Deep MRI Reconstruction with Generative Vision Transformers

Yilmaz Korkmaz[1,2], Mahmut Yurt[1,2], Salman Ul Hassan Dar[1,2], Muzaffer Özbey[1,2], and Tolga Cukur[1,2(✉)]

[1] Department of Electrical and Electronics Engineering, Bilkent University, Ankara, Turkey
cukur@ee.bilkent.edu.tr
[2] National Magnetic Resonance Research Center (UMRAM), Bilkent University, Ankara, Turkey

Abstract. Supervised training of deep network models for MRI reconstruction requires access to large databases of fully-sampled MRI acquisitions. To alleviate dependency on costly databases, unsupervised learning strategies have received interest. A powerful framework that eliminates the need for training data altogether is the deep image prior (DIP). To do this, DIP inverts randomly-initialized models to infer network parameters most consistent with the undersampled test data. However, existing DIP methods leverage convolutional backbones, suffering from limited sensitivity to long-range spatial dependencies and thereby poor model invertibility. To address these limitations, here we propose an unsupervised MRI reconstruction based on a novel generative vision transformer (GVTrans). GVTrans progressively maps low-dimensional noise and latent variables onto MR images via cascaded blocks of cross-attention vision transformers. Cross-attention mechanism between latents and image features serve to enhance representational learning of local and global context. Meanwhile, latent and noise injections at each network layer permit fine control of generated image features, improving model invertibility. Demonstrations are performed for scan-specific reconstruction of brain MRI data at multiple contrasts and acceleration factors. GVTrans yields superior performance to state-of-the-art generative models based on convolutional neural networks (CNNs).

Keywords: MRI reconstruction · Transformer · Generative · Attention · Unsupervised

1 Introduction

Magnetic resonance imaging (MRI) is pervasive in non-invasive assessment of tissue morphology. However, its inherently slow acquisition process limits practical utility in many clinical applications, so there is emergent interest in accelerated MRI methods. Deep neural networks (DNN) have revolutionized accelerated MRI by offering state-of-the-art reconstruction performance

N. Haq et al. (Eds.): MLMIR 2021, LNCS 12964, pp. 54–64, 2021.
https://doi.org/10.1007/978-3-030-88552-6_6

2 Theory

2.1 Deep Unsupervised MRI Reconstruction

MRI acquisitions can be accelerated via undersampling in the Fourier domain (i.e., k-space):

$$F_u C m = y_s \tag{1}$$

where F_u is the partial Fourier operator defined on the set of sampled k-space locations, C denotes sensitivity maps of coil elements, m is the underlying MR image and y_s are collected k-space data. To reconstruct the image m given y_s, the underdetermined system in Eq. 1 must be solved. To improve the conditioning of the problem, prior information on the MR image is incorporated as a regularization term:

$$\widehat{m} = \underset{m}{\operatorname{argmin}} \|y_s - F_u C m\|_2^2 + H(m) \tag{2}$$

where $\widehat{m}$ is the reconstruction, and $H(m)$ is the regularization. In deep learning reconstructions, the regularization function that is typically implemented as a projection through a CNN architecture that suppresses aliasing artifacts. In the supervised learning setup, model training is performed on a large dataset of fully-sampled ground truth MRI data. The CNN weights are learned to effectively map undersampled MRI data onto high-quality MR images that resemble ground truth data as closely as possible.

To mitigate reliance on fully-sampled ground truths, the deep image prior (DIP) framework observes that CNNs performing filtering with local kernels can serve as native image regularizers. As such, DIP-based reconstructions randomly initialize the network inputs and weights. Without any training, inference is performed directly on each given test subject starting with the untrained network model. In the case of MRI, to ensure that the network output maintains fidelity to the physical signal model, network inputs and weights are adapted to ensure maximal consistency to the acquired k-space data [3,11]. This process is known as model inversion, and the resulting reconstruction can then be formulated as:

$$\theta^* = \underset{\theta}{\operatorname{argmin}} \|F_u C d_\theta(z) - y_s\|_1 \tag{3}$$

where θ are network weights, z are latent variables, $d_\theta(z)$ is the projection from latents onto the reconstruction. Network weights and latents are randomly initialized, and the optimization in Eq. 3 is performed over θ, while z is fixed. The reconstructed image can be obtained as:

$$\widehat{m} = d_{\theta^*}(z) \tag{4}$$

2.2 Generative Vision Transformers

DIP-type MRI reconstruction eliminates the need to collect experimentally costly ground truth acquisitions, and model pre-training on large databases.

