A APPENDIX

Our main manuscript presents PIRATE and PIRATE+ as the first PnP based method in DIR. Our experimental results show the effectiveness of the architectures of PIRATE and PIRATE+ and demonstrate their state-of-the-art performance. The additional materials in this section supplement our experimental results and further support the conclusions mentioned above. In appendix, we provide the visualization of warped images and segmentation masks on CANDI dataset as a supplement to the results on OASIS dataset in main manuscript. We also provide additional results on the EMPIRE10 lung CT dataset. In addition to the image similarity and DSC used in main manuscript, we illustrate the warped grids on OASIS and CANDI datasets to further evaluate the performance using the smoothness of registration fields. We design an experiment by using total variation (TV) denoiser (Rudin et al., 1992) in PIRATE to demonstrate the compatibility of PIRATE with non-learning based denoisers. Moreover, we show the visualization of warped images and segmentation masks from PIRATE using dynamic and fixed step size on OASIS dataset, which matches the numerical results in main manuscript. To further evaluate our methods, we include the analysis of computational efficiency and scalability. To clarify the acronyms, we provide a table of acronyms and corresponding full names in this paper. We also provide the corner case of our method and the number of function evaluation(NFE) along with the training iterations of PIRATE+.

Fig. 5 shows the visual results for warped images and correlated warped segmentation masks from PIRATE+ and benchmarks on the CANDI dataset. In the top and bottom row, the results of PIRATE+ in ROI are more consistent with the fixed image with fewer artifacts than other benchmarks.

Table. 4 shows the numerical results of PIRATE, PIRATE+ and benchmarks on the EMPIRE10 lung CT dataset. PIRATE+ shows superior performance compared with other benchmarks. The results of PIRATE show that the pre-trained AWGN denoiser can provide competitive results without refinement of the DEQ.

Fig. 6 and Fig. 7 show the illustration of warped images and correlated warped segmentation masks from PIRATE+ and benchmarks on the EMPIRE10 lung CT dataset. In both warped images and segmentation masks, the results of PIRATE+ are more consistent with the reference than other benchmarks.

Fig. 8 shows the visual results of warped grid from OASIS and CANDI datasets of PIRATE and different baselines. The results show that PIRATE maintains the smoothness of the registration fields.

Fig. 9 shows the visual results of using dynamic step size and fixed step size from OASIS and CANDI datasets. It shows that dynamic step size can provide results more similar to the fixed images and segmentation masks with less artifact.

Fig. 10 shows the visual results of using TV denoiser and DnCNN denoiser from OASIS dataset. Table 5 shows the numerical result of using TV denoiser and DnCNN denoiser from OASIS and CANDI datasets. The results show that TV denoiser can still provide reasonable result, although with some performance degradation.

Table. 6 shows the numerical results of memory cost, running time in training and inference for PIRATE, PIRATE+, and benchmarks on OASIS and CANDI datasets. Table. 6 shows that PIRATE and PIRATE+ has lower memory complexity compared to other DL baselines.

Fig. 11 shows the visual results of memory cost and inference time of PIRATE and PIRATE+ with different size of inputs. Fig. 11 shows that PIRATE and PIRATE+ have similar scalability.

Table. 7 shows acronyms and corresponding full names in this paper.

Fig. 12 shows the visual results of the corner case of PIRATE+ (with the least improvement in Dice) compared with results from other benchmarks on this case. Note that the result of PIRATE+ still maintains state-of-the-art performance in the corner case.

Fig. 13 shows the visual results of NFE along with the training iterations of PIRATE+. The figure shows that the NFE converges in the training process.

