Indescribable Multi-modal Spatial Evaluator

Lingke Kong
Manteia Tech
kid_liet@163.com

X. Sharon Qi
University of California, Los Angeles
xqi@mednet.ucla.edu

Qijin Shen
Fuzhou University
qijinshen@foxmail.com

Jiacheng Wang
Xiamen University
jiachengw@stu.xmu.edu

Jingyi Zhang
Xiamen University
zhangjingyi1@stu.xmu.edu.cn

Yanle Hu*
Mayo Clinic Arizona
Hu.Yanle@mayo.edu

Qichao Zhou*
Manteia Tech
zhouqc@manteiatech.com

Overview

In this study, we developed a self-supervised approach, Indescribable Multi-model Spatial Evaluator (***IMSE***), to evaluate the spatial error problem of multi-modal images.

IMSE visualization spatial error

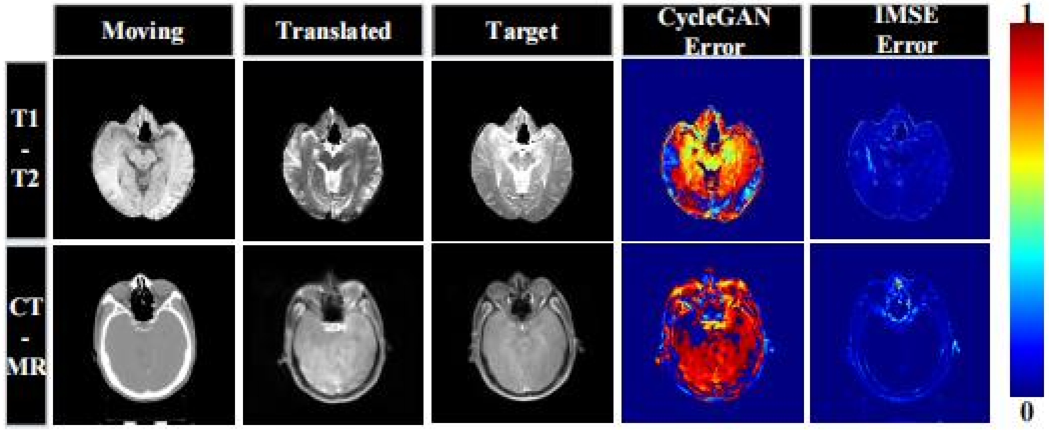

Figure 1. The distributions of the Moving images translated by CycleGAN still have large residual differences from the Target images. IMSE gives smaller error assessment values in the overlapping regions (converging to blue).

IMSE used for registration

IMSE can replace the traditional manually designed similarity operator such as mutual information to **optimize the registration based on deep learning and traditional registration**.

IMSE used for image-to-image translation

IMSE can be used to establish a new image-to-image translation paradigm. The translated image can have both the **spatial features of the source image** and **the modal features of the reference image.**

IMSE used for Spatial Error

IMSE can provide provide a convenient way to evaluate and **visualize registration results.**