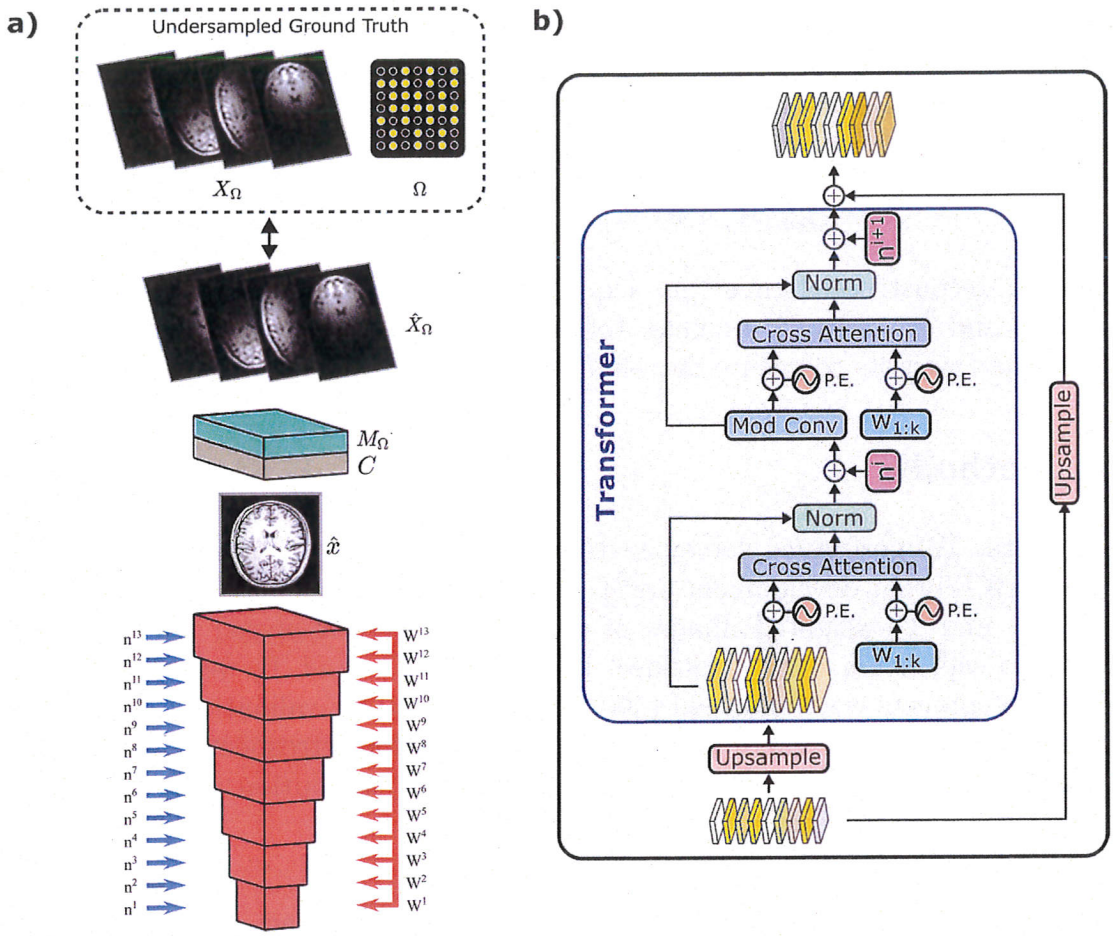

Fig. 1. GVTrans is a deep generative network that maps noise variables (n) and latents (W) onto high-quality MR images. The multi-layer architecture progressive increases image resolution. Within each layer, upsampled feature maps are input to a cross-attention transformer module (see the right panel). For inference on test data, the generated MR images are masked with the same sampling pattern as in the under-sampled acquisition. Network parameters are optimized to ensure consistency between reconstructed and original k-space samples.