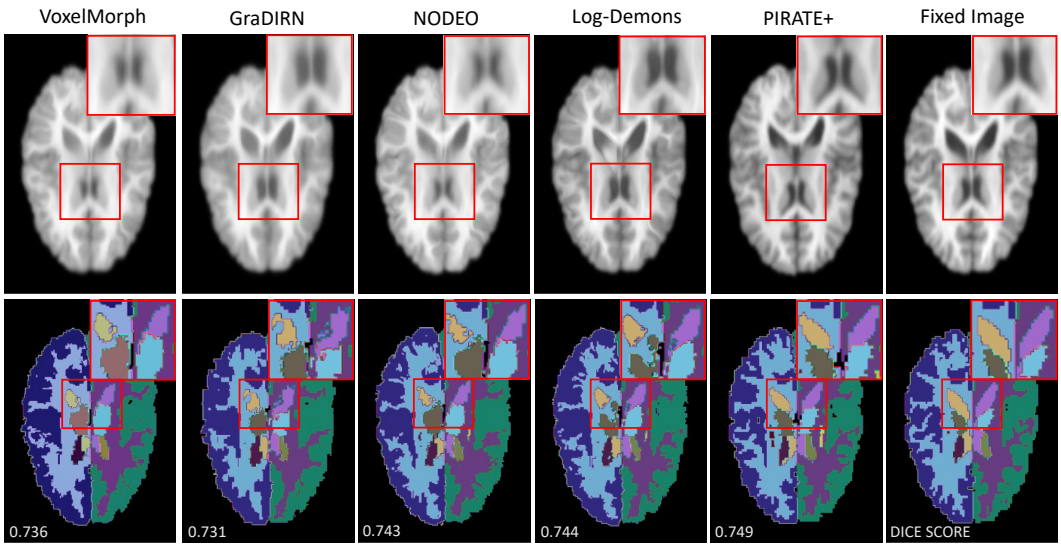

Figure 5: The visual results of warped images (top) and correlated warped segmentation masks (bottom) from PIRATE+ and selected benchmarks on the CANDI dataset. The regions of interest are highlighted within a red box. The DSC for each image is displayed in the bottom row. Note that the result of PIRATE+ is more consistent with the fixed image with fewer artifacts comparing with other baselines.

Table 4: Numerical results of DSC, ratio of negative JD, and inference time for PIRATE, PIRATE+, and benchmarks on the EMPIRE10 lung CT dataset. The variances are shown in parentheses. Note that the **best** and the second-best result are highlighted in red and blue. The result shows that PIRATE+ performs state-of-the-art performance comparing with other baselines. Moreover, PIRATE can achieve competitive performance without DEQ

Method	Avg. DSC ↑	Neg. JD % ↓	Time (s)
SyN	0.9246 (0.045)	0.0854 (0.008)	37.28
VoxelMorph	0.9681 (0.022)	0.0521 (0.012)	**0.16**
SYMNet	0.9701 (0.012)	0.0542 (0.009)	0.28
GraDIRN	0.9644 (0.012)	0.0542 (0.009)	0.44
Log-Demons	0.9772 (0.019)	0.0571 (0.011)	43.29
NODEO	0.9802 (0.028)	**0.0441 (0.008)**	82.17
PIRATE	0.9797 (0.032)	0.0546 (0.009)	29.77
PIRATE+	**0.9811 (0.027)**	0.0467 (0.007)	45.32

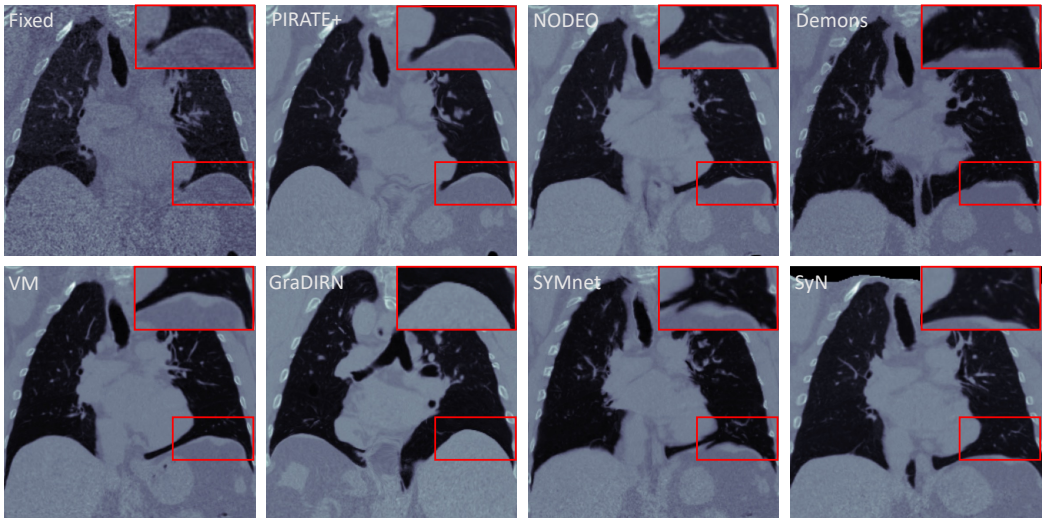

Figure 6: The visual results of warped images from PIRATE+ and selected benchmarks on the EMPIRE10 lung CT dataset. The regions of interest are highlighted within a red box. Note that the result of PIRATE+ is more consistent with the fixed image with fewer artifacts comparing with other baselines.

