

Supplementary materials

1. Robustness to source image quality

Here we present the robustness test results to the quality degradations of source images, for the 5- and 9-spoke under-sampling scenarios of real-time imaging. We first present the DICE coefficients and center-of-mass errors (COMEs) of the cardiac dataset in Fig. S-1. The quality of the prior images was degraded by removing 20%, 50%, and 80% of radial spokes from the fully-sampled k-space readout trajectory. For comparison, we also present the metrics between the fully-sampled source and target images. The results of KS-RegNet using fully-sampled source images are also included for comparison.

While there is a downward trend of registration accuracy for most of the subjects as the degradation factor increases, no significant reduction of registration accuracy is observed.

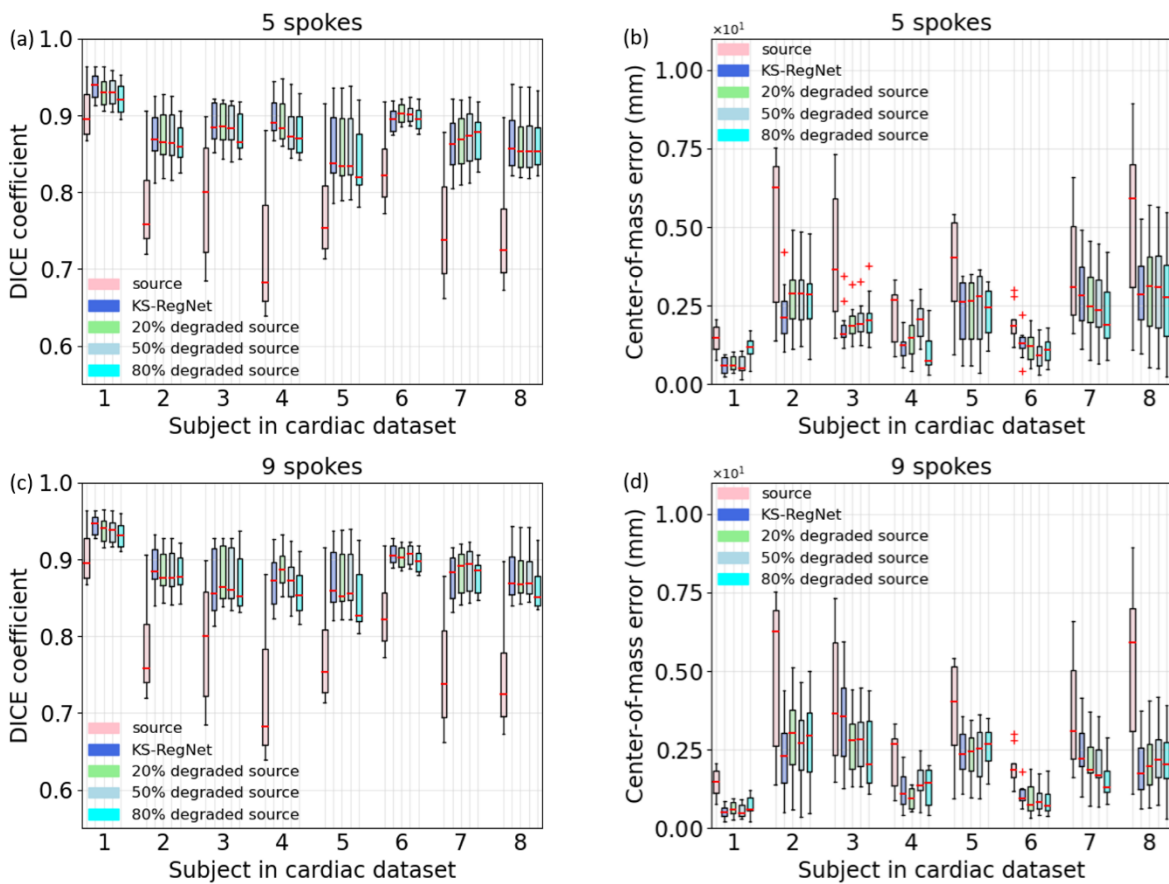


Figure S-1. Robustness test results of KS-RegNet to the quality variations of source images on the cardiac dataset. The spoke numbers are given in the subfigure title. The quality of source images was controlled by under-sampling their k-space data by 20%, 50% and 80%, respectively. For comparison, the first and second boxplots of each subject show the metrics between the source and target images, and the results of KS-RegNet using fully-sampled, non-degraded prior images, respectively.

