MRI Cross-Modality Image-to-Image Translation

Nature Research

Scientific Reports 2020

이한얼

Data and Visual Analytics Lab

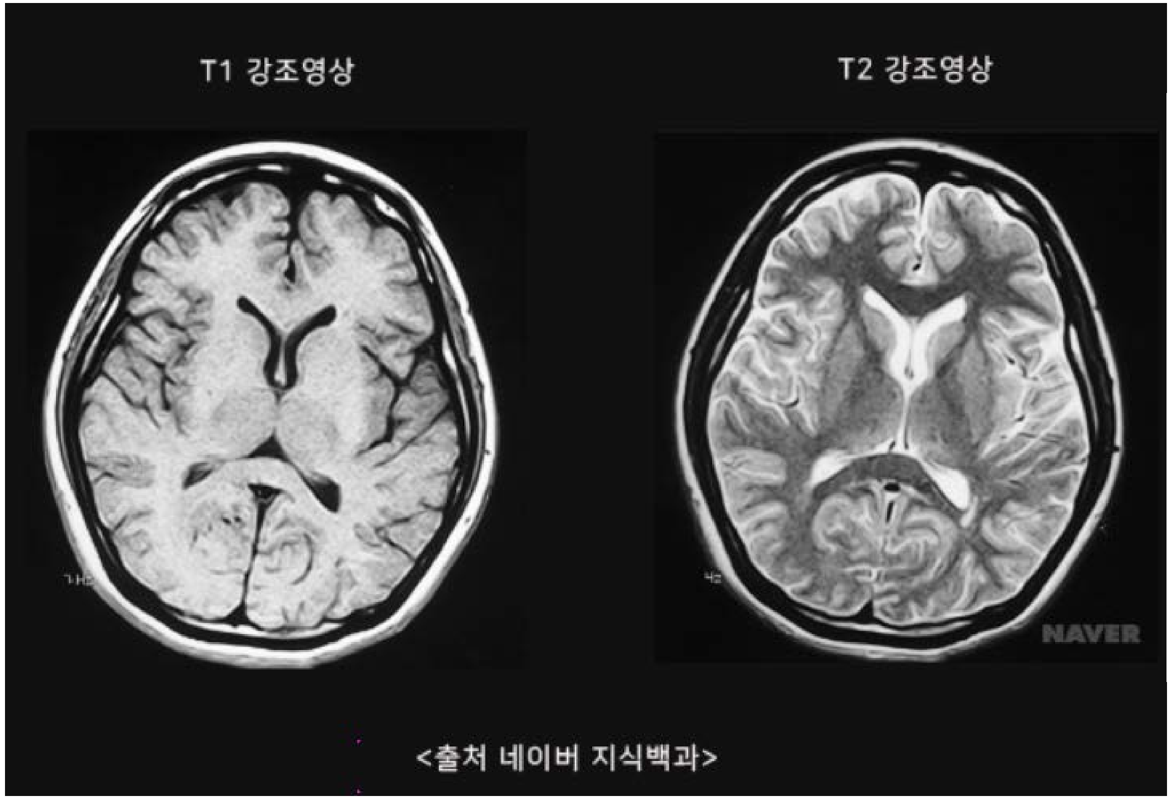
Contribution

- Present a new approach for cross-modality MR image generation using IMT network
- Contribution
  1) Introduce the end-to-end Image Modality Translation (IMT) network for cross-modality MRI generation to synthesize translated modalities from given modalities. A comprehensive comparison is provided with five datasets representing real-world clinical applications, each has its unique characteristics in data size, patient cohort and disease status. The results show that our IMT framework can cope with a variety of brain MRI modality translation tasks using the same objective and architecture.
  2) Registration: We proposed a registration method which is able to leverage our IMT framework to augment the fixed images space with translated modalities for atlas-based registration. Registering moving images to fixed images and weighted fusion process enable us to make the most of cross-modality information without adding any extra data.
  3) Segmentation: Translated multichannel segmentation (TMS), performs cross-modality image segmentation by means of FCNs. We input two identical given modalities and one corresponding translated modality into separate channels, which allows us to adopt and fuse cross-modality information and improve the segmentation performance without using any extra data.

Provided a comprehensive comparison with five datasets representing real-world clinical applications, each has its unique characteristics in data size, patient cohort and disease status

Link : https://www.nature.com/articles/s41598-020-60520-6

MRI, T1, T2, T1-Flair, T2-Flair



구분	T1	T2
물	검은색	흰색
지방	흰색	검은색
뇌실	검은색	흰색
백색질	흰색	검은색
회색질	회색	회색
석회화,뼈	검은색	검은색

- T1 강조영상은 짧은 TR과 짧은 TE를 이용한 스핀에코 기법으로 서 조직의 T1 이완시간의 차이를 신호 차이로 반영하는 기법
- T2강조영상은 긴 TR과 긴 TE를 이용한 스핀에코 기법으로서 조직의 T2이완시간의 차이를 신호 차이로 반영 하는 기법
- FLAIR는 180도 반전펄스를 먼저 가하는 반전회복 (inversion recovery) 기법의 일종으로서 뇌척수액의 신호를 억제하기 위하여 2500 msec 정도의 반전시간을 적용

About details of MRI Link : https://www.i-mri.org/Synapse/Data/PDFData/0040JKSMRM/jksmrm-13-9.pdf

MRI Cross-Modality Image-to-Image Translation

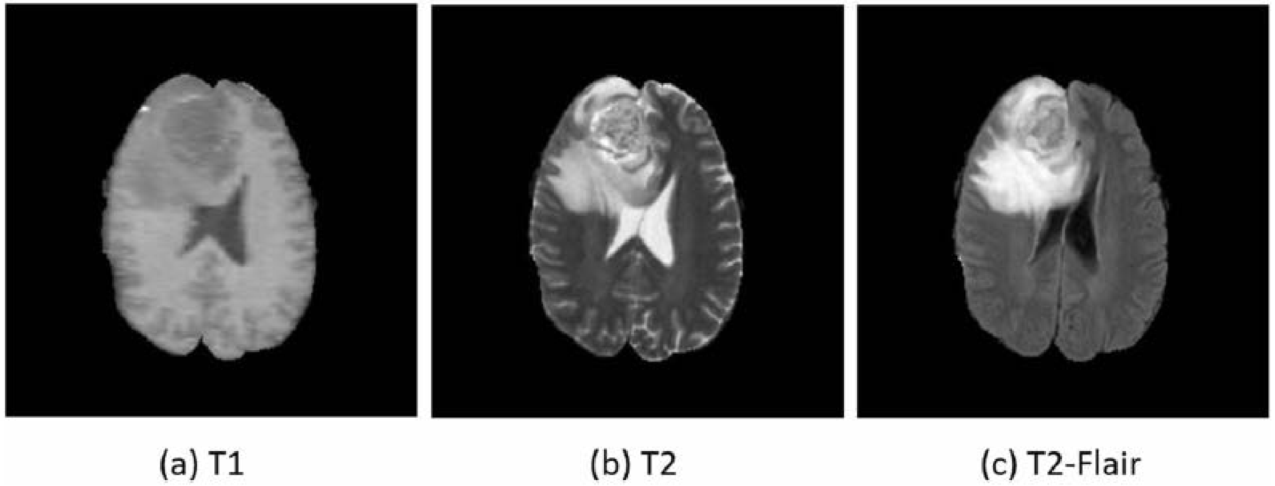


Figure 1. Examples of three different modalities: (**a**) T1, (**b**) T2, and (**c**) T2-Flair.