Figure 7: The visual results of warped segmentation masks from PIRATE+ and selected benchmarks on the EMPIRE10 lung CT dataset. The red region is the mask from fixed image, and the gray region is the misalignment of the warped mask. The DSC for each image is displayed in the bottom. Note that the result of PIRATE+ is more consistent with the fixed mask comparing with other baselines.

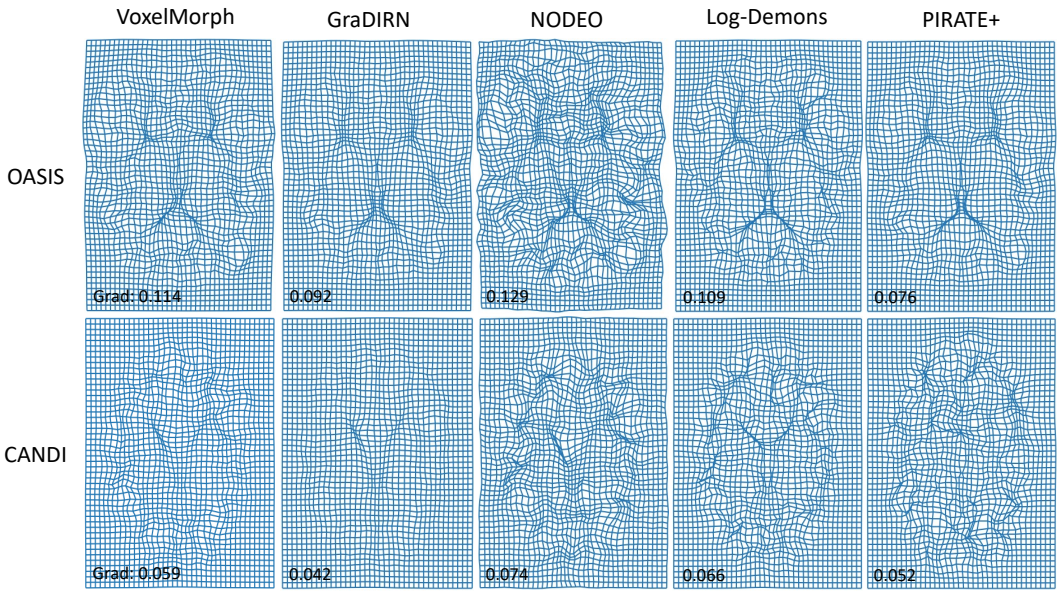

Figure 8: The visual results of warped grid from OASIS (top) and CANDI (bottom) of PIRATE+ and different baselines. The gradient loss is shown at the bottom of each image.

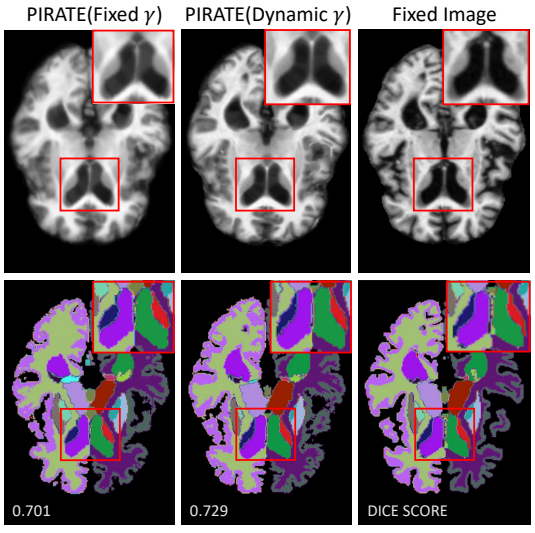

Figure 9: The visual results of warped images (top) and correlated warped segmentation masks (bottom) from PIRATE(fixed γ) and PIRATE(dynamic γ) on the OASIS dataset. The regions of interest are highlighted within a red box. The DSC for each image is displayed in the bottom row.

Table 5: Numerical results of DSC, Jacobian determinant, and inference time for PIRATE with total variation (TV) denoiser and DnCNN denoiser on OASIS and CANDI datasets. The variances are shown in parentheses. Note that the **optimal** results are highlighted in red.

