Decomposition-Based Variational Network for Multi-Contrast MRI Super-Resolution and Reconstruction (Supplementary material)

Pengcheng Lei[1] Faming Fang[1,2*] Guixu Zhang[1] Tieyong Zeng[3]
[1]School of Computer Science and Technology, East China Normal University
[2]KLATASDS-MOE [3]Department of Mathematics, Chinese University of Hong Kong
pengchenglei1995@163.com {fmfang,gxzhang}@cs.ecnu.edu.cn zeng@math.cuhk.edu.hk

Figure 1. Visualization of the undersampling mask M: (a) the center cropping mask used in the MRI SR task; and (b) the cartesian sampling mask used in the MRI reconstruction task.

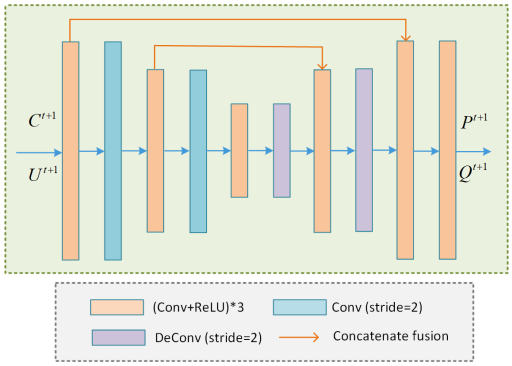

Figure 2. The structure of the modified U-Net for denoising.

1. Undersampling mask

In figure 1, we visualize the center cropping mask and the cartesian sampling mask used in our model. Figure (a) is the center cropping mask [2, 4] used in the MRI SR task, and figure (b) is the cartesian sampling mask used in the MRI reconstruction task.

2. The denoising U-Net

In this paper, we employ a variant of U-net as our deep denoising network. As shown in Figure 2, our denoising U-Net consists of three encoder blocks and three decoder blocks. The first two encoder blocks contain three "Conv-ReLU" layers and a "Conv" layer with stride 2 for downsampling. The third encoder block contains three "Conv-ReLU" layers. The first two decoder blocks contain a "Deconv" layer for upsampling and three "Conv-ReLU" layers. The third encoder block contains three "Conv-ReLU" layers. Besides, skip connections are used to fuse the information between the encoders and decoders. To reduce the number of network parameters and the effect of overfitting, the denoising networks in different iteration stages share the same network parameters.

*Corresponding author

Table 1. The number of network parameters and average testing time of recent SOTA multi-contrast MRI SR and reconstruction methods.

Task	Method	Parameters	Testing Time
SR	CUNet [1]	0.2M	51ms
	MCSR [7]	3.5M	18ms
	MASA [6]	4.0M	52ms
	MINet [2]	11.9M	80ms
	MCMRSR [4]	3.5M	250ms
Rec	MDUNet [11]	5.2M	23ms
	MTrans [3]	86.1M	47ms
	Restormer [12]	26.1M	83ms
	Ours-S	1.4M	41ms
	Ours-L	5.7M	93ms

3. Model efficiency analysis

In our main manuscript, we have compared the performance and the parameter numbers of our methods with existing SOTA methods. In Table 1, we further report the running times of them. The running times are tested on an RTX3090 GPU. Compared with MINet [2], MCMRSC [4] and Restormer [12], our MC-VarNet-S has advantages in terms of performance, number of parameters, and computation. Since our method is based on iterative unrolling, the running time increases with the number of iterations. MCSR [7] and MTrans [3] run faster than our MC-VarNet-S, however, our model has great advantages in terms of

model performance and the number of parameters. Overall, our model can better balance the model performance, the number of parameters, and the computational complexity compared with existing SOTA methods.