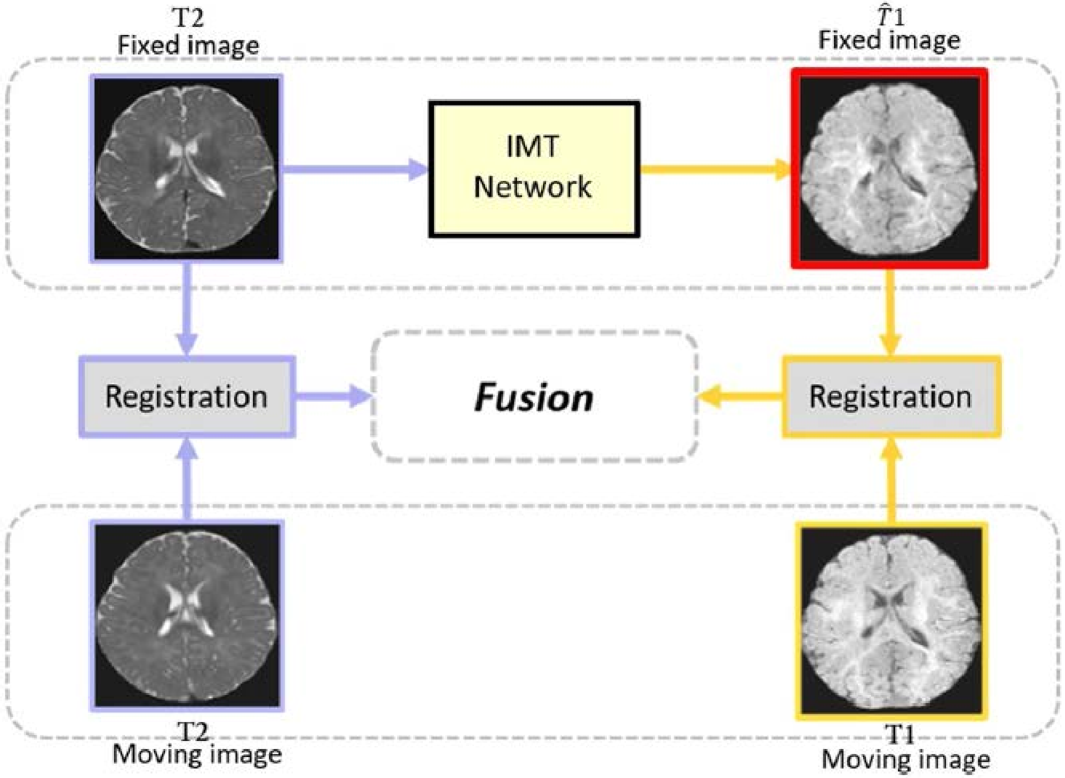


Figure 3. Overview of our approach for cross-modality registration. Inputting a given-modality image (T2) to IMT framework yields a translated modality ($\hat{T}1$). Then T2 (moving) is registered to T2 (fixed), T1 (moving) is registered to $\hat{T}1$ (fixed). The deformation generated in the registration process are finally combined in a weighted fusion process, obtaining our final registration result. The red box indicates our translated images.

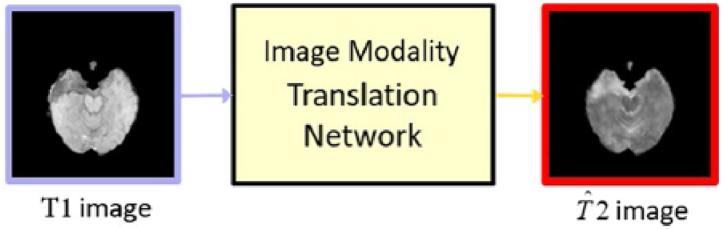


Figure 2. Overview of our IMT network. It learns to generate translated modality images ($\hat{T}2$) from given modality images (T1). The red box indicates our translated images.

Methods

Training

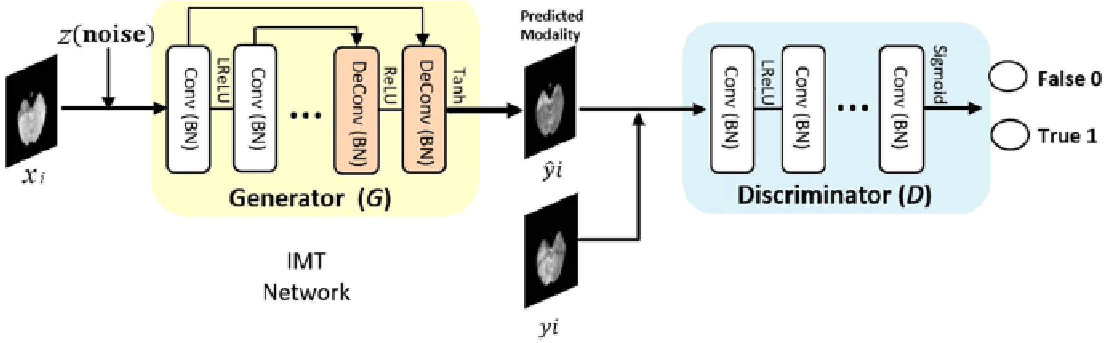


Figure 4. Overview of our end-to-end IMT network for cross-modality generation. Notice that our training set is denoted as $S = \{(x_i, y_i), i = 1, 2, 3, \ldots, n\}$, where x_i and y_i refer to the ith input given-modality image and its corresponding target-modality image. The training process involves two aspects. On the one hand, given an input image x_i and a random noise vector z, generator G aims to produce indistinguishable images $\hat{y}_i$ from the real images y_i. On the other hand, discriminator D evolves to distinguish between translated-modality images $\hat{y}_i$ generated by G and the real images y_i. The output of D is 0 or 1, where 0 represents synthesized images and 1 represents the real data. In the generation process, translated-modality images can be synthesized through the optimized G.

$$\mathscr{L}_{cGAN}(G, D) = \mathbb{E}_{x,y\sim p_{data}(x,y)}[\log D(x, y)] + \mathbb{E}_{x\sim p_{data}(x), zp_z(z)}[\log(1 - D(x, G(x, z))],$$

$$\mathscr{L}_{L1}(G) = \mathbb{E}_{x,y\sim p_{data}(x,y), z\sim p_z(z)}[\|y - G(x, z)\|_1].$$

$$\mathscr{L} = \mathscr{L}_{cGAN}(G, D) + \lambda \mathscr{L}_{L1}(G)$$

$$G^* = arg\ \min_G \max_D \mathscr{L}_{cGAN}(G, D) + \lambda \mathscr{L}_{L1}(G)$$

Network architecture

- Generator
  - The architecture of G has 8 convolutional layers, each of which contains a convolution, a Batch Normalization, and a leaky ReLu activation (a slope of 0.2) with numbers of filters at 64, 128, 256, 512, 512, 512, 512, and 512 respectively
  - Following them are 8 deconvolutional stages, each of which includes a deconvolution, a Batch Normalization, and an unleaky ReLu (a slope of 0.2) with numbers of filters at 512, 1024, 1024, 1024, 1024, 512, 256, and 128 respectively
  - Tanh activation function
- Discriminator
  - The architecture of *D* contains four stages of convolution-BatchNorm-ReLu with the kernel size of (4,4)
  - The numbers of filters are 64, 128, 256, and 512 for convolutional layers. Lastly, a sigmoid function is used to output
  - The confidence probability that the input data comes from real MR images rather than generated images

Application

1) Cross-modality image registration

2) Cross-modality image segmentation

3) Cross-modality image generation

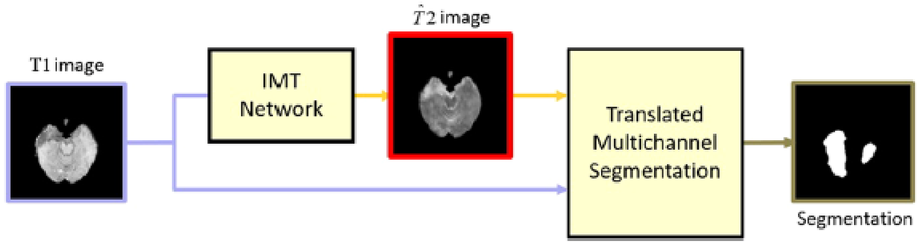


Figure 5. Flowchart of our approach for cross-modality segmentation. First, we input a given-modality image to our IMT network to generate a translated-modality image. For instance, given a T1 image, $\hat{T}2$ images can be generated with our method. Second, two identical given-modality images and one corresponding translated-modality image are fed to channels 1, 2, and 3 and segmented by FCN networks. Under the standard FCN-32s, standard FCN-16s, and standard FCN-8s settings, we output our segmentation results. The red box indicates our translated images.