Figure S-2 presents the robustness test results on the abdominal dataset. The results show a similar trend as the cardiac dataset. From Figs. S-1 and S-2, we see that the performance of KS-RegNet is insensitive to the

source image quality degradations in both cardiac and abdominal studies, which demonstrates the robustness of KS-RegNet to image quality variations of the source/prior images.

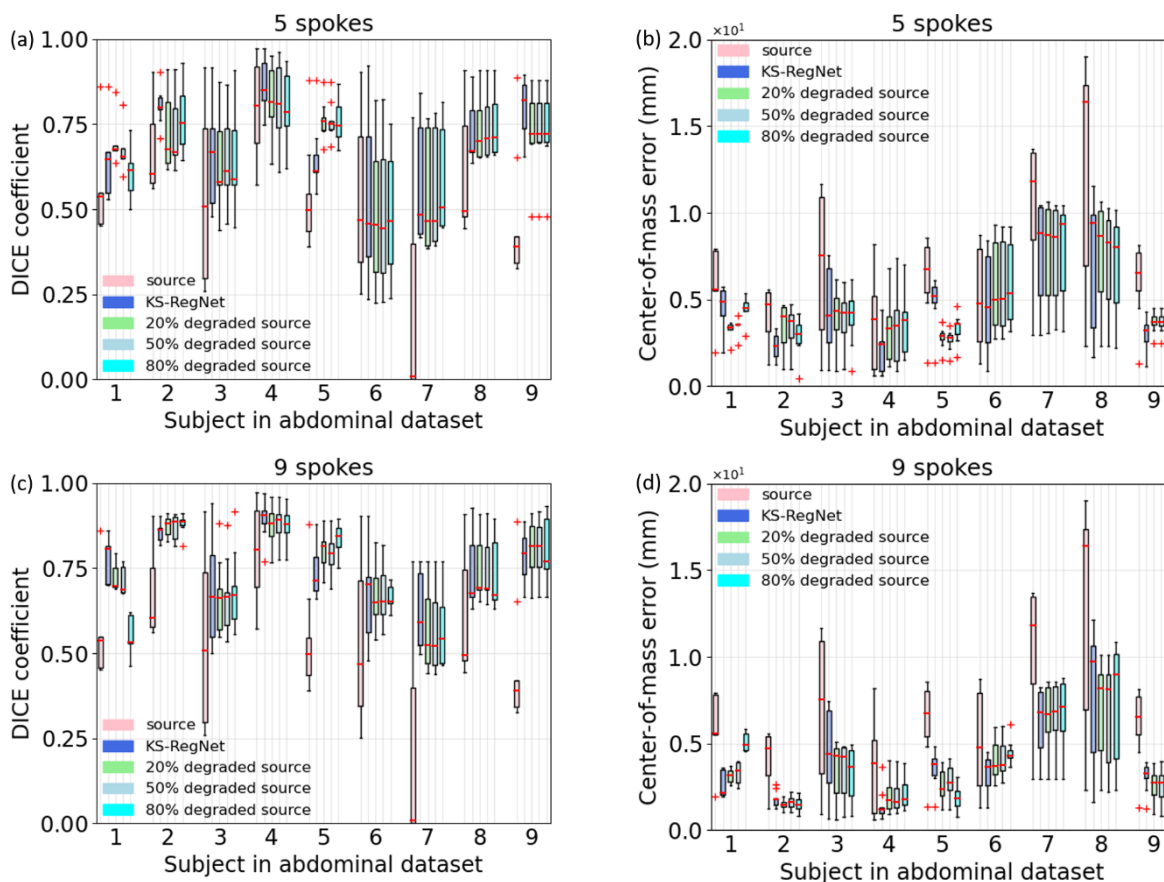


Figure S-2. Robustness test results of KS-RegNet to the quality variations of source images on the abdominal dataset. The spoke numbers are given in the subfigure title. The quality of source images was controlled by under-sampling their k-space data by 20%, 50% and 80%, respectively. For comparison, the first and second boxplots of each subject show the metrics between the source and target images, and the results of KS-RegNet using fully-sampled, non-degraded prior images, respectively.