affine-transformed global latent variables ($w_s^i \in \mathbb{R}^u$ derived from w_g). To do this, scaled convolution kernels θ_G^i are used to filter the feature maps:

$$X''^i = \begin{bmatrix} \sum_s X'^{i-1,s} \circledast \theta_G^{i,s,1} \\ \vdots \\ \sum_s X'^{i-1,s} \circledast \theta_G^{i,s,v} \end{bmatrix} + \begin{bmatrix} \alpha^{i,1} n^{i,1} \\ \vdots \\ \alpha^{i,v} n^{i,v} \end{bmatrix} \tag{7}$$

where $\theta_G^{i,u',v'} \in \mathbb{R}^{r \times r}$ is the convolution kernel for the $u'th$ input channel and $v'th$ output channel, and s is the channel index. Furthermore, noise variables are injected onto feature maps $n^{i,v'} \in \mathbb{R}^{m \times h}$ is spatially-varying noise on the $v'th$ channel of ith layer and $\alpha^{i,v}$ is a learnable scalar.

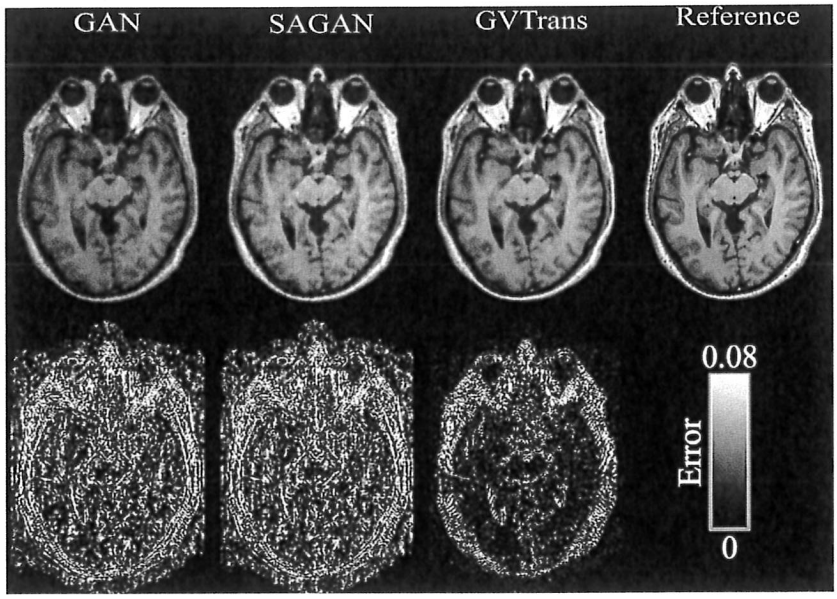

Fig. 2. Representative T_1-weighted MRI reconstructions and respective error maps in IXI at R = 4. Results are shown along with the reference image.

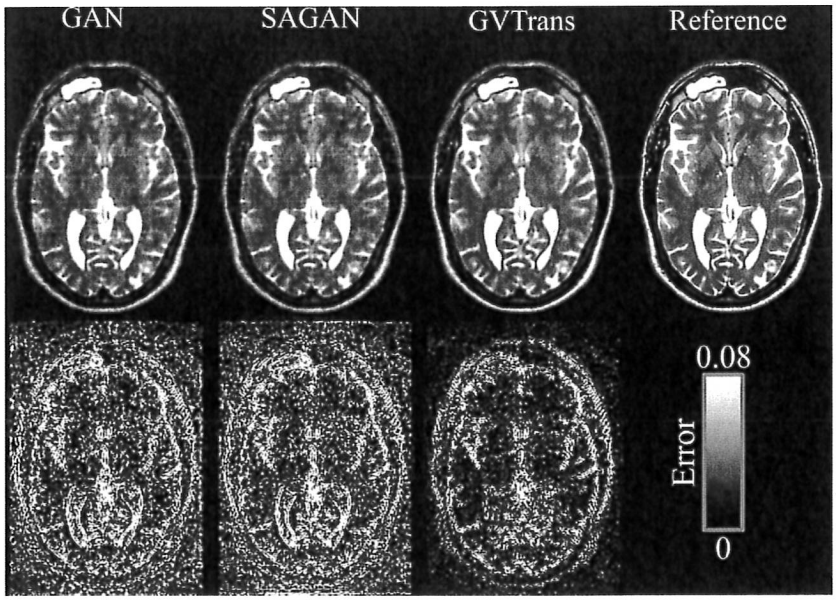

Fig. 3. Representative T_2-weighted MRI reconstructions and respective error maps in IXI at R = 4. Results are shown for along with the reference image.