Dataset

1) BraTs2015

2) Iseg2017

3) MRBrain13

4) ADNI

5) RIRE

(1)*BraTs2015*: The BraTs2015 dataset[60] contains multi-contrast MR images from 220 subjects with high-grade glioma, including T1, T2, T2-Flair images and corresponding labels of tumors. We randomly select 176 subjects for training and the rest for testing. 1924 training images are trained for 600 epochs with batch size 1. 451 images are used for testing.

(2)*Iseg2017*: The Iseg2017 dataset[61] contains multi-contrast MR images from 23 infants, including T1, T2 images and corresponding labels of Grey Matter (gm) and White Matter (wm). We randomly select 18 subjects for training and remaining 5 subjects for testing. 661 training images are trained for 800 epochs with batch size 1. 163 images from the 5 subjects are used for testing.

(3)*MRBrain13*: The MRBrain13 dataset[62] contains multi-contrast MR images from 20 subjects, including T1 and T2-Flair images. We randomly choose 16 subjects for training and the remaining 4 for testing. 704 training images are trained for 1200 epochs with batch size 1. 176 images are used for testing.

(4)*ADNI*: The ADNI dataset[30] contains T2 and PD images (proton density images, tissues with a higher concentration or density of protons produce the strongest signals and appear the brightest on the image) from 50 subjects. 40 subjects are randomly selected for training and the remaining 10 for testing. 1795 training images are trained for 400 epochs with batch size 1. 455 images are used for testing.

(5)*RIRE*: The RIRE dataset[63] includes T1 and T2 images collected from 19 subjects. We randomly choose 16 subjects as for training and the rest for testing. 477 training images are trained for 800 epochs with batch size 1. 156 images are used for testing.

Experiments

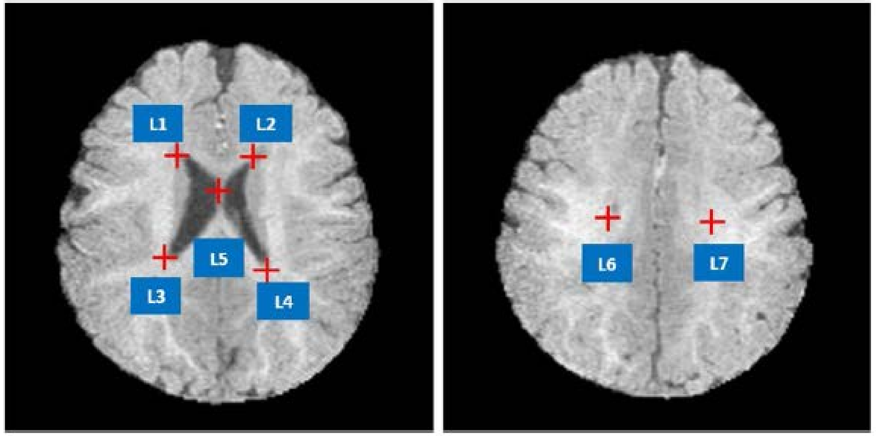


Figure 6. Illustration of the seven landmarks selected for cross-modality registration. L1: right lateral ventricle superior, L2: left lateral ventricle superior, L3: right lateral ventricle inferior, L4: left lateral ventricle inferior. L5: middle of the lateral ventricle, L6: right lateral ventricle posterior, L7: left lateral ventricle posterior.

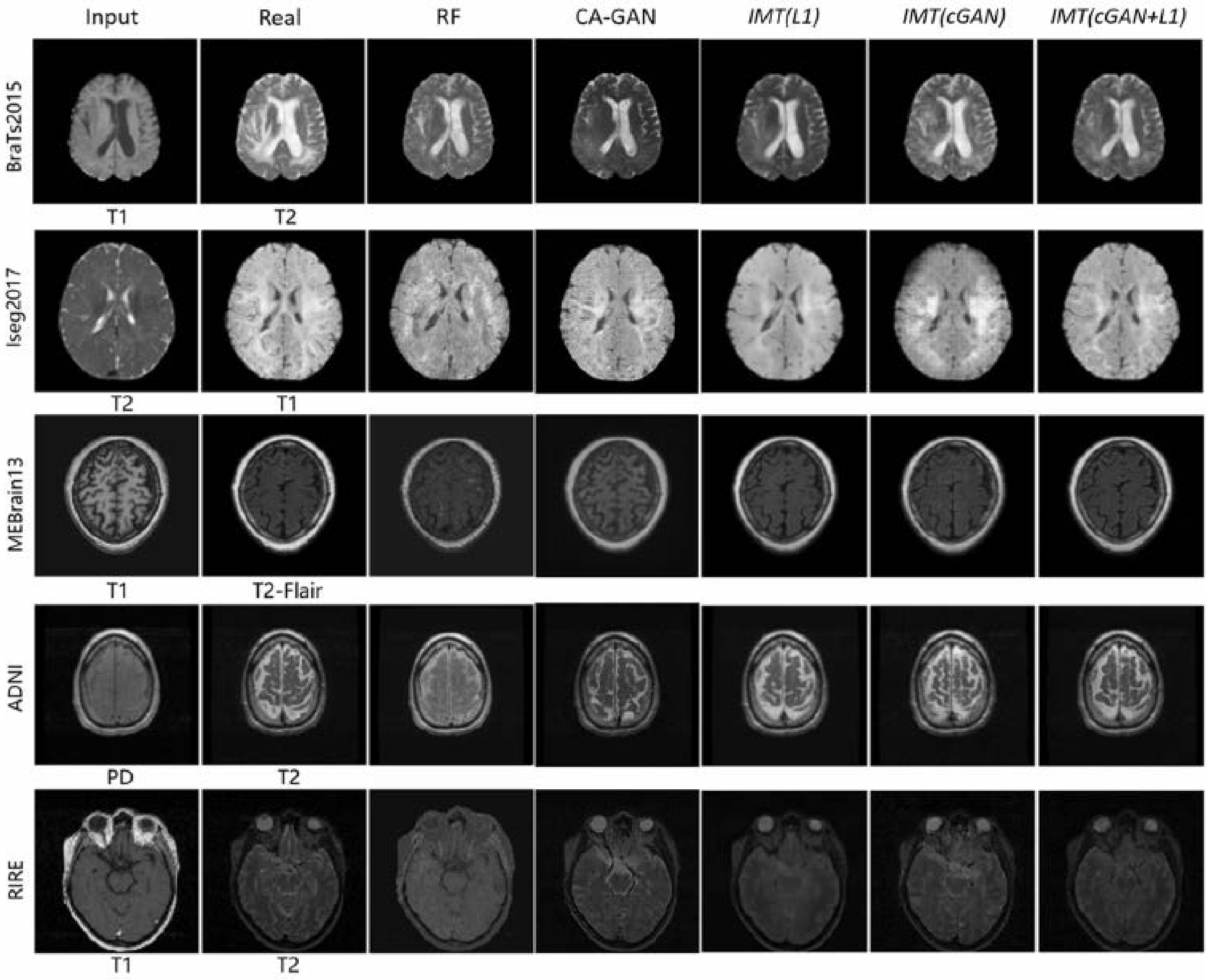


Figure 7. Samples of cross-modality generation results on five publicly available datasets including *BraTs2015*[60], *Iseg2017*[61], *MRBrain13*[62], *ADNI*[30], and *RIRE*[63]. Results are selected from top performing examples (relatively low MAE, high PSNR, high MI, and high PSNR collectively) with four approaches. The right five columns show results of the random-forests-based method (RF)[5], the Context-Aware GAN (CA-GAN)[30] and IMT framework with different loss functions ($L1$, $cGAN$, $cGAN + L1$).