Tables S-1 and S-2 summarize the mean (\pm s.d.) DICE coefficient and COME of the robustness tests on the cardiac and abdominal datasets, respectively. We also performed Wilcoxon signed-rank tests between the results of KS-RegNet with and without degraded source images. The p-values of the tests are also presented in the Tables. Although some tests appear statistically significant, the actual metric differences are very limited.

Table S-1. Mean (\pm s.d.) DICE coefficients, COME, and Wilcoxon signed-rank test results on the cardiac dataset. The Wilcoxon sign-rank tests are between the results of KS-RegNet with and without degraded source images.

Number of spokes	Degradation factor	Mean (\pm s.d.)		p-value	
		DICE coefficient	COME (mm)	DICE coefficient	COME
5	20%	0.884 \pm 0.041	1.39 \pm 1.18	0.291	0.060
	50%	0.883 \pm 0.040	1.41 \pm 1.19	0.001	0.028
	80%	0.876 \pm 0.038	1.40 \pm 1.09	$< 10^{-4}$	0.007
9	20%	0.891 \pm 0.035	1.38 \pm 1.09	0.814	0.374
	50%	0.889 \pm 0.036	1.40 \pm 1.09	0.237	0.472
	80%	0.879 \pm 0.037	1.40 \pm 1.10	$< 10^{-4}$	0.049
13	20%	0.897 \pm 0.033	1.13 \pm 0.91	0.031	0.257
	50%	0.896 \pm 0.033	1.16 \pm 0.89	0.549	0.526
	80%	0.884 \pm 0.034	1.22 \pm 0.90	$< 10^{-4}$	0.001

Table S-2. Mean (\pm s.d.) DICE coefficients, COME, and Wilcoxon signed-rank test results on the abdominal dataset. The Wilcoxon sign-rank tests are between the results of KS-RegNet with and without degraded source images.

Number of spokes	Degradation factor	Mean (\pm s.d.)		p-value	
		DICE coefficient	COME (mm)	DICE coefficient	COME
5	20%	0.679 \pm 0.162	4.69 \pm 2.62	0.017	0.289
	50%	0.681 \pm 0.160	4.63 \pm 2.58	0.045	0.502
	80%	0.682 \pm 0.154	4.75 \pm 2.46	0.151	0.224
9	20%	0.747 \pm 0.127	3.66 \pm 2.36	0.702	0.104
	50%	0.745 \pm 0.128	3.69 \pm 2.35	0.721	0.177
	80%	0.741 \pm 0.138	3.82 \pm 2.58	0.836	0.456
13	20%	0.754 \pm 0.123	3.41 \pm 2.18	0.107	0.084
	50%	0.757 \pm 0.119	3.40 \pm 2.18	0.027	0.107
	80%	0.748 \pm 0.110	3.58 \pm 2.13	0.052	0.026

2. Data augmentation using the synthesized phase maps

Since the real-valued MR images in the abdominal dataset were augmented by synthesized phase maps to create complex-valued images, the network performance can depend on the degree of the phase-map augmentation. Here we compared the liver tumor registration accuracy for KS-RegNet trained with different degrees of phase map augmentation. Three scenarios of augmentation were considered: without the phase-map augmentation, with 20 phase maps, and with 40 phase maps. Figure S-3 presents the DICE coefficients and COMEs of KS-RegNet with various numbers of synthesized phase maps at three different under-sampling factors, and Table S-3 summarizes the mean (\pm s.d.) and p-values of the Wilcoxon signed-rank test between different degrees of augmentation.

The results show that the registration accuracy improves with more phase maps in the data augmentation, and except for the 13-spoke case, the p-values between the cases of the 20 and 40 phase maps are smaller

than 0.05. Furthermore, the 5-spoke trajectory is most benefited from the increase of the phase maps, which may indicate the importance of phase augmentation for severely under-sampled cases.