IMSE

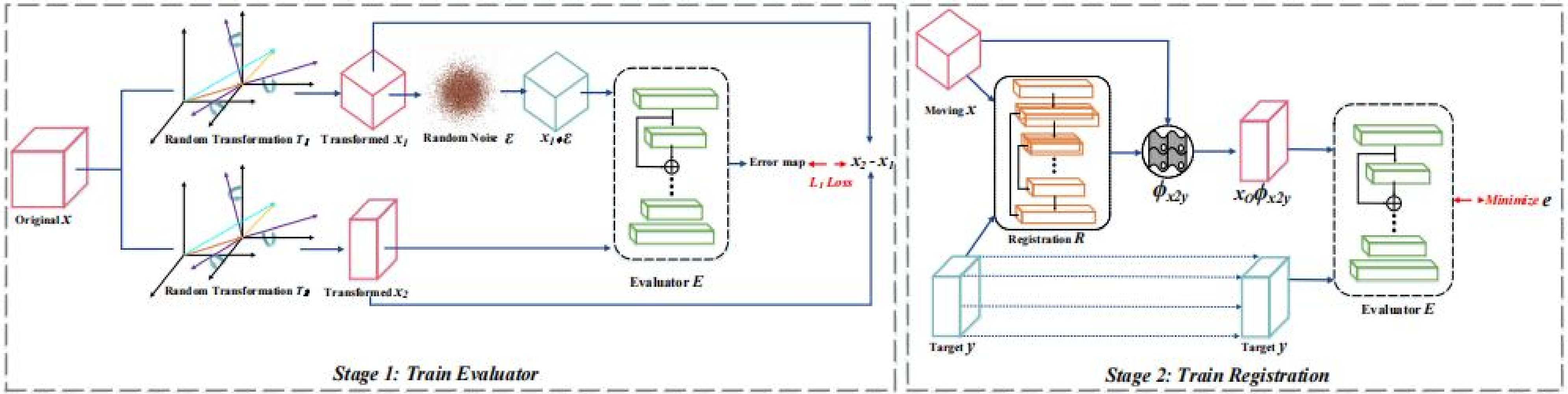

Fig 2. A general overview of the IMSE process. It is divided into two main parts, the training evaluator and the training registration

Stage 1： Train Evaluator

$$x_1 = x \circ T_1. \qquad x_2 = x \circ T_2.$$

$$\min_E \mathcal{L}_{L_1}(E) = \mathbb{E}_{x_1,x_2}\left[\|E(x_2, x_1+\varepsilon), (x_2-x_1)\|_1\right].$$

x: original image
T: random spatial transformations
E: evaluator
ε : random noise(Shuffle Remap)

Stage 2： Train Registration

$$\min_R \mathcal{L}_{sim}(R) = \mathbb{E}_{x,y}\left[\|E(x \circ R(x,y), y)\|_1\right].$$

$$\min_R \mathcal{L}_{smooth}(R) = \mathbb{E}_{x,y}\left[\|\nabla R(x,y)\|^2\right].$$

x and y: images with different modalities
R: registration network
∇R: regularization constraint

Shuffle Remap

Algorithm 1 Pseudocode of Shuffle Remap in PyTorch style.

```
# X: the input image and the range is [-1,1]
# r_min: Minimum number of random control points
# r_max: Maximum number of random control points

# number of randomly generated control points
control_point=random.randint(r_min, r_max)
# normalize to the range of the image distribution
dist=torch.rand(control_point)*(1-(-1))+(-1)
# sort from small to large
dist=torch.sort(dist)
# Add endpoint -1 and 1
dist=torch.cat([torch.tensor([-1]), dist])
dist=torch.cat([dist, torch.tensor([1])])
# shuffle the distribution and generate empty new image
shuffle_remap=torch.randperm(control_point+1)
new_X=torch.zeros_like(x)
for i in range(control_point +1):
    target_part=shuffle_remap[i]
    min1, max1=dist[i], dist[i+1]
    min2, max2=dist[target_part], dist[target_part+1]
    # get the coordinates corresponding to the distribution
    coord=torch.where((min1<=X)&(X<max1))
    # Eq.(8)
    new_X[coord]=((X[coord]-min1)/(max1-min1))*
                    (max2-min2)+min2
return   new_X
```