Experiments-MAE, PSNR, SSIM, MI

$$MAE = \frac{1}{m \times n} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} \| \hat{y}(i,j) - y(i,j) \|$$

$$MSE = \frac{1}{NM} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} e(m,n)^2$$

$$PSNR = 10log\frac{s^2}{MSE}$$

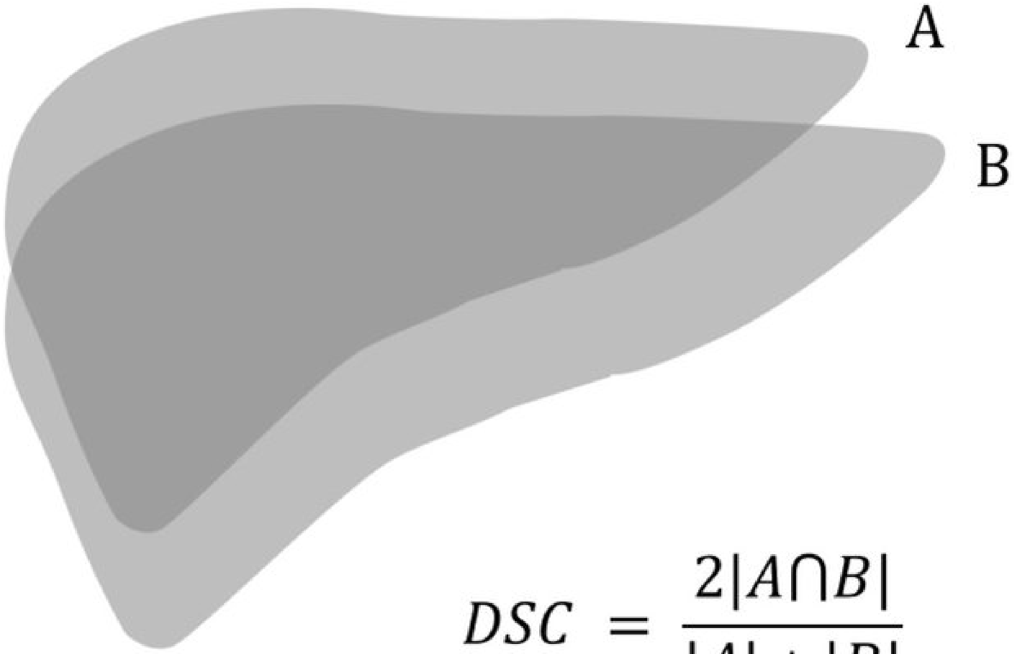
S → 해당 영상의 최대값으로서, 해당 채널의 최대값에서 최소값을 빼서 구함
8bit grayscale 영상의 경우 255(255-0) 이 된다.

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(2\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}$$

$$I(y;\hat{y}) = \sum_{m \in y} \sum_{n \in \hat{y}} p(m,n) \log\left(\frac{p(m,n)}{p(m)p(n)}\right)$$

- 평균 절대값 오차 (MAE), 평균 제곱 오차 (MSE)
- 최대 신호대 잡음비 (PSNR)
  - 신호가 가질 수 있는 최대 신호에 대한 잡음의 비를 나타냄
  - 화질 손실 정보를 평가할 때 사용 (손실이 적을수록 높은 값을 가짐)
- 구조적 유사 지수 (SSIM) : 압축 및 변환에 의해 발생하는 왜곡에 대해 원본 영상에 대한 유사도를 측정하는 방법 (MSE, PSNR 방법보다 더 정확하게 비교)
- Mutual Information
  - m. n : intensity

Experiments-Dice similarity coefficient, Distance Between Corresponding Landmarks (Dist)



$$DSC = \frac{2|A \cap B|}{|A| + |B|}$$

DSC: Dice similarity coefficient

(2) *Distance Between Corresponding Landmarks (Dist)*: The second metric is adopted to measure the capacity of algorithms to register the brain structures. The registration error on a pair of images is defined as the average Euclidean distance between a landmark in the warped image and its corresponding landmark in the fixed image. To compute the Euclidean distance, all 2D-slices after registration are stacked into 3D images.

Experiments

Datasets	Transitions	RF				CA-GAN				IMT											
										cGAN + L1				cGAN				L1			
		MAE ↓	PSNR ↑	MI ↑	SSIM ↑	MAE ↓	PSNR ↑	MI ↑	SSIM ↑	MAE ↓	PSNR ↑	MI ↑	SSIM ↑	MAE ↓	PSNR ↑	MI ↑	SSIM ↑	MAE ↓	PSNR ↑	MI ↑	SSIM ↑
BraTs2015	T1 → T2	**6.025**	**24.717**	0.617	**0.910**	11.947	19.738	0.787	0.826	8.292	22.560	0.862	0.866	10.692	20.301	0.788	0.575	8.654	22.517	**0.901**	0.880
	T2 → T1	**7.921**	**23.385**	0.589	**0.893**	16.587	17.462	0.661	0.723	9.937	22.518	0.777	0.854	15.430	18.507	0.673	0.723	10.457	22.374	**0.818**	**0.896**
	T1 → T2-Flair	**8.176**	**23.222**	0.609	**0.873**	13.999	19.157	0.722	0.756	**7.934**	**22.687**	0.833	0.837	11.671	19.969	0.749	0.797	8.462	22.642	**0.879**	0.857
	T2 → T2-Flair	**7.318**	**23.138**	0.610	**0.875**	12.658	18.848	0.756	0.749	8.858	21.664	0.848	0.836	10.469	20.656	0.817	0.823	8.950	21.791	**0.928**	0.860
Iseg2017	T1 → T2	3.955	28.028	0.803	0.902	12.175	21.992	0.804	0.690	**3.309**	**29.979**	0.931	0.887	8.028	22.860	0.782	0.748	3.860	28.874	**0.993**	**0.913**
	T2 → T1	11.466	22.342	0.788	**0.808**	17.151	18.401	0.789	0.662	**9.586**	**23.610**	**0.868**	0.745	17.311	18.121	0.777	0.620	10.591	**23.325**	**0.880**	0.754
MRBrain13	T1 → T2-Flair	7.609	24.780	1.123	0.863	13.643	19.503	0.805	0.782	**6.064**	**26.495**	1.066	0.823	9.906	22.616	1.009	0.785	6.505	**26.299**	**1.185**	**0.881**
ADNI	PD → T2	9.485	24.006	1.452	0.819	16.575	19.008	0.674	0.728	6.757	26.477	1.266	0.812	7.211	26.330	1.184	0.779	**4.898**	**29.089**	**1.484**	**0.891**
	T2 → PD	5.856	29.118	1.515	**0.880**	17.648	18.715	0.659	0.713	**4.590**	**31.014**	1.381	0.856	5.336	29.032	1.282	0.820	5.055	30.614	**1.536**	**0.881**
RIRE	T1 → T2	38.047	12.862	0.694	0.501	18.625	18.248	**0.724**	0.749	**5.250**	**28.994**	0.636	0.736	13.690	21.038	0.513	0.506	9.105	**28.951**	**0.698**	**0.760**
	T2 → T1	17.022	19.811	0.944	0.622	23.374	16.029	0.650	0.728	**9.035**	**24.043**	0.916	0.692	13.964	20.450	0.737	0.538	9.105	**24.003**	**0.969**	**0.741**

Table 1. Generation performance on five publicly available datasets evaluated by MAE, PSNR, MI, and SSIM. The bold entries in this table indicate the algorithm which gets the best performance in each task. The standard for choosing the best algorithm is to have statistical significance over the other algorithms (p-value < 0.05). If an algorithm gets the best evaluation metrics but has no statistical significance over the others (p-value > 0.05), all of them will be regarded as the best algorithms. The result show that our IMT approach outperforms both Random Forest (RF) based method[5] and Context-Aware GAN (CA-GAN)[30] method on most datasets.