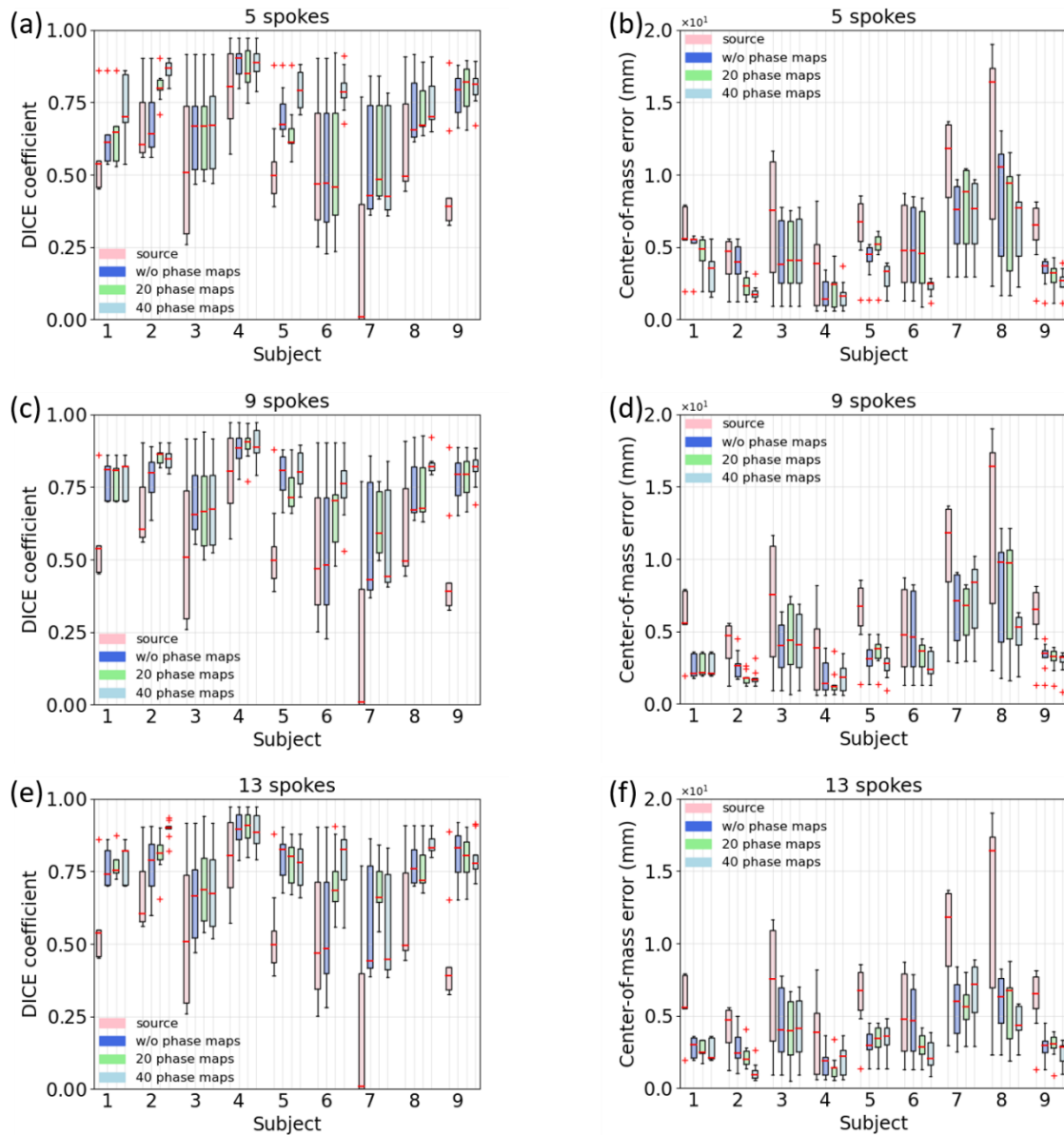


Figure S-3. Comparison of liver tumor registration accuracy with various numbers of synthesized phase maps used in the data augmentation, at three under-sampling factors. The spoke numbers are given in the subfigure titles.