$$x' = \frac{x - p_i}{p_{i+1} - p_i} * (p_{j+1} - p_j) + p_j.$$

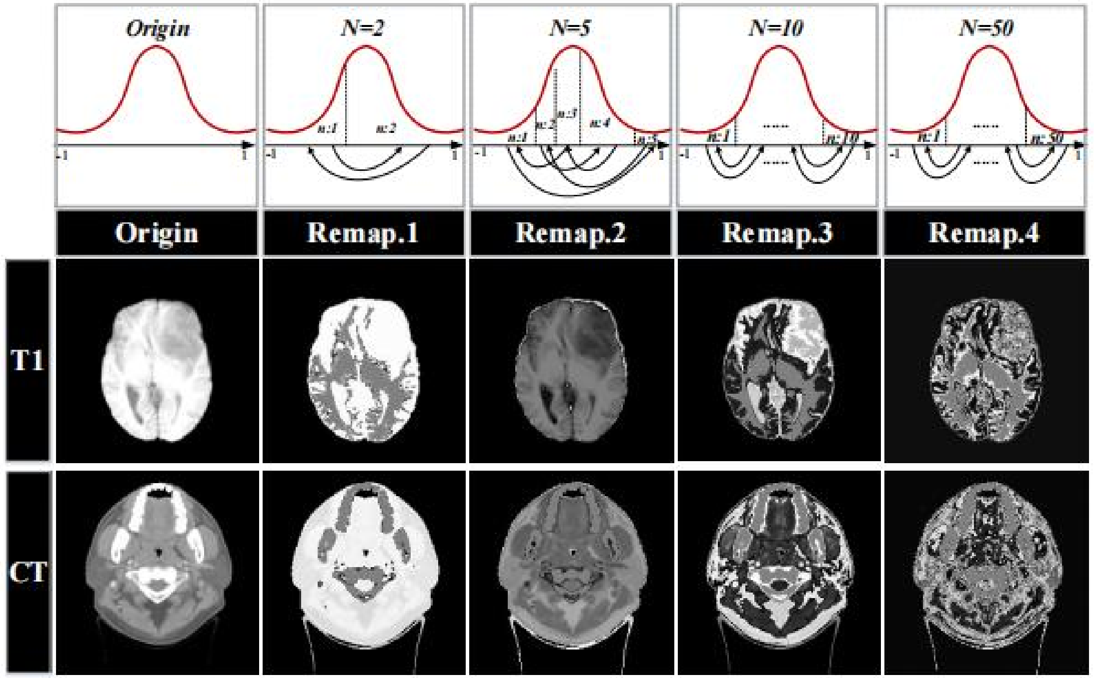

Fig 3. Examples of T1 and CT after Shuffle Remap.

It is worth noting that we only provide a style enhancement method, and Shuffle Remap is not irreplaceable in the IMSE architecture. Shuffle Remap is a pure style enhancement method, which can be easily combined with other domain generalization methods.

IMSE used for registration

IMSE for Registration Based on Neural Network

Moving → Target	Methods	2D Dice ↑	2D HD95 ↓	2D $\|\nabla\phi\|_2$ ↓	3D Dice ↑	3D HD95 ↓	3D $\|\nabla\phi\|_2$ ↓
	Initial	0.68 ± 0.08	4.17± 1.76	✗	0.81± 0.05	4.13± 1.53	✗
T1 → T2	NCC	0.75 ± 0.05	2.71± 1.33	0.0026	0.84 ± 0.03	3.38 ± 1.03	0.0036
	MI	0.82 ± 0.04	1.70± 1.29	0.0102	0.86± 0.04	2.79 ± 1.44	0.0162
	MIND	0.83 ± 0.04	1.66± 1.27	**0.0023**	0.88± 0.02	2.70 ± 1.01	0.0034
	CycleGAN	0.85 ± 0.03	1.37± 0.95	0.0061	0.88± 0.02	2.94 ± 0.93	0.010
	RegGAN	0.86 ± 0.03	1.25 ± 0.90	0.0091	0.89± 0.01	2.85 ± 0.83	0.0090
	IMSE(BC)	0.83 ± 0.04	1.72 ± 1.30	0.0125	0.84± 0.04	3.21 ± 1.33	0.0105
	IMSE(SR)	**0.89 ± 0.02**	**1.06± 0.87**	**0.0023**	**0.91± 0.01**	**2.36 ± 0.77**	**0.0032**
T2 → T1	NCC	0.74 ± 0.04	2.76± 1.35	0.0026	0.84± 0.03	3.07 ± 1.04	0.0038
	MI	0.79 ± 0.05	1.58± 1.31	0.0103	0.88± 0.04	2.33 ± 1.41	0.0166
	MIND	0.81 ± 0.03	2.15± 1.28	0.0023	0.88± 0.02	2.42 ± 1.05	0.0035
	CycleGAN	0.86 ± 0.04	1.19± 0.94	0.0056	0.88± 0.02	2.95 ± 0.95	0.009
	RegGAN	0.86 ± 0.03	1.20± 0.90	0.0071	0.89± 0.01	2.71 ± 0.80	0.0085
	IMSE(BC)	0.80 ± 0.04	1.87 ± 1.34	0.0122	0.85± 0.03	3.01 ± 1.26	0.0097
	IMSE(SR)	**0.89 ± 0.02**	**0.85± 0.86**	**0.0022**	**0.91± 0.01**	**2.21 ± 0.75**	**0.0025**