4. MC-VarNet without decomposition

In our ablation experiments, we compare our model with the model without decomposition. The model without decomposition can be presented as:

$$\begin{aligned}\underset{\{x_1,C\}}{\operatorname{argmin}}&\frac{1}{2}\|M\mathcal{F}(x_1)-k_{x_1}\|_2^2+\frac{1}{2}\|C-\hat{x}_2\|_2^2\\&+\frac{\gamma}{2}\|Ax_1-BC\|_2^2+\lambda_1\mathcal{R}(x_1)+\lambda_2\psi_2(C).\end{aligned}\tag{1}$$

In this model, we directly extract common features from the reference image. Using the half-quadratic splitting algorithm, the model can be reformulated as:

$$\begin{aligned}\underset{\{x_1,C,P\}}{\operatorname{argmin}}&\frac{1}{2}\|M\mathcal{F}(x_1)-k_{x_1}\|_2^2+\frac{1}{2}\|C-\hat{x}_2\|_2^2\\&+\frac{\gamma}{2}\|Ax_1-BC\|_2^2+\frac{\alpha}{2}\|P-C\|_2^2\\&+\lambda_1\mathcal{R}(x_1)+\lambda_2\psi_2(P).\end{aligned}\tag{2}$$

This model can be solved via the following iterative scheme:

$$\begin{aligned}x_1^{t+1}=\underset{x_1}{\operatorname{argmin}}&\frac{1}{2}\|M\mathcal{F}(x_1)-k_{x_1}\|_2^2+\frac{\gamma^t}{2}\|Ax_1-BC^t\|_2^2\\&+\lambda_1\mathcal{R}(x_1),\end{aligned}\tag{3}$$

$$\begin{aligned}C^{t+1}=\underset{C}{\operatorname{argmin}}&\frac{1}{2}\|C-\hat{x}_2\|_2^2+\frac{\gamma^t}{2}\|Ax_1^t-BC\|_2^2\\&+\frac{\alpha^t}{2}\|P^t-C\|_2^2,\end{aligned}\tag{4}$$

$$P^{t+1}=\underset{P}{\operatorname{argmin}}\frac{\alpha}{2}\|P-C^{t+1}\|_2^2+\lambda_2\psi_2(P).\tag{5}$$

Similar to our main manuscript, we can get the solutions of these sub-problems as:

$$\begin{aligned}x_1^{t+1}=&x_1^t-\mu^tF^*(F(x_1^t)-k_{x_1})\\&-\mu^t(\gamma^tA^\top(Ax_1^t-BC^t)+\lambda_1\phi(x_1^t)),\end{aligned}\tag{6}$$

$$C^{t+1}=(I+\gamma^tB^\top B+\alpha^tI)^{-1}(\hat{x}_2+\gamma^tB^\top Ax_1^t+\alpha^tP^t),\tag{7}$$

$$P^{t+1}=\mathcal{N}_C(C^{t+1};\Theta_C).\tag{8}$$

Then we unfold these iterations into a deep network as we have done in our main manuscript. Since this model has only one auxiliary variable, its number of parameters is fewer than our main model. For a fair comparison, we enlarge the denoising U-Net to ensure the two models have a similar number of parameters. As shown in Table 3 of our main manuscript, the model with decomposition achieves the best performance than the others. Compared with the model without decomposition, our model separates the reference image into a common component and a unique component, avoiding the interference of inconsistent information (contrast information and some unaligned structures) on the target image. We think this is the key to the good performance of our model.

In figure 3, we compare the reconstruction errors of our model with different multi-contrast fusion strategies. Especially, we highlight the reconstruction results for regions where the reference image and target image are inconsistent. From the figure, we can find that our reconstruction result has fewer reconstruction errors in those inconsistent areas. It indicates that our decomposition-based model has advantages in dealing with inconsistent cases.