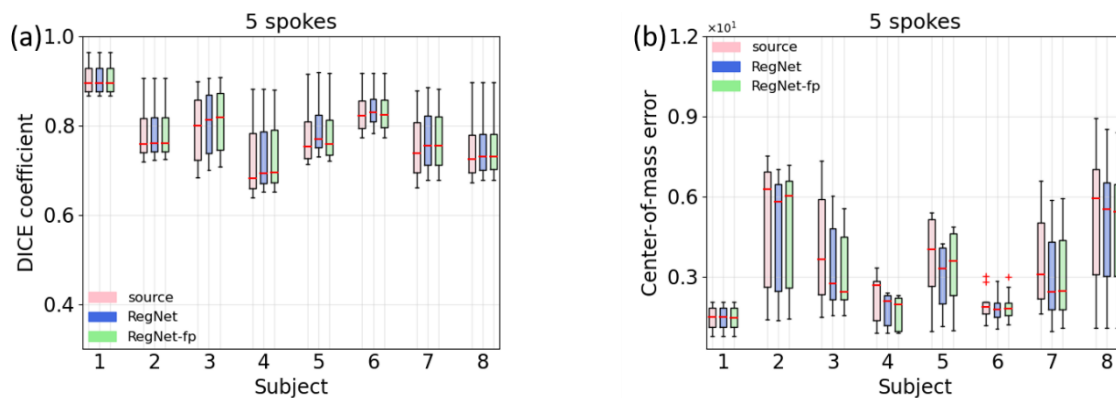
Table S-3. Mean (\pm s.d.) DICE coefficients and COMEs of different levels of phase-map augmentation, and p-values of the Wilcoxon signed-rank test between the three levels of augmentation.

Number of spokes	Metric	w/o phase maps	20 phase maps	40 phase maps	p-value (0-20)	p-value (0-40)	p-value (20-40)
5	DICE	0.683 \pm 0.175	0.694 \pm 0.167	0.755 \pm 0.146	0.024	< 10 ⁻³	< 10 ⁻³
	COME (mm)	4.79 \pm 2.97	4.61 \pm 2.91	3.72 \pm 2.56	0.068	< 10 ⁻³	< 10 ⁻³
9	DICE	0.719 \pm 0.171	0.751 \pm 0.128	0.771 \pm 0.137	0.051	< 10 ⁻³	0.030
	COME (mm)	4.22 \pm 2.80	3.91 \pm 2.72	3.43 \pm 2.17	0.038	0.001	0.008
13	DICE	0.727 \pm 0.168	0.766 \pm 0.106	0.776 \pm 0.142	0.007	0.006	0.489
	COME (mm)	3.83 \pm 2.20	3.48 \pm 2.00	3.28 \pm 2.04	0.005	0.034	0.271

3. RegNet with fully-sampled source images

Here we compare the registration accuracy of RegNet with and without accessing fully-sampled source images. The ablation study of KS-RegNet shows that, when the input channels contain the fully-sampled source images, the registration accuracy improves and the model are more robust. However, it is unclear whether the accessibility of fully-sampled source image benefits RegNet. The RegNet variant with accessing to the fully-sampled prior is called RegNet-fp. Figure S-4 compares the registration accuracy of RegNet and RegNet-fp for the subjects in the cardiac dataset. Table S-4 summarizes the mean (\pm s.d.) DICE coefficients and COMEs and the p-values of the Wilcoxon signed-rank test between the two networks.

The 13-spoke case shows a minute improvement of the registration accuracy with p-values < 0.05 when the network is able to access the fully-sampled source image. On the other hand, for the 5- and 9-spoke cases, one can see the registration accuracy even slightly decreases for RegNet-fp. Since the prior information is not utilized to define the similarity loss during the network training, it seems that RegNet-fp is unable to fully utilize this prior information to improve the registration accuracy. Moreover, as the image quality between the fully- and under-sampled source images diverts apart, this additional input channel of high-quality source image seems potentially confusing the registration network when the sampling ratio is very low (5- and 9- spokes).