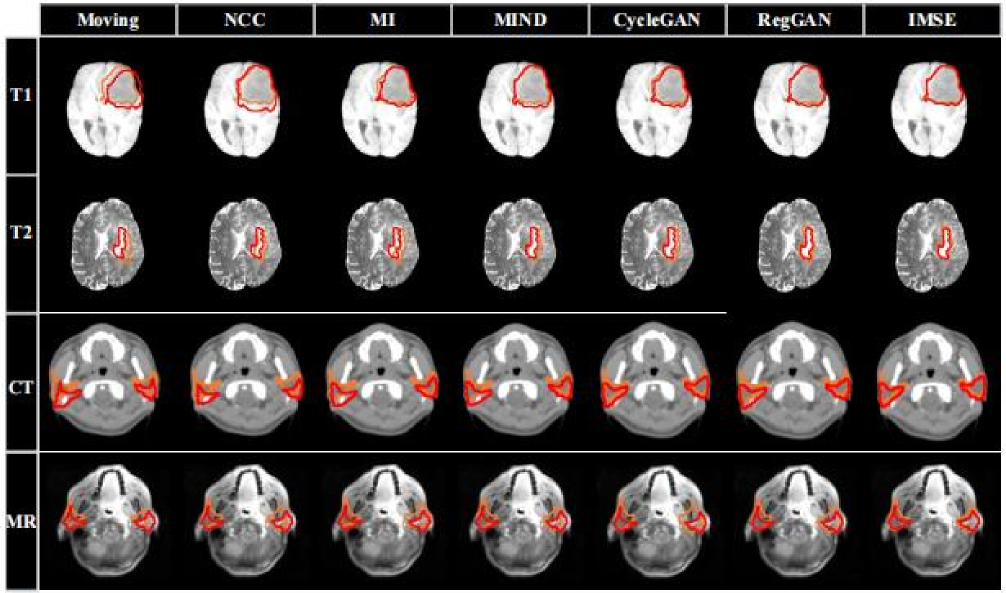

Fig 4. Registration results of various methods based on the T1-T2 and CT-MR dataset. Initial indicates the results before registration. The source data used to train the IMSE were T1 and CT.

IMSE for Traditional Registration

IMSE can be readily integrated into the traditional registration process, such as replacing the L_{sim} function with a trained evaluator. By first initializing a deformation field of 0 and then optimizing it through similarity loss (Similarity operator or IMSE) and regularization loss.

$$\hat{\phi} = \arg\min_{\phi} \mathcal{L}_{sim}\left(M\left(\phi\right), T\right).$$

$$\min_{R} \mathcal{L}_{smooth}\left(R\right) = \mathbb{E}_{x,y}\left[\left\|\nabla R\left(x,y\right)\right\|^2\right].$$

Moving→Target	Methods	Dice ↑	HD95 ↓	$\|\nabla\phi\|_2$ ↓
T2 → T1	NCC	0.84 ± 0.03	3.40± 1.14	0.016
	MI	0.85 ± 0.03	3.26± 1.22	0.06
	MIND	0.89 ± 0.02	3.01± 1.16	**0.002**
	IMSE	**0.91± 0.01**	**2.62± 0.81**	**0.002**
MR → CT	NCC	0.54 ± 0.03	5.18± 1.20	0.011
	MI	0.55 ± 0.02	5.29± 1.18	0.05
	MIND	0.55 ± 0.02	4.93± 1.24	0.008
	IMSE	**0.61± 0.01**	**4.37± 0.81**	**0.007**
T1 → T1	MAE	0.64 ± 0.03	8.38± 1.43	0.018
	IMSE	**0.67 ± 0.02**	**7.4± 1.22**	**0.005**
CT → CT	MAE	0.56 ± 0.02	4.92± 1.07	0.012
	IMSE	**0.59 ± 0.02**	**4.80±0.87**	**0.005**