5. Analysis of MC-VarNet model

Analysis of the channel expansion operation. In this paper, we propose to expand the channel numbers of the input images from 1 to c_i and reconstruct c_i images simultaneously in the iteration stages. Similar operations can be found in [8, 10]. To be specific, in our model, we copy each image along the channel dimension to expand the channel numbers. In the DCL of our model, we perform the data-consistency operation on the c_i reconstruction images simultaneously. In Table 4 of our main manuscript, we have compared the model with and without channel expansion operation and the significant performance improvements fully demonstrate its effectiveness. Next, we will analyze the feasibility of this operation on model-based methods. There is a common sense in the DCNN model that the more channels of the intermediate features, the better performance of the model. Model-based methods need to reconstruct an intermediate image in each iteration stage. From the perspective of an end-to-end network, the channel reduction in the intermediate step will result in information loss, which is unfavorable to the final reconstruction results. By contrast, our channel expansion operation avoids this problem and successfully combines model-based prior information with powerful deep neural networks.

Compared with DCNN methods: Existing DCNN-based multi-contrast MRI SR and reconstruction methods often manually design fusion rules to fuse the multi-contrast information. By contrast, our MC-VarNet model is constructed under the guidance of the optimization algorithm with a well-designed data fidelity term. As we mentioned above, the multi-contrast MR images are generated by different scan settings, thus they usually have their unique contrast information and some inconsistent structure information. Simple fusion strategies may transfer useless information from the reference image to the target image, thus af-

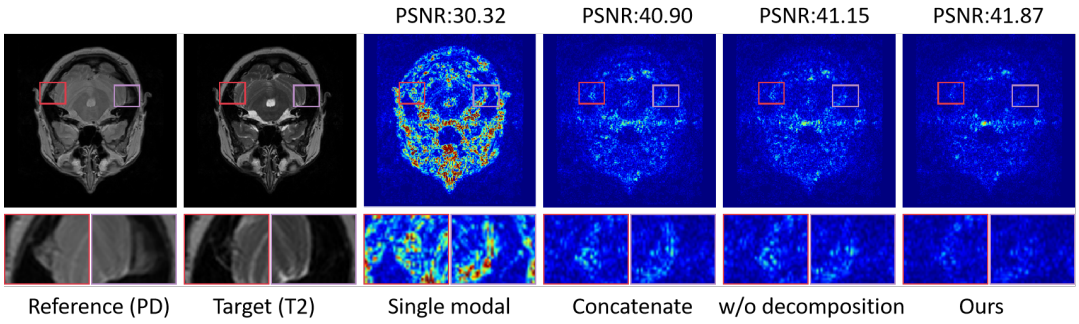

Figure 3. The reconstruction error comparison of our model with different multi-contrast fusion strategies.

fecting the reconstruction of the target image. Considering that, we decompose the multi-contrast images into common components and unique components, and only the common components are used to guide the target images. Compared with existing DCNN-based methods, our model develops priors and constraints plugged into the model according to the characteristics of multi-contrast MR images. By unfolding the iterative solutions into deep networks, the proposed model is not only interpretable but also powerful. We think that's why our approach is better than existing methods.

Visualization of the intermediate components. In Figure 4, we visualize the intermediate reconstruction results of our model on the IXI test set for multi-contrast SR and reconstruction tasks. Since we reconstruct c_i images in the intermediate stages, we use the weighted average layer (WAL) to compress them into one channel and visualize them. As one can see, with the increase in the number of iterations, the PSNR of the reconstruction results gradually increased. The decomposed common components are mainly the consistent high-frequency details extrasted from the reference image. The decompoased unique components are mainly the low-frequency contrast information and some inconsistent structure information.

6. Experimental settings

In this section, we will introduce the experimental details of the comparison methods.

Multi-contrast guided SR. We compare our MC-VarNet model with various MRI guided-SR methods, including MCSR [7], CUNet [1][1], MASA [6][2], MINet [2][3] and MCMRSR [4][4]. Note that MCSR and MINet are two CNN-based MRI guided-SR methods. MASA is a SOTA method for natural images guided-SR. CUNet is a deep-unfolding multi-modal restoration method. MCMRSR is a transformer-based MRI guided-SR method. All these models are trained 50 epochs with a learning rate of 1×10^{-4} using the Adam optimizer. It should be noticed that when reproducing the methods MASA and MCMRSR on the BrainTS dataset, we pad the input images from 240×240 to 256×256, since their input sizes are required to be a multiple of 64. This may be the reason for the poor performance of these two methods on the BrainTS dataset.