Method	Accuracy		Dice
	all	tumor	tumor
T1 → T2	0.955	0.716	0.757
T2 (real)	0.965	0.689	0.724
T2 → T1	0.958	0.663	0.762
T1 (real)	0.972	0.750	0.787
T1 → T2-Flair	0.945	0.729	0.767
T2 → T2-Flair	0.966	0.816	0.830
T2-Flair (real)	0.986	0.876	0.899

Table 2. Segmentation results of IMT images on *BraTs2015* evaluated by FCN-score. The gap between translated images and the real images can evaluate the generation performance of our method. Note that "all" represents mean accuracy of all pixels (the meanings of "all" are the same in the following tables). We achieve close segmentation results between translated-modality images and target-modality images.

Method	Accuracy			Dice	
	all	gm	wm	gm	wm
T1 → T2	0.892	0.827	0.506	0.777	0.573
T2 (real)	0.920	0.829	0.610	0.794	0.646
T2 → T1	0.882	0.722	0.513	0.743	0.569
T1 (real)	0.938	0.811	0.663	0.797	0.665

Table 3. Segmentation results of IMT translated images on *Iseg2017* evaluated by FCN-score. Note that and "wm" indicate gray matter and white matter respectively. The minor gap between images and the target-modality images shows decent generation performance of our framework.

Experiments

Datasets	Modalities	Structures	Dice		Dist	
			ANTs	Elastix	ANTs	Elastix
Iseg2017	T2	wm	0.508	0.475	2.105	2.836
		gm	0.635	0.591		
	$\hat{T}1$	wm	0.503	0.469	1.884	2.792
		gm	0.622	0.580		
	$T2+\hat{T}1$	wm	**0.530**	**0.519**	**1.062**	**2.447**
		gm	**0.657**	**0.648**		
	T1	wm	0.529	0.500	1.136	2.469
		gm	0.650	0.607		
	$\hat{T}2$	wm	0.495	0.457	2.376	3.292
		gm	0.617	0.573		
	$T1+\hat{T}2$	wm	**0.538**	**0.527**	**1.097**	**2.116**
		gm	**0.664**	**0.650**		
	T1 + T2	wm	0.540	0.528	1.013	2.109
		gm	0.666	0.651		
MRBrain13	T2-Flair	wm	0.431	0.412	3.417	3.642
		gm	0.494	0.463		
	$\hat{T}1$	wm	0.468	0.508	3.159	3.216
		gm	0.508	0.487		
	T2-Flair + $\hat{T}1$	wm	**0.473**	**0.492**	**2.216**	**2.659**
		gm	**0.530**	**0.532**		
	T1	wm	0.484	0.534	2.524	2.961
		gm	0.517	0.510		
	$\hat{T}2$*-Flair*	wm	0.431	0.410	3.568	3.726
		gm	0.497	0.458		
	$T1+\hat{T}2$ *-Flair*	wm	**0.486**	**0.505**	**2.113**	**2.556**
		gm	**0.534**	**0.540**		
	T2-Flair + T1	wm	0.486	0.503	2.098	2.508
		gm	0.534	0.539		

Table 4. Registration results evaluated by Dist and Dice on *Iseg2017* and *MRBrain13*. The bold entries indicate the experiments which used the combination of the real and the translated images in another modality generated by the real images.

Experiments-Dice similarity coefficient, Distance Between Corresponding Landmarks (Dist)

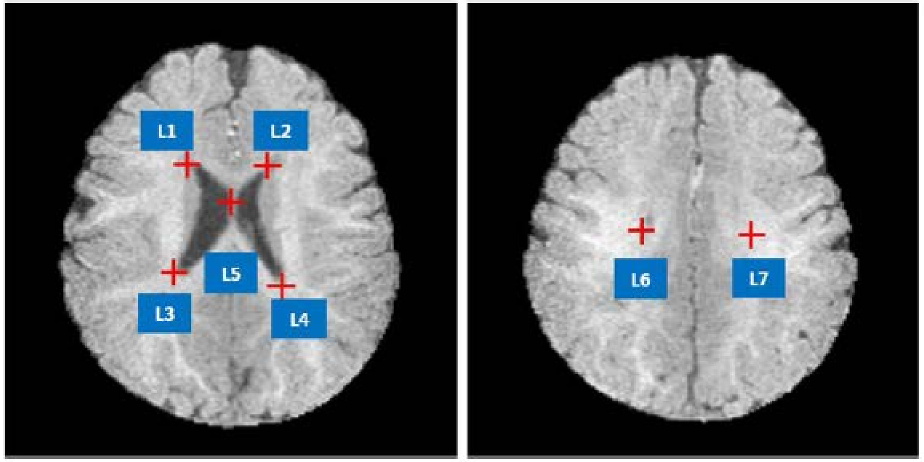


Figure 6. Illustration of the seven landmarks selected for cross-modality registration. L1: right lateral ventricle superior, L2: left lateral ventricle superior, L3: right lateral ventricle inferior, L4: left lateral ventricle inferior. L5: middle of the lateral ventricle, L6: right lateral ventricle posterior, L7: left lateral ventricle posterior.