IMSE for Image-to-image Translation

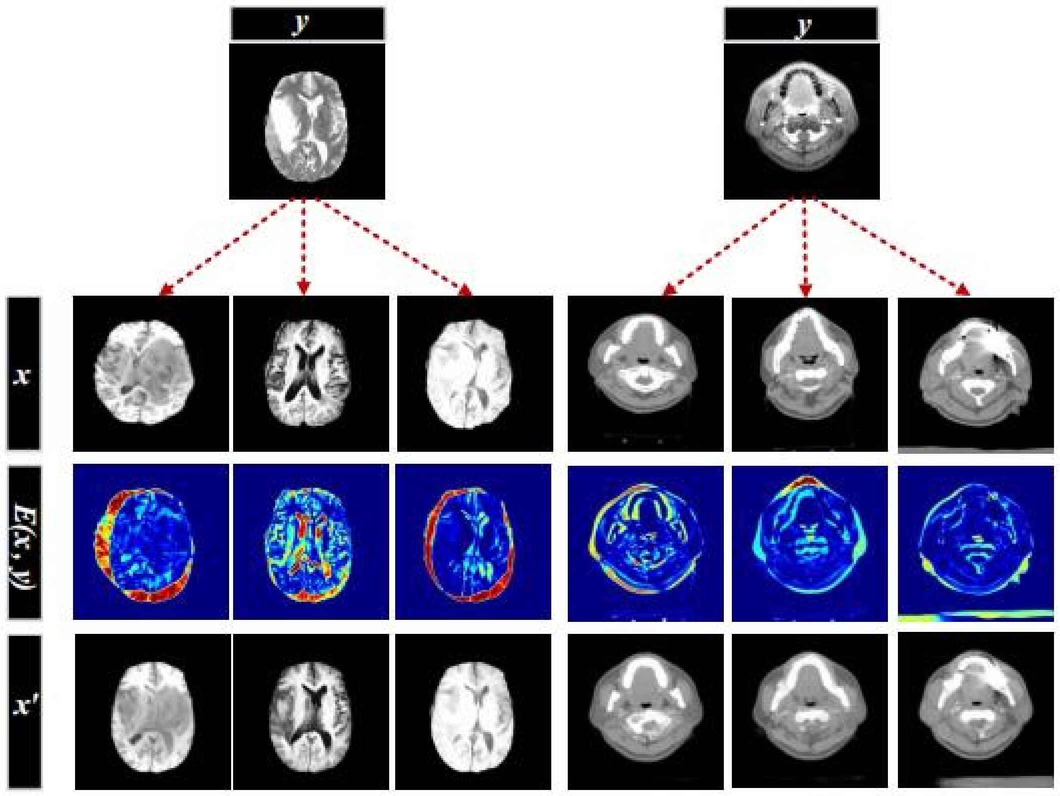

Fig 5 . IMSE is used for image to image translation. Where, y is the source image, x is different reference images, E(x, y) is the prediction result of IMSE and x'= x − E(x, y).

Source→Target	Methods	NMAE ↓	PSNR ↑	SSIM ↑
T2 → T1	CycleGAN	0.088	24.1	0.89
	RegGAN	0.071	25.5	0.90
	IMSE	**0.029**	**91.6**	**0.96**
MR → CT	CycleGAN	0.049	22.9	0.88
	RegGAN	0.041	24.1	0.89
	IMSE	**0.022**	**41.3**	**0.93**

IMSE based image-to-image translation differs from GAN based translation in two aspects:

1) The result of image-to-image translation based on GAN is unique whereas IMSE based depends on the characteristics of the reference image.

2) GAN based image-to-image translation requires two modal data whereas IMSE based translation only uses one modal data for training.

