

基于生成对抗网络的多模态MR图像缺失模态合成

徐盼盼,张东,袁达龙

武汉大学物理科学与技术学院,湖北 武汉 430072

【摘要】提出一种基于生成对抗网络的多模态磁共振(MR)图像缺失模态合成方法,在不同缺失情况下均能预测出完整的模态序列。该模型将生成对抗网络作为模态合成的基本框架,生成对抗网络由生成器和判别器2个部分组成。其中生成器有3个阶段:特征提取、特征融合和图像生成。在特征提取过程中,使用U-Net网络进行编码。然后将4个模态的特征进行融合,并将融合结果输入混合注意力模块。这样使网络能够自适应地调整不同模态的权重,最后进行图像生成。判别器部分由4个相同的判别器组成,对4种不同的模态进行判别,其网络结构基于PatchGan进行改进。通过与其他具有代表性的模态合成算法比较,此方法无论在视觉效果还是客观评价指标上都取得了更好的效果,其中峰值信噪比均在23 dB以上,结构相似性均在0.99以上。

【关键词】磁共振成像;生成对抗网络;混合注意力模块;多模态合成

【中图分类号】R318;TP391

【文献标志码】A

【文章编号】1005-202X(2023)07-0827-06

Incomplete multimodality MR image synthesis based on generative adversarial network

XU Panpan, ZHANG Dong, YUAN Dalong

School of Physics and Technology, Wuhan University, Wuhan 430072, China

Abstract: An incomplete modality synthesis method for multimodality magnetic resonance (MR) images based on generative adversarial network (GAN) is proposed to predict the complete modal sequence in different scenarios. The proposed model uses GAN which consists of generator and discriminator as the backbone of the network. The generator contains feature extraction, feature fusion and image generation. U-Net is used for encoding during feature extraction, and the features of the 4 modalities are fused before being input into the hybrid attention module, so that the network can adaptively adjust the weights for different modalities and finally generate images. The discriminator consists of 4 same discriminators which network structure is improved from PatchGan to discriminate different modalities. Compared with other state-of-the-art modality synthesis algorithms, the proposed method achieves better results in both visual effects and objective criteria, with peak signal-to-noise ratios higher than 23 dB and structural similarity index measurements above 0.99.

Keywords: magnetic resonance imaging; generative adversarial network; hybrid attention module; multimodal synthesis

前言

磁共振成像(MRI)是一种常用的无创成像技术,广泛应用于病理诊断、图像分割等任务^[1-5]。MRI有多种模态,如 T_1 加权(T_1)、 T_1 增强(T_{1ce})、 T_2 加权(T_2)和液体衰减反转恢复序列(Flair)。然而在现实情况中,由于扫描费用、患者身体状况等原因,医生常用于诊断肿瘤的某些模态无法得到。另外,在已经获得的模态序列中,也会因为存在着噪声和人工干扰

等因素导致图像无法使用。正如前面提到的,不同模态所包含的信息是不同的,模态序列缺失会造成MR图像分析、诊断和研究判断失误的风险。因此,在影像诊断、治疗规划和研究工作中,合成缺失或受损的模态是十分必要的。

目前医学图像合成主要有两类方法:第1种方法是根据图像配准对相关图像图谱配准来合成,然而图谱一般建立在健康人的图像之上,当异常组织存在时,会造成图像的重大变形。第2种是基于深度学习的方法,可以缓解变形问题,如字典学习^[6]和卷积神经网络(CNN)^[7]等。通过搭建合适的网络框架,对病理样本进行有效的特征提取,通过这种方式合成肿瘤影像。Yu等^[8]提出基于条件生成对抗网络(cGAN)和局部自适应融合的方法,从 T_1 合成Flair图像。该方法将边缘

【收稿日期】2023-02-08

【基金项目】国家重点研发计划973项目(2011CB707900)

【作者简介】徐盼盼,硕士,研究方向:医学图像处理, E-mail: panpan_w@outlook.com

信息进行融合,以反映图像的纹理结构,并绘制出图像中不同对象的边界,但该方法只考虑了T₁到Flair之间的合成。Zhou等^[9]提出多模态MR图像合成网络生成缺失的MR图像,该方法引入了二维和三维自监督任务,在融合特征时使用通道注意机制,使网络能够自适应地调节不同的融合操作,以学习所有模态的共同表示。现有的MR图像合成方法多数需要对不同模态的缺失情况进行指定,针对MR图像的4种模态,需要训练14种模型。针对上述问题,本研究提出了一种基于生成对抗网络的多模态MR缺失模态合成算法。将不存在的模态使用全零图像替代,与存在模态拼接输入网络。

因此生成器可以灵活处理不同情况的输入,只需1个网络模型就能应对14种缺失情况。

1 方法

1.1 算法框架

图1为本文的算法流程,整体算法流程大致可以分为3个步骤。(1)数据准备:包含数据获取、数据集的划分和数据预处理。(2)搭建基于生成对抗网络的图像生成网络,并选取合适的损失函数对网络参数进行训练。(3)在测试数据集上用训练好的模型进行预测,并分析预测结果。