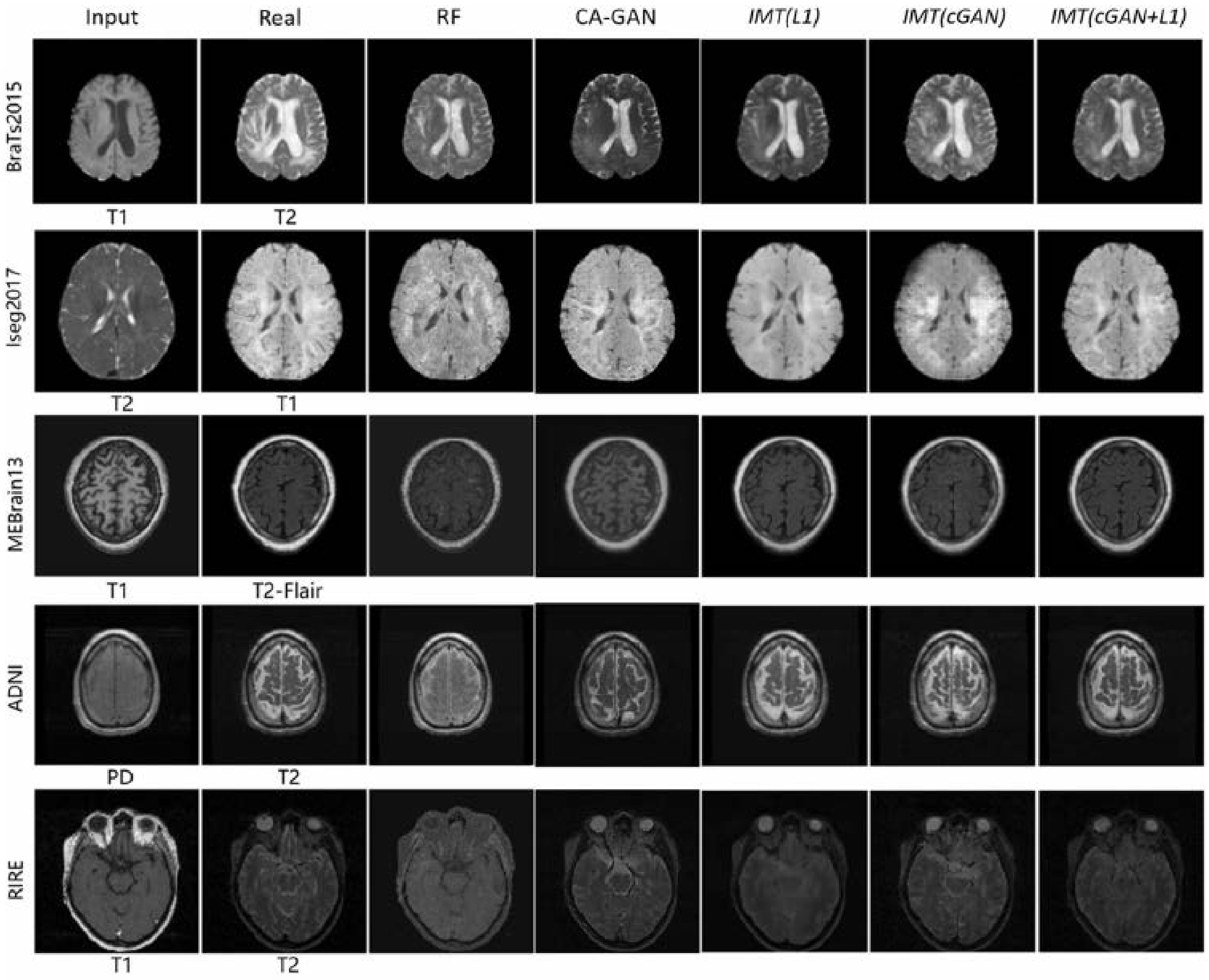


Figure 7. Samples of cross-modality generation results on five publicly available datasets including *BraTs2015*[60], *Iseg2017*[61], *MRBrain13*[62], *ADNI*[30], and *RIRE*[63]. Results are selected from top performing examples (relatively low MAE, high PSNR, high MI, and high PSNR collectively) with four approaches. The right five columns show results of the random-forests-based method (RF)[5], the Context-Aware GAN (CA-GAN)[30] and IMT framework with different loss functions ($L1$, $cGAN$, $cGAN + L1$).

Results

Datasets	Modalities	Structures	Dice	Dist
Iseg2017	T2	wm	0.823	0.475
		gm	0.859	
	$\hat{T}1$	wm	0.882	0.183
		gm	0.910	
	T2 + $\hat{T}1$	wm	**0.883**	**0.190**
		gm	**0.857**	
	T1	wm	0.868	0.179
		gm	0.898	
	$\hat{T}2$	wm	0.807	0.218
		gm	0.846	
	T1 + $\hat{T}2$	wm	**0.868**	**0.186**
		gm	**0.898**	
	T1 + T2	wm	0.868	0.184
		gm	0.898	
MRBrain13	T2-Flair	wm	0.976	0.182
		gm	0.976	
	$\hat{T}1$	wm	0.966	0.181
		gm	0.968	
	T2-Flair + $\hat{T}1$	wm	**0.971**	**0.180**
		gm	**0.974**	
	T1	wm	0.976	0.179
		gm	0.981	
	$\hat{T}2$-Flair	wm	0.985	0.180
		gm	0.983	
	T1 + $\hat{T}2$-Flair	wm	**0.985**	**0.179**
		gm	**0.985**	
	T2-Flair + T1	wm	0.978	0.178
		gm	0.982	

Table 5. Results of our additional registration experiments evaluated by Dist and Dice on *Iseg2017* and *MRBrain13* implemented by ANTS. The bold entries indicate the experiments which used the combination of the real and the translated images in another modality generated by the real images.

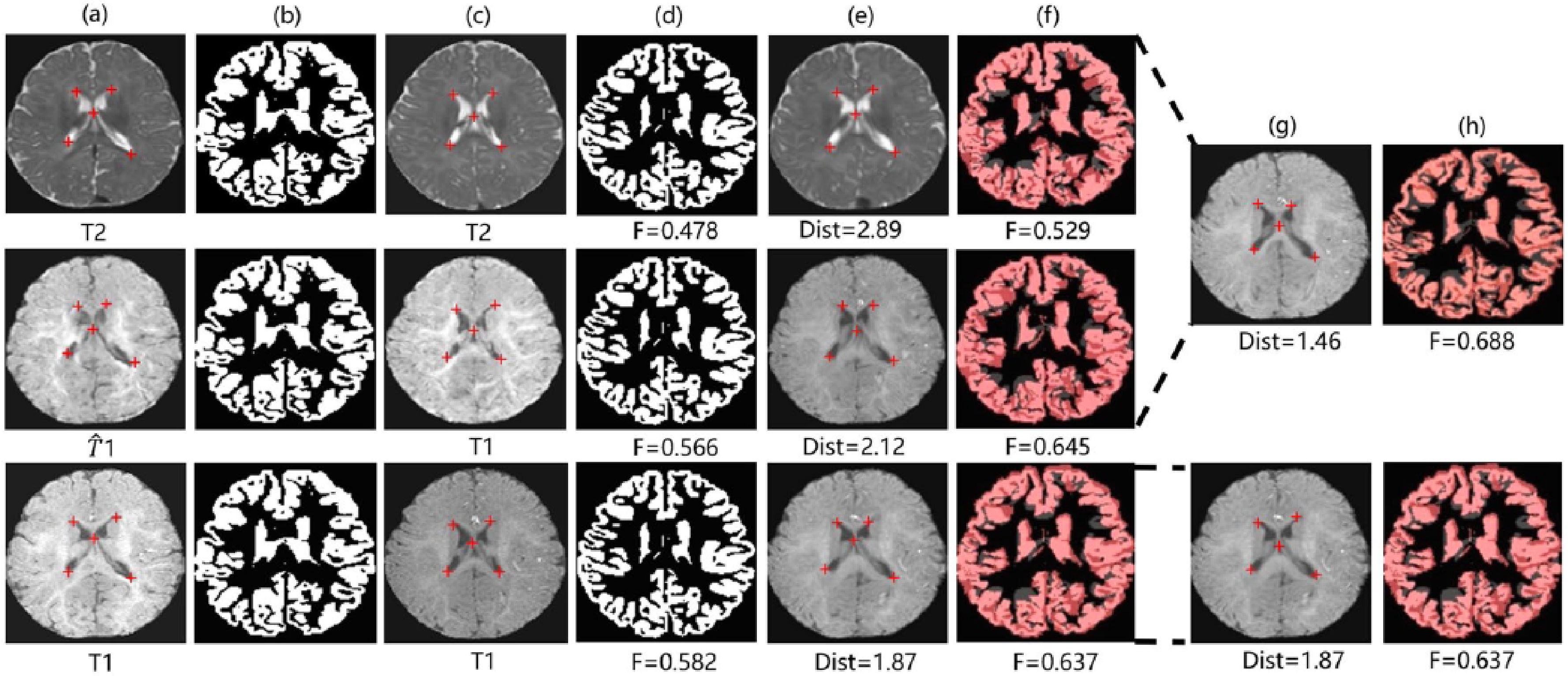


Figure 8. Samples of registration results of our method: (**a**) Fixed image, (**b**) Ground truth segmentation label of fixed image, (**c**) Moving image, (**d**) Ground truth segmentation label of moving image, (**e**) Warped image (moving image warped by the best traditional registration algorithm (ANTs), (**f**) Warped ground truth segmentation label of moving image, (**g**) Fused image, (**h**) Segmentation prediction of fused image. The pink, dark red, grey areas in (**f**) denote true regions, false regions, and missing regions respectively. The red crosses denote landmarks in the fixed and moving images.