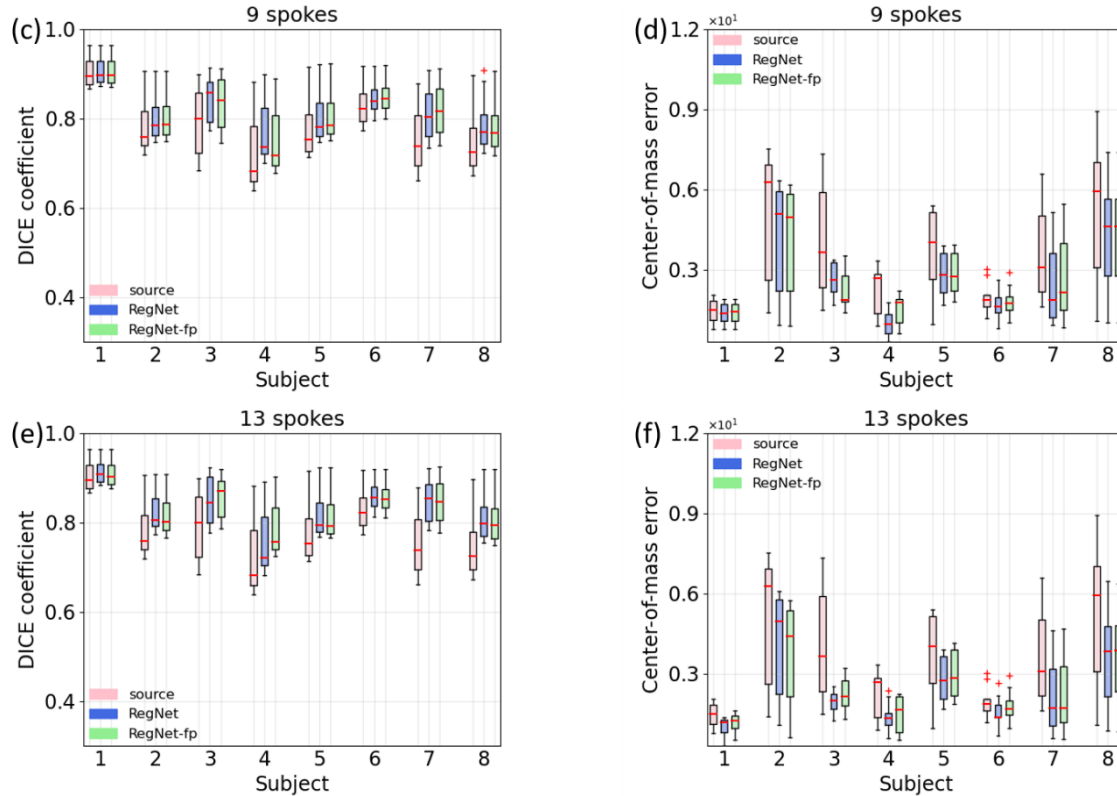


Figure S-4. Comparison of the registration accuracy of RegNet and RegNet-fp for the cardiac dataset. RegNet-fp stands for the RegNet with fully-sampled prior images as additional network inputs.

Table S-4. Mean (\pm s.d.) DICE coefficients and COMEs of RegNet and RegNet-fp for the subjects in the cardiac dataset, and p-values of the Wilcoxon signed-rank test between the two variant of RegNet.

Number of spokes	Metric	RegNet	RegNet-fp	p-value
5	DICE	0.797 \pm 0.051	0.796 \pm 0.014	0.822
	COME (mm)	1.82 \pm 1.85	1.83 \pm 1.86	0.426
9	DICE	0.821 \pm 0.040	0.819 \pm 0.013	0.886
	COME (mm)	1.61 \pm 1.56	1.62 \pm 1.54	0.341
13	DICE	0.835 \pm 0.042	0.836 \pm 0.033	0.001
	COME (mm)	1.49 \pm 1.41	1.52 \pm 1.392	0.032

Figure S-5 presents the liver tumor registration accuracy of RegNet and RegNet-fp, and the mean (\pm s.d.) DICE coefficients and COMEs are summarized in Table S-5. Like the cardiac dataset, RegNet-fp has better registration accuracy for the 13-spoke case, but the registration error increases when the under-sampling factor increases for some cases (5- and 9- spokes).