Multi-contrast guided reconstruction. We compare our MC-VarNet model with zero-filling (ZF), UNet [9], MDUNet [11], MUSC [5][5], MTrans [3][6] and Restormer [12][7]. Specifically, UNet is a CNN-based method, and MUSC is a deep unfolding multi-scale CDic model. Restormer is a SOTA transformer-based reconstruction method. MDUNet is a multi-contrast MRI reconstruction method. All these models are trained 50 epochs with a learning rate of 1×10^{-4} using adam optimizer. MTrans is a transformer-based multi-contrast MRI reconstruction method. For the MTrans model, we set the window size as 8 on the two contrast images. Following their original paper, we use the SGD optimizer with a learning rate of 1×10^{-4} and train the model 50 epochs.

[1] https://github.com/cindydeng1991/TPAMI-CU-Net
[2] https://github.com/dvlab-research/MASA-SR
[3] https://github.com/chunmeifeng/MINet
[4] https://github.com/XAIMI-Lab/McMRSR
[5] https://github.com/liutianlin0121/MUSC
[6] https://github.com/chunmeifeng/MTrans
[7] https://github.com/swz30/Restormer

References

[1] Xin Deng and Pier Luigi Dragotti. Deep convolutional neural network for multi-modal image restoration and fusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(10):3333–3348, 2020.

[2] Chun-Mei Feng, Huazhu Fu, Shuhao Yuan, and Yong Xu. Multi-contrast MRI super-resolution via a multi-stage integration network. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 140–149. Springer, 2021.

[3] Chun-Mei Feng, Yunlu Yan, Geng Chen, Yong Xu, Ying Hu, Ling Shao, and Huazhu Fu. Multi-modal transformer for accelerated MR imaging. *IEEE Transactions on Medical Imaging*, pages 1–1, 2022.