Performance Evaluation. Reconstruction quality was evaluated by measuring peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) between the reconstructed and reference fully-sampled MR images. In Tables, metrics are reported as mean ± std across test subjects. Statistical differences between reconstruction methods were assessed via Wilcoxon signed-rank tests.

5 Discussion

Here, we introduced an unsupervised reconstruction model based on generative vision transformers. GVTrans can perform DIP-type reconstructions, so it bypasses the need for a priori model training. Instead, randomly initialized models are inverted to optimize network parameters that maximize consistency of the reconstruction to the acquired k-space data. While DIP-type reconstructions have longer inference times compared to reconstructions with offline-trained network models, the scan-specific nature of GVTrans can improve generalizability. Furthermore, as opposed to classical models backboned with CNN architectures that only preserve local dependencies, GVTrans leverages vision transformers to capture long-range dependencies. In the reported experiments, GVTrans was observed to significantly improve image quality compared to DIP reconstructions implemented via CNN-based models with and without self-attention blocks. These results suggest that MRI reconstructions can benefits from incorporation of broader spatial context across images beyond only local context.

6 Conclusion

In this study, we introduced a novel unsupervised MRI reconstruction approach by embedding vision transformers into a generative network, and performing scan-specific reconstructions as inspired by the deep image prior framework. GVTrans offers improved image quality compared to CNN-based reconstructions with and without self-attention mechanisms, and it can flexibly adapt its model to individual test subjects. Therefore, GVTrans is a promising candidate for improving the applicability and generalizability of deep MRI reconstructions.

References

1. Adler, J., Öktem, O.: Learned primal-dual reconstruction. IEEE Trans. Med. Imaging **37**(6), 1322–1332 (2018)
2. Aggarwal, H.K., Mani, M.P., Jacob, M.: MoDL: model-based deep learning architecture for inverse problems. IEEE Trans. Med. Imaging **38**(2), 394–405 (2019)
3. Biswas, S., Aggarwal, H.K., Jacob, M.: Dynamic MRI using model-based deep learning and SToRM priors: MoDL-SToRM. Magn. Reson. Med. **82**(1), 485–494 (2019)
4. Dar, S.U.H., Yurt, M., Shahdloo, M., Ildız, M.E., Tınaz, B., Çukur, T.: Prior-guided image reconstruction for accelerated multi-contrast MRI via generative adversarial networks. IEEE J. Sel. Top. Sig. Process. **14**(6), 1072–1087 (2020)
5. Dar, S.U.H., Özbey, M., Çatlı, A.B., Çukur, T.: A transfer-learning approach for accelerated MRI using deep neural networks. Magn. Reson. Med. **84**(2), 663–685 (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. Gabbay, A., Hoshen, Y.: Style generator inversion for image enhancement and animation. arXiv preprint arXiv:1906.11880 (2019)

25. Wang, S., et al.: Accelerating magnetic resonance imaging via deep learning. In: IEEE 13th International Symposium on Biomedical Imaging (ISBI), pp. 514–517 (2016). https://doi.org/10.1109/ISBI.2016.7493320
26. Yaman, B., Hosseini, S.A.H., Moeller, S., Ellermann, J., Uğurbil, K., Akçakaya, M.: Self-supervised learning of physics-guided reconstruction neural networks without fully sampled reference data. Magn. Reson. Med. **84**(6), 3172–3191 (2020)
27. Yu, S., et al.: DAGAN: deep de-aliasing generative adversarial networks for fast compressed sensing MRI reconstruction. IEEE Trans. Med. Imaging **37**(6), 1310–1321 (2018)
28. Zhang, H., Goodfellow, I., Metaxas, D., Odena, A.: Self-attention generative adversarial networks. In: Proceedings of the 36th International Conference on Machine Learning, pp. 7354–7363 (2019)
29. Zhu, B., Liu, J.Z., Rosen, B.R., Rosen, M.S.: Image reconstruction by domain transform manifold learning. Nature **555**(7697), 487–492 (2018)