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

1. Training Details

All the experiments were implemented in Pytorch software on 64-bit Ubuntu Linux system with 96GB RAM and 24GB Nvidia Titan RTX GPU. All the images were normalized to [-1, 1]. We train all methods using the Adam optimizer with the learning rate of 1e-4 and (β_1 ,β_2) = (0.5, 0.999). In the case of 2D and 3D, the batch size is set to 8 and 1 respectively, and the weight decay is 1e-4. The weight of all similarity losses is set to 1.

Normalized Cross-correlation (NCC) is used to describe the correlation between two vectors or samples of the same dimension. As shown in Eq 1.

$$\frac{1}{N}\sum_{x,y}\frac{1}{\sigma_f\sigma_t}\left(f\left(x,y\right)-\mu_f\right)\left(t\left(x,y\right)-\mu_t\right) \tag{1}$$

Where, $f()$ and $t()$ are two vectors or samples, and n is vector dimension or window size, σ is the standard deviation of various samples, μ is the mean value of each sample.

Mutual Information(MI) is used to describe the degree of interdependence between variables. As shown in Eq 2.

$$\sum_{x,y}p\left(x,y\right)log\frac{p\left(x,y\right)}{p\left(x\right)p\left(y\right)} \tag{2}$$

$p(x)$ and $p(y)$ are the probability distributions of values in X and Y, respectively.

Modal Independent Neighborhood Descriptor(MIND) is used to describe the local modal characteristics around each voxel. First, D_p is defined as the similarity measure between image patches.

$$D_p(I,x_1,x_2)=\frac{1}{|P|}\sum_t(I(x_1+t)-I(x_2+t))^2 \tag{3}$$

I is the image, x_1 and x_2 are the two locations in the image, and P is the displacement set from voxel with patch size $p*p*p$ to the center of the patch. Therefore, D_p calculates the mean square deviation between two image patches centered on x_1 and x_2. Next is the definition of MIND:

$$MIND_p(I,x,r)=exp\left(\frac{-D_p(I,x,x+r)}{V(I,x)}\right) \tag{4}$$

Where r is the distance vector. $V(I,x)$ is the evaluation of local variance. We make MIND a Gaussian function of D_p, that is, low response when patches are not similar, and high response when patches are similar. Finally, we want to align the mean value of the absolute difference between the MIND of the two images.

$$\frac{1}{|R|}\sum_r|MIND(A,x,r)-MIND(B,x,r) \tag{5}$$

CycleGAN uses two downsampling convolution blocks, nine residual blocks, two up-sampling deconvolution blocks and four discriminator layers. The CycleGAN was developed which was based on the assumption that the generator G from the source domain X to the target domain Y ($G : X \to Y$) was the reverse of the generator F from Y to X ($F : Y \to X$).

$$\begin{split}\min_G\min_F\mathcal{L}_{Cyc}\left(G,F\right)=\mathbb{E}_x\left[\|F\left(G\left(x\right)\right)\right.\\ \left.-x\|_1+\mathbb{E}_y\left[\|G\left(F\left(y\right)\right)-y\|_1\right]\right.\end{split} \tag{6}$$

$$\begin{split}\min_G\max_D\mathcal{L}_{Adv}\left(G,D\right)=\mathbb{E}_y\left[log\left(D\left(y\right)\right)\right]\\ +\mathbb{E}_x\left[log\left(1-D\left(G\left(x\right)\right)\right)\right]\end{split} \tag{7}$$

*Corresponding author.

RegGAN based on "loss-correction" shown in Eq 8. The solution corrects the output of the generator $G(x_n)$ by modeling a noise transition ϕ to match the noise distribution. Previously. Also, the RegGAN add the adversarial loss between the generator and the discriminator (Eq 7),

$$\hat{G} = \arg\min_{G} \frac{1}{N} \sum_{n=1}^{N} \mathcal{L}\left(\phi \circ G\left(x_n\right), \widetilde{y}_n\right) \tag{8}$$

2. Weight of smooth loss

In addition to smoothness loss weight, all our comparison methods adopt the same network structure and the same parameter settings. Because the loss magnitude obtained by different methods of calculating similarity is also different, it is necessary to adjust the ratio between similarity and smoothness for each method. The settings are shown in Table 1.

Method / Weight	NCC	MI	MIND	CycleGAN	RegGAN	IMSE
λ	5	8	1	1	1	1

Table 1. Different smoothness loss weights for different methods.