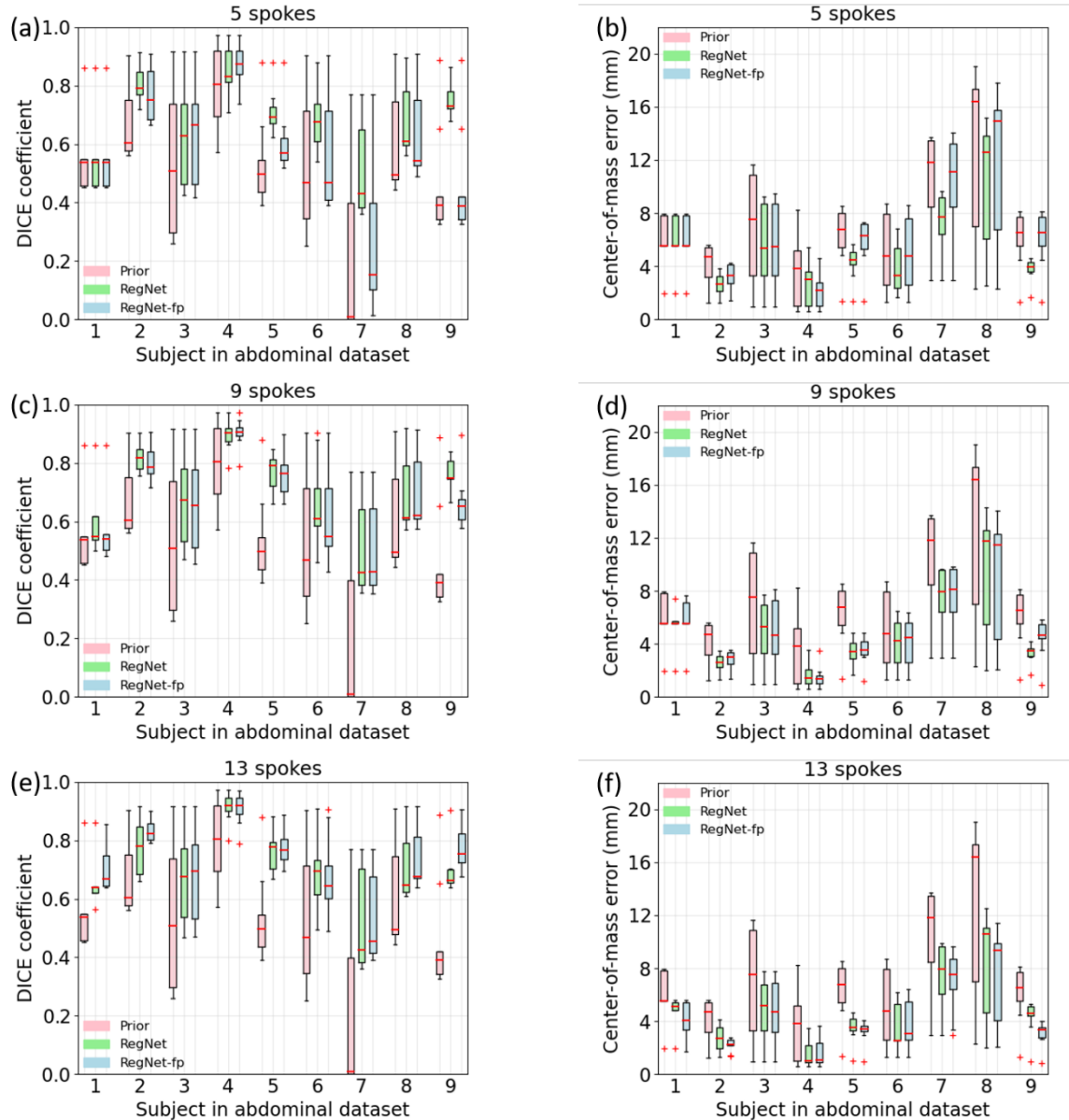


Figure S-5. Comparison of the registration accuracy of RegNet and RegNet-fp for the abdominal dataset.

Table S-5. Mean (\pm s.d.) DICE coefficients and COMEs of RegNet and RegNet-fp for the subjects in the abdominal dataset, and p-values of the Wilcoxon signed-rank test between the two variant of RegNet.

Number of spokes	Metric	RegNet	RegNet-fp	p-value
5	DICE	0.693 \pm 0.151	0.594 \pm 0.228	$< 10^{-3}$
	COME (mm)	5.14 \pm 3.35	6.30 \pm 4.22	$< 10^{-3}$
9	DICE	0.713 \pm 0.156	0.692 \pm 0.163	$< 10^{-3}$
	COME (mm)	4.62 \pm 3.23	4.78 \pm 3.19	0.058
13	DICE	0.712 \pm 0.149	0.735 \pm 0.145	$< 10^{-3}$
	COME (mm)	4.54 \pm 2.93	4.17 \pm 2.72	$< 10^{-3}$