Method	OASIS		CANDI		Time (s)
	Avg. DSC ↑	JD ↓	Avg. DSC ↑	JD ↓	
PIRATE(TV)	0.7617 (0.054)	0.1281 (0.002)	0.7481 (0.049)	0.0992 (0.001)	118.20
PIRATE(DnCNN)	0.7948 (0.033)	0.0567 (0.002)	0.7624 (0.020)	0.0833 (0.002)	**32.01**

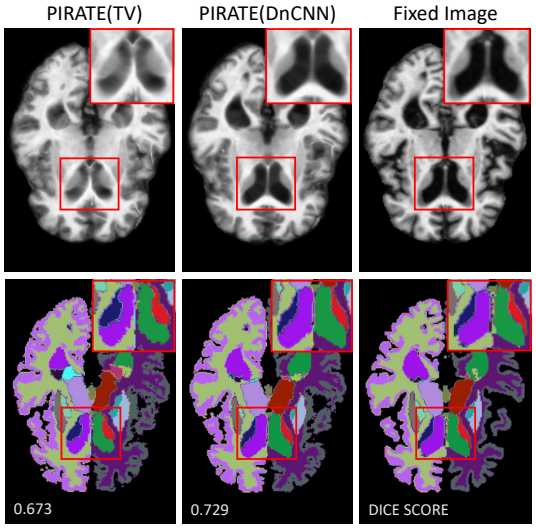

Figure 10: The visual results of warped images (top) and correlated warped segmentation masks (bottom) from PIRATE(TV) and PIRATE(DnCNN) on the OASIS dataset. The regions of interest are highlighted within a red box. The DSC for each image is displayed in the bottom row.

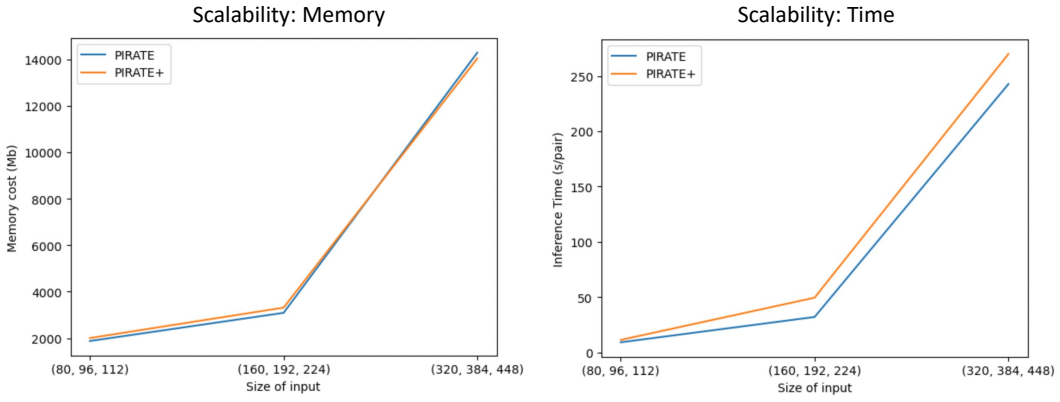

Figure 11: The illustrations of memory cost and inference time of PIRATE and PIRATE+ with different size of inputs. This figure shows that PIRATE and PIRATE+ have similar scalability.

Table 6: Numerical results of memory cost, running time in training and inference for PIRATE, PIRATE+, and benchmarks on OASIS and CANDI datasets. This table shows that PIRATE and PIRATE+ has lower memory complexity compared to other DL baselines.

Method	Inference		Training	
	Memory (MB) ↓	Running time (s) ↓	Memory (MB) ↓	Running time (s/epoch) ↓
SyN	2252	42.52	\	\
VoxelMorph	7642	0.17	12846	78
SYMNet	5976	0.32	16448	185
GraDIRN	4386	0.49	11642	112
Log-Demons	2709	46.72	\	\
NODEO	5534	89.01	\	\
PIRATE	3090	32.01	5408	30.02
PIRATE+	3314	49.47	8866	1821.90

Table 7: Acronyms and corresponding full names in this paper.

