018)

37. Zhao, Z., Zhang, T., Xie, T., Wang, Y., Zhang, Y.: K-space transformer for under-sampled MRI reconstruction. In: BMVC (2022)

38. Zhu, X., Su, W., Lu, L., Li, B., Wang, X., Dai, J.: Deformable DETR: deformable transformers for end-to-end object detection. arXiv preprint: arXiv:2010.04159 (2020)