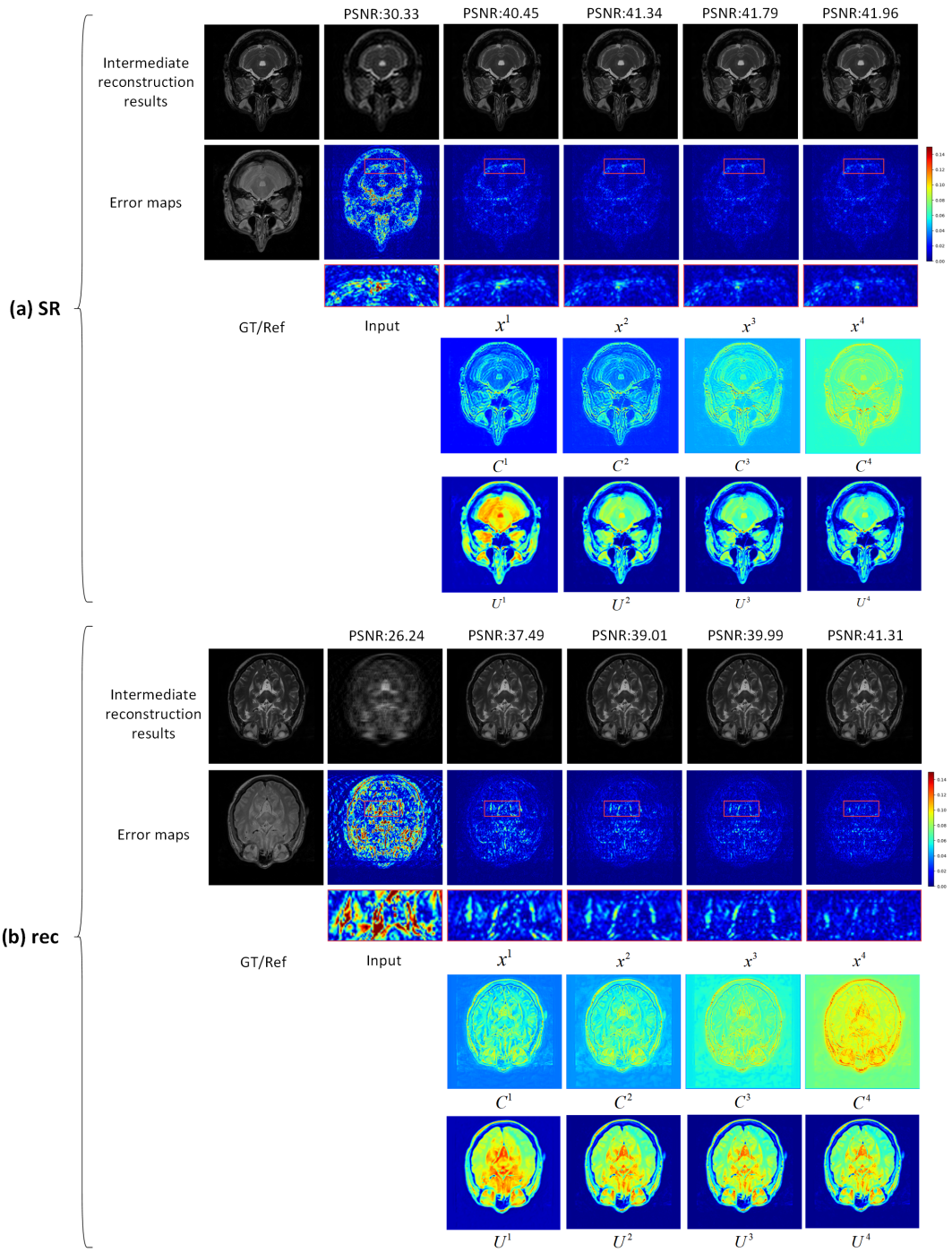

Figure 4. Visualization of the intermediate reconstruction results ($T \in \{1, 4\}$) on the IXI test set for multi-contrast (a) SR and (b) reconstruction tasks.

[4] Guangyuan Li, Jun Lv, Yapeng Tian, Qi Dou, Chengyan Wang, Chenliang Xu, and Jing Qin. Transformer-empowered multi-scale contextual matching and aggregation for multi-contrast MRI super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20636–20645, 2022.

[5] Tianlin Liu, Anadi Chaman, David Belius, and Ivan Dokmanić. Learning multiscale convolutional dictionaries for image reconstruction. *IEEE Transactions on Computational Imaging*, 8:425–437, 2022.

[6] Liying Lu, Wenbo Li, Xin Tao, Jiangbo Lu, and Jiaya Jia. MASA-SR: Matching acceleration and spatial adaptation for

reference-based image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6368–6377, 2021.

[7] Qing Lyu, Hongming Shan, Cole Steber, Corbin Helis, Chris Whitlow, Michael Chan, and Ge Wang. Multi-contrast super-resolution MRI through a progressive network. *IEEE Transactions on Medical Imaging*, 39(9):2738–2749, 2020.

[8] Cong Quan, Jinjie Zhou, Yuanzheng Zhu, Yang Chen, Shanshan Wang, Dong Liang, and Qiegen Liu. Homotopic gradients of generative density priors for mr image reconstruction. *IEEE Transactions on Medical Imaging*, 40(12):3265–3278, 2021.

[9] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 234–241. Springer, 2015.

[10] Hong Wang, Yuexiang Li, Deyu Meng, and Yefeng Zheng. Adaptive convolutional dictionary network for CT metal artifact reduction. In *The 31st International Joint Conference on Artificial Intelligence*, pages 1401–1407, 2022.

[11] Lei Xiang, Yong Chen, Weitang Chang, Yiqiang Zhan, Weili Lin, Qian Wang, and Dinggang Shen. Deep-learning-based multi-modal fusion for fast MR reconstruction. *IEEE Transactions on Biomedical Engineering*, 66(7):2105–2114, 2018.

[12] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5728–5739, 2022.