Results

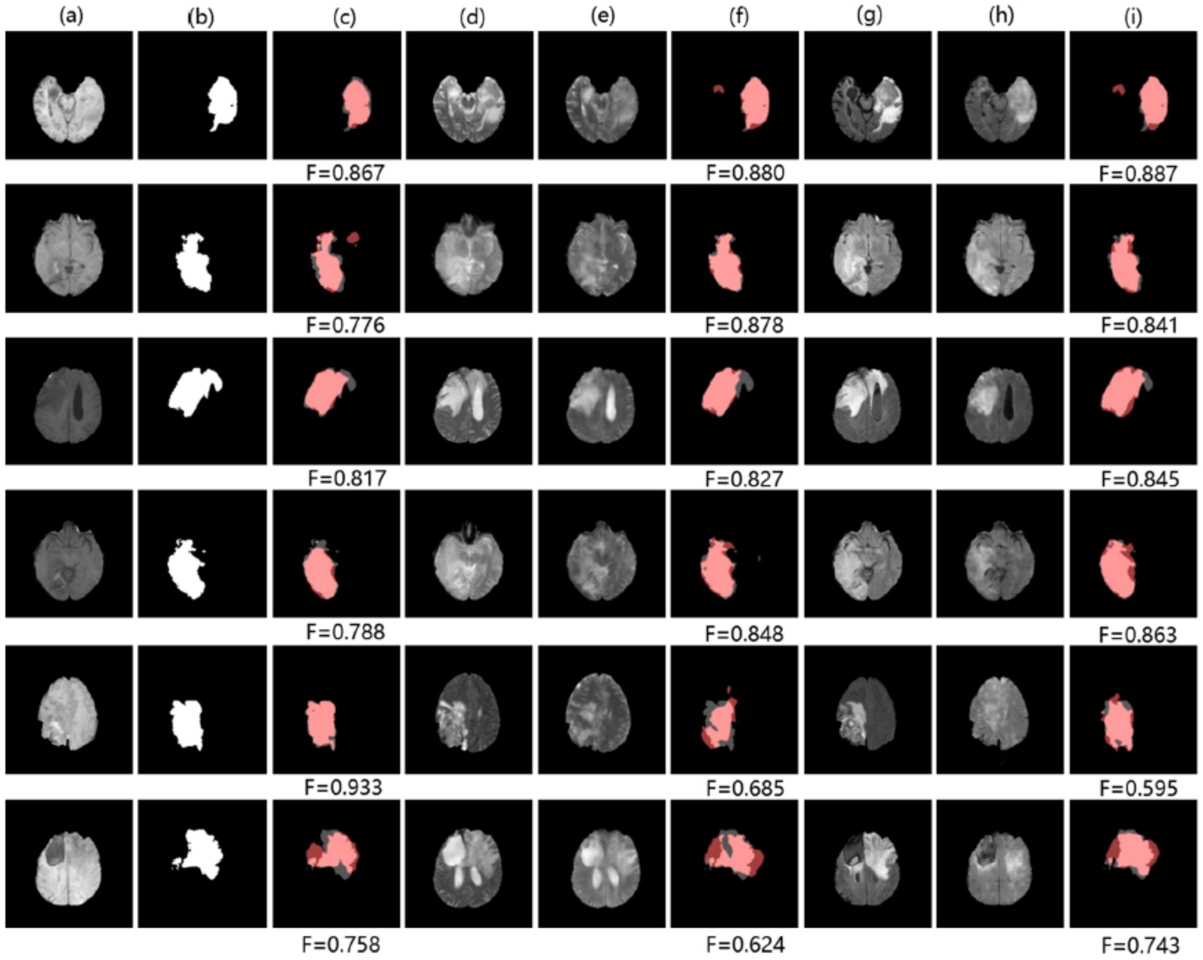


Figure 9. Samples of tumor segmentation results on *BraTs2015*: (**a,d,e,g,h**) denote T1 image, T2 image, $\hat{T}2$ image, T2-Flair image, $\hat{T}2$-*Flair* image. (**b**) Denotes ground truth segmentation label of T1 image. (**c,f,i**) Denote tumor segmentation results of T1 image using the FCN method, TMS (adding cross-modality information from $\hat{T}2$ image), TMS (adding cross-modality information from $\hat{T}2$-*Flair* image). Pink: true regions. Grey: missing regions. Dark red: false regions.

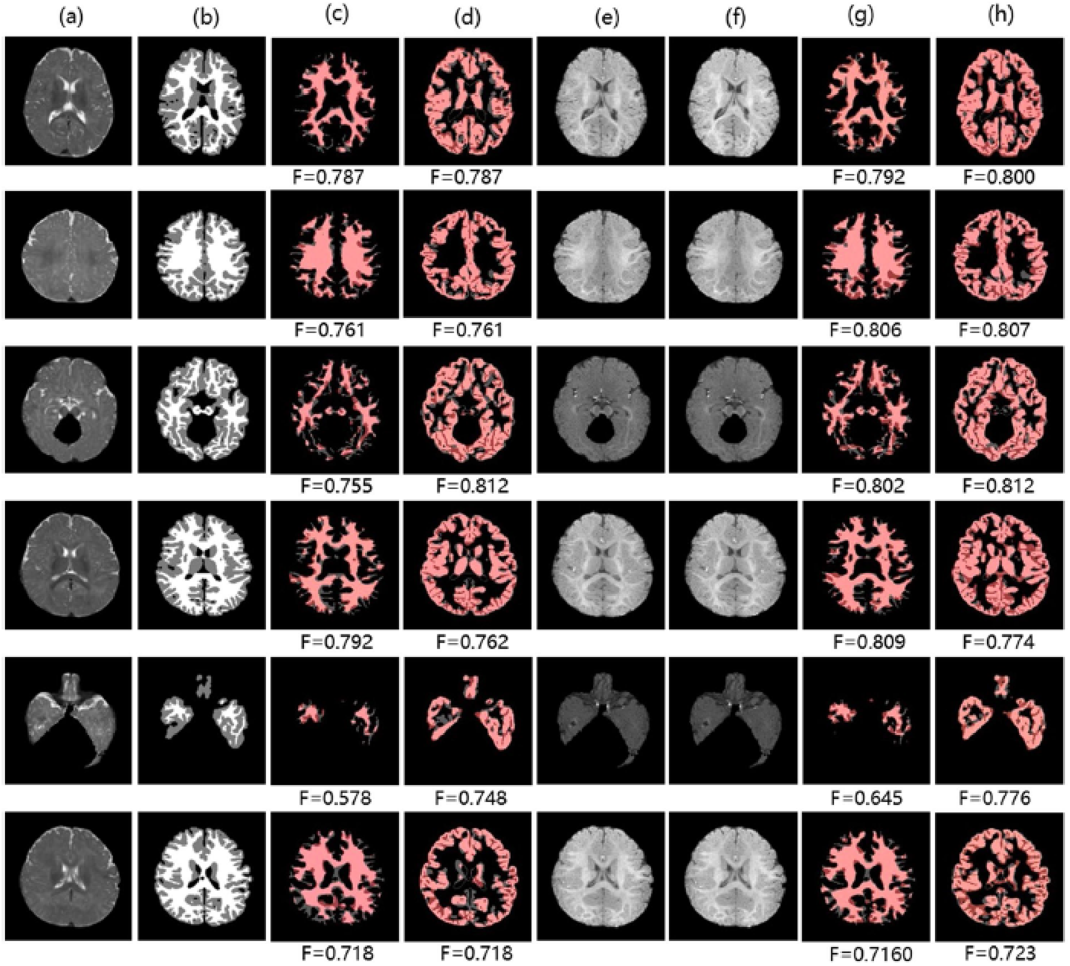


Figure 10. Samples of brain structure segmentation results on *Iseg2017*: (**a,e,f**) denote T2 image, T1 image, $\hat{T}1$ image. (**b**) Denotes ground truth segmentation label of T2 image. (**c,d**) Denote white matter and gray matter segmentation results of T2 image using the FCN method respectively. (**g,h**) Denote white matter and gray matter segmentation results of T2 image using TMS (adding cross-modality information from $\hat{T}1$ image) respectively. Pink: true regions. Grey: missing regions. Dark red: false regions.