IMSE for Spatial Error

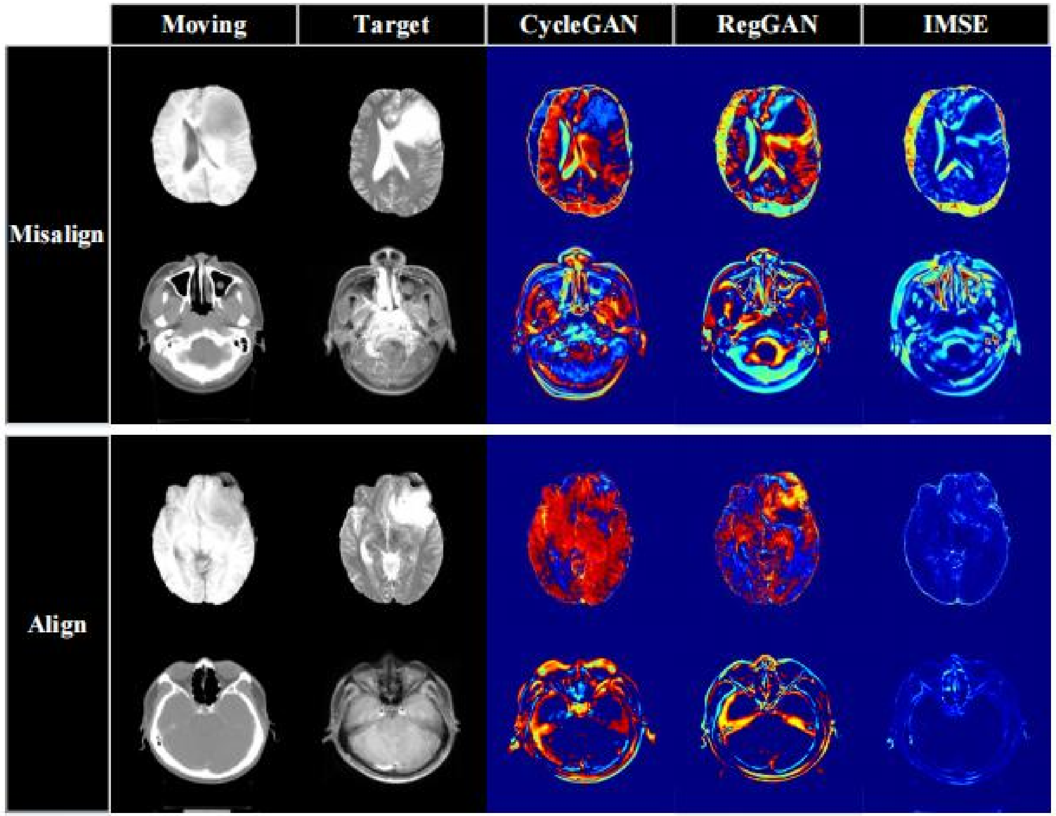

Fig 6 . Demonstration of errors estimated by the CycleGAN, RegGAN and IMSE methods in both misaligned and aligned cases.

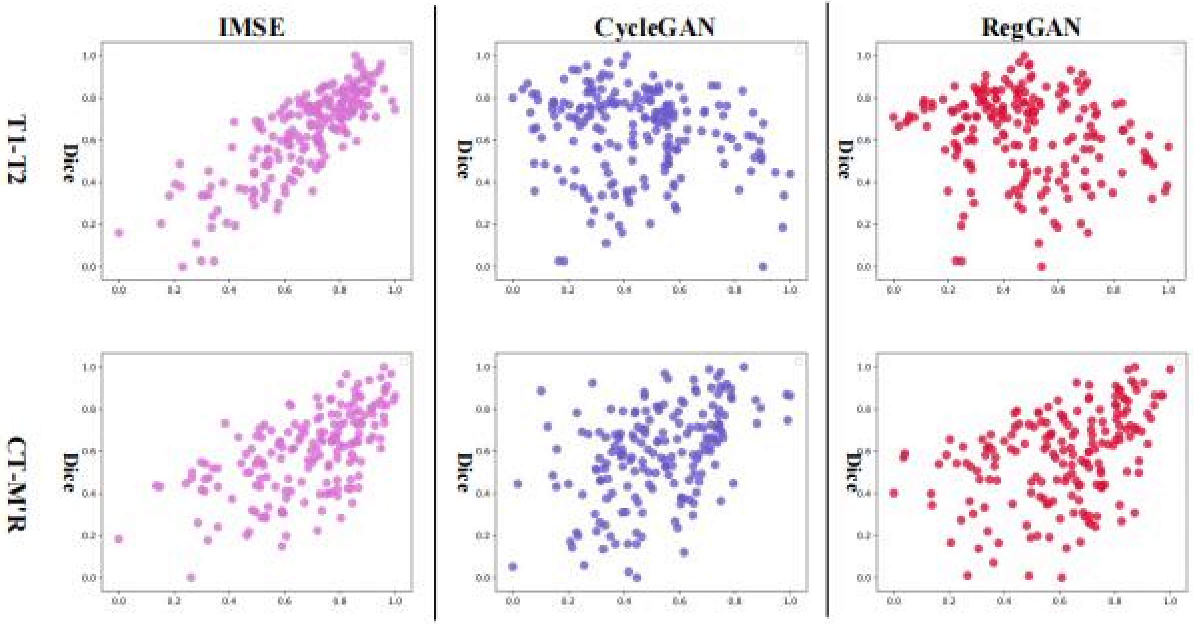

Fig 7 . Correlation of estimated spatial errors with Dice for the IMSE, CycleGAN, RegGAN methods.

The main advantage of IMSE is that it focuses only on differences in spatial location while ignoring differences in multi-modal distribution caused by different image acquisition mechanisms.