Acronyms	Full name	Acronyms	Full name
DIR	deformable image registration	DL	deep learning
CNN	convolutional neural network	PIRATE	plug-and-play image registration network
DEQ	deep equilibrium models	MBDL	model-based deep learning
PnP	plug-and-play	NCC	normalized cross-correlation
GCC	global cross-correlation	DSC	Dice similarity coefficient
JD	Jacobian determinant	MSE	mean squared error
STN	spatial transform network	FCN	fully convolutional network
DU	deep unfolding	SD-RED	steepest descent regularization by denoising
AWGN	additive white Gaussian noise	NODE	neural ordinary differential equations
ASM	adjoint sensitivity method	MMSE	minimum mean squared error
JFB	Jacobian-Free deep equilibrium models	ROI	region of interest
TV	total variation		

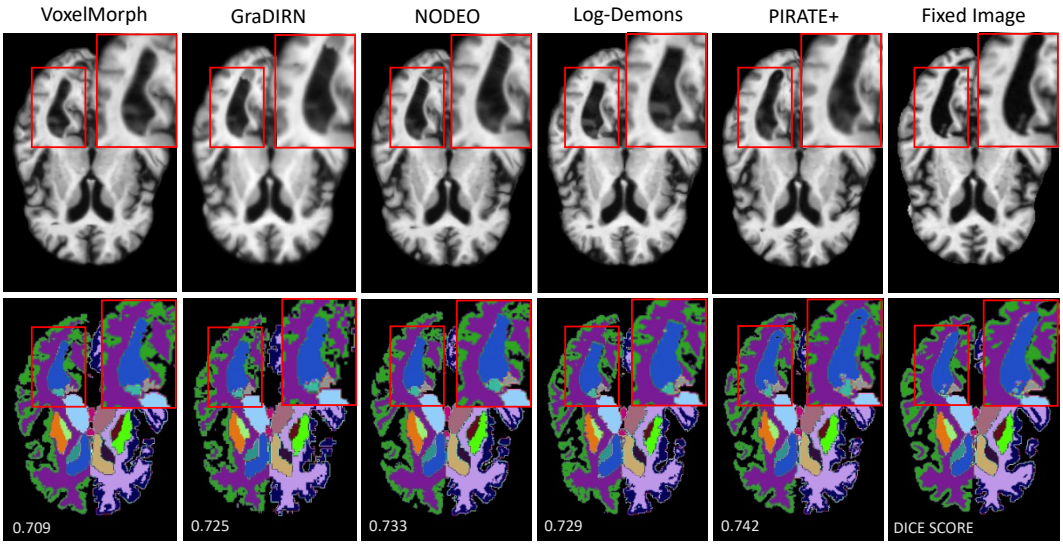

Figure 12: The visual results of warped images (top) and correlated warped segmentation masks (bottom) from the corner case of PIRATE+ (with the least improvement (*not* absolute value) in Dice) compared with results from other benchmarks on this case. The regions of interest are highlighted within a red box. The DSC for each image is displayed in the bottom row. Note that the result of PIRATE+ is still more consistent with the fixed image on this corner case.

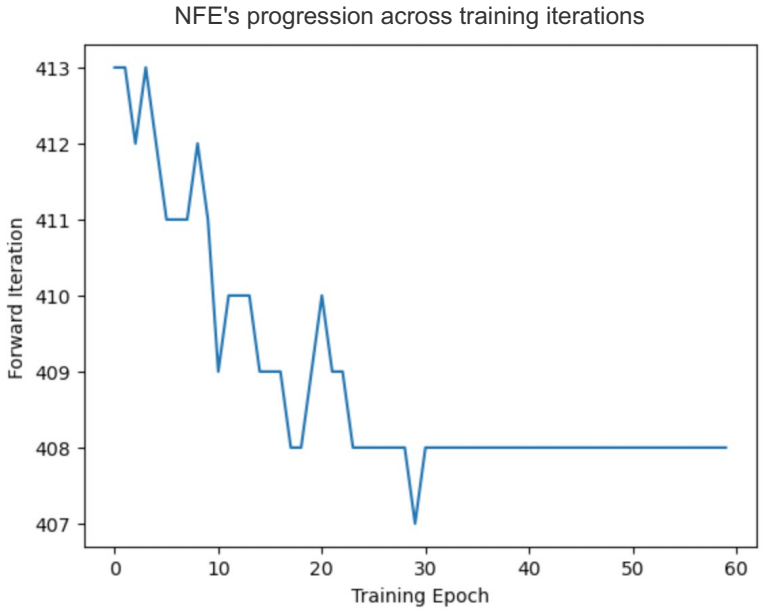

Figure 13: The plot for NFE along with the training iterations of PIRATE+. The results show a convergence behaviour.