Results

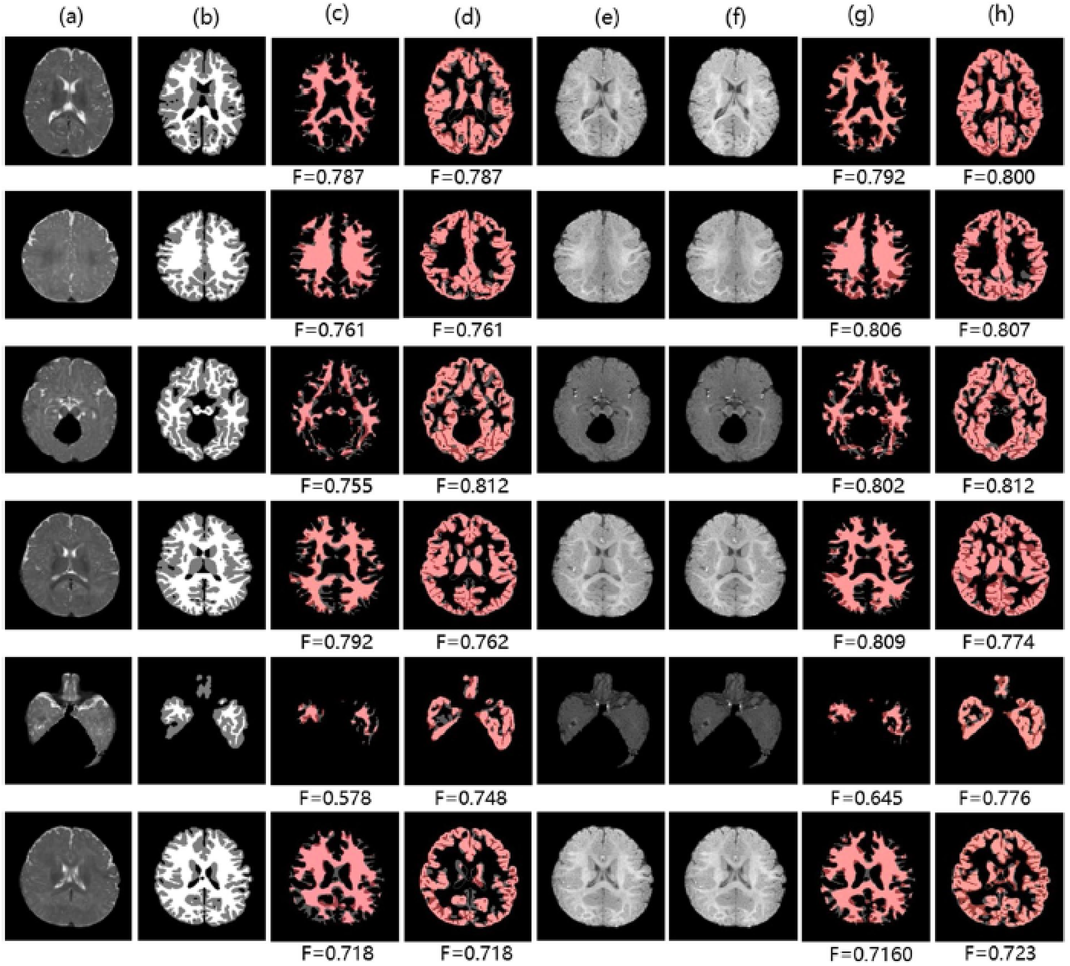


Figure 10. Samples of brain structure segmentation results on *Iseg2017*: (**a,e,f**) denote T2 image, T1 image, $\hat{T}1$ image. (**b**) Denotes ground truth segmentation label of T2 image. (**c,d**) Denote white matter and gray matter segmentation results of T2 image using the FCN method respectively. (**g,h**) Denote white matter and gray matter segmentation results of T2 image using TMS (adding cross-modality information from $\hat{T}1$ image) respectively. Pink: true regions. Grey: missing regions. Dark red: false regions.

	Dice(tumor)	Δ
T1	0.760	—
T1 + $\hat{T}$2	**0.808**	**6.32%**
T1 + T2	0.857	—
T1 + $\hat{T}$2-Flair	**0.819**	**7.89%**
T1 + T2-Flair	0.892	—

Table 6. Tumor segmentation results of TMS on *Brats2015*. "T1 + $\hat{T}2$" and "T1 + $\hat{T}$2-*Flair*" in bold font indicate our approach (TMS) where inputs are both T1 and $\hat{T}2$ images or T1 and $\hat{T}$2-*Flair* images. "T1" indicates the traditional FCN method where inputs are only T1 images. "T1 + T2" and "T1 + T2-Flair" indicate the upper bound. Δ indicates the increment between TMS and the the traditional FCN method.

	Dice(wm)	Δ	Dice(gm)	Δ
T2	0.649	—	0.767	—
T2 + $\hat{T}$1	**0.669**	**3.08%**	**0.783**	**2.09%**
T2 + T1	0.691	—	0.797	—

Table 7. Brain structure segmentation results of TMS on *Iseg2017*. "T2 + $\hat{T}1$" in bold font indicates our method (TMS) where inputs are both T2 and $\hat{T}1$ images. "T2" indicates the traditional FCN method where inputs are only T2 images. "T2 + T1" indicates the upper bound.

Results

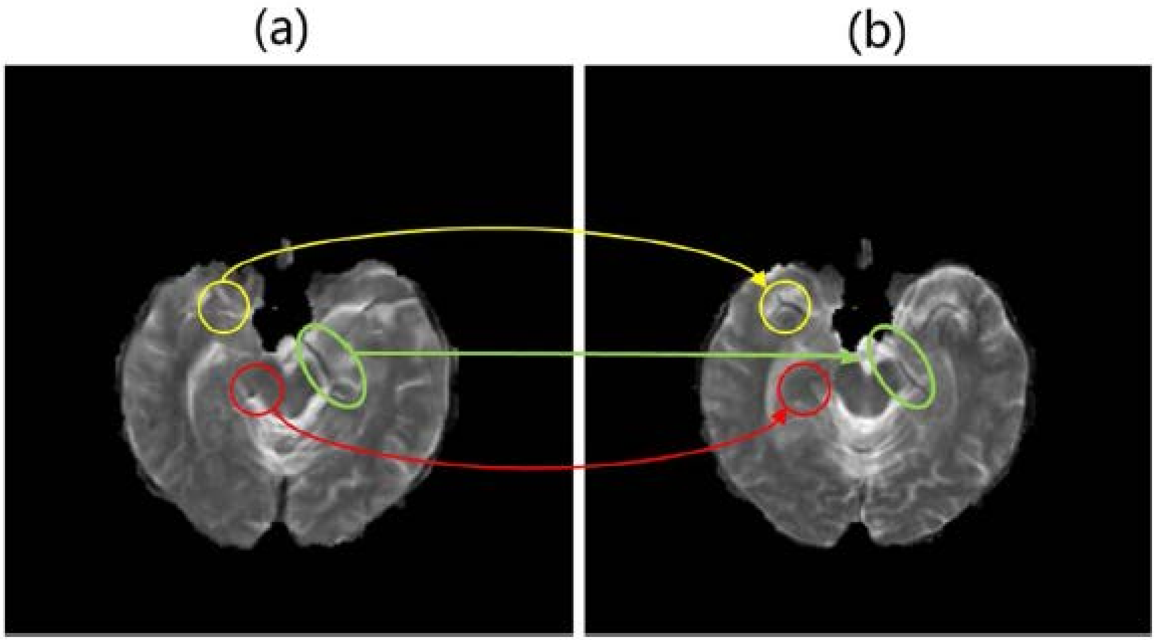


Figure 11. An abortive sample in our generation results: (**a**) $\widehat{T}2$. (**b**) T2. Circles in $\widehat{T}2$ indicate some misdescription of tiny structures. Circles in different colors indicate different problems.

Thank you!