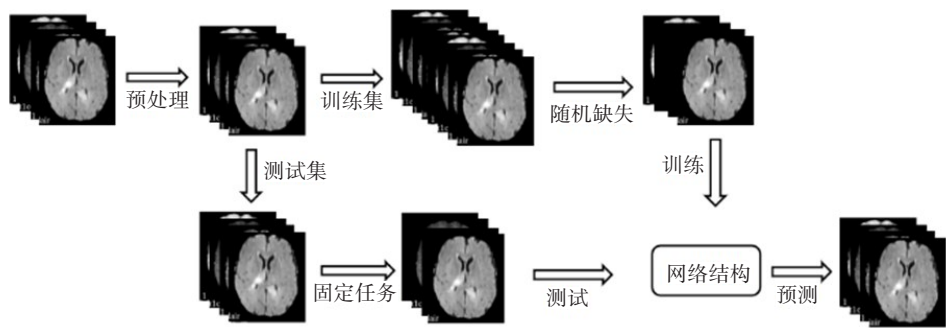


图1 算法流程
Figure 1 Algorithm flowchart

1.2 网络结构

设计的网络结构如图2所示。该网络主要由2部分组成,分别为左边的生成器和右边的判别器,也就是比较经典的生成对抗网络。生成器又分为3个部分:特征提取、特征融合和图像生成。特征提取和图像生成的部分由U形网络(U-Net)^[11]构成,特征融

合使用逐元素取最大值和混合注意力机制(Convolutional Block Attention Module, CBAM)^[10],判别器部分由PatchGan修改而来。网络的输入由随机缺失后的4个模态拼接而来,缺失的图像使用全零图像代替,输出为预测的4个完整的模态。

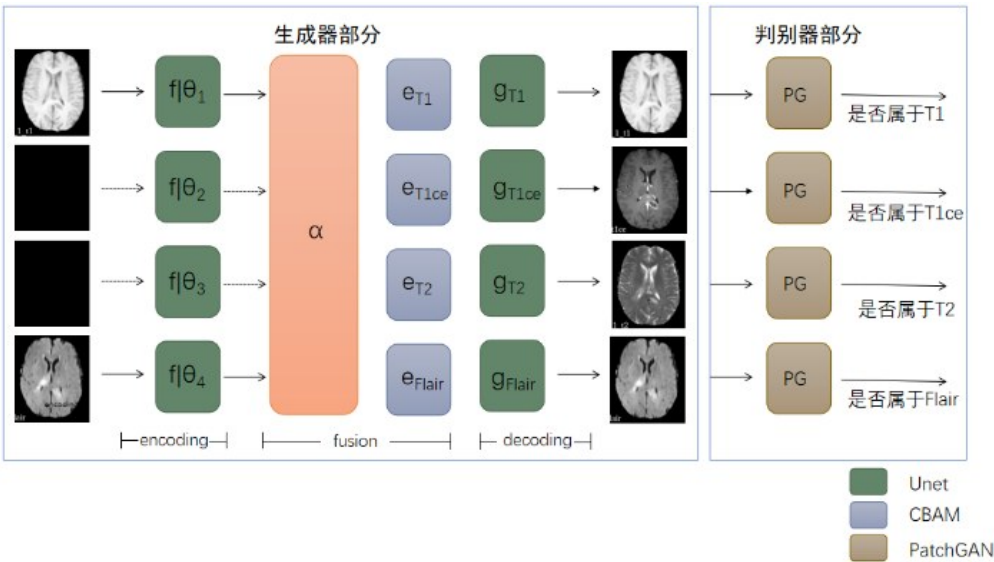


图2 网络结构示意图
Figure 2 Schematic diagram of network structure

为模型的每种输入模态准备了一个独立的编码器,编码部分采用的是U-Net网络。U-Net是一种在图像分割、图像重建中常见的网络结构,在2015年由Ronneberger等提出。其网络结构如图3所示,包含编码器和解码器两个部分。编码器部分进行特征提

取,由2个3×3的卷积、ReLU激活函数和1个2×2的最大池化组成。解码器部分进行上采样,由1个上采样的卷积、特征拼接和2个3×3的卷积组成。图像生成部分与编码部分类似。

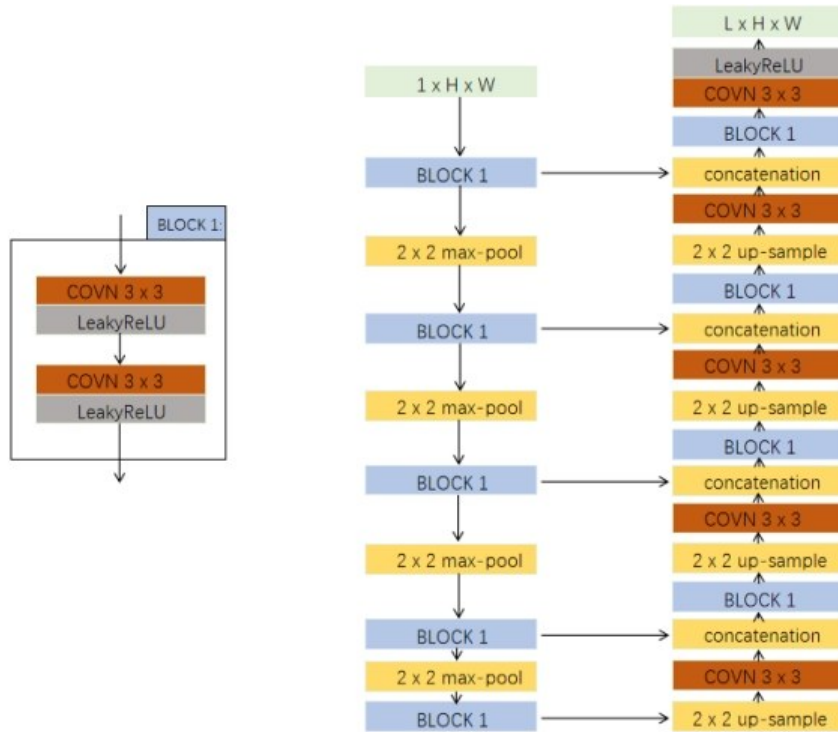


图3 U-Net网络架构图

Figure 3 U-Net architecture

在融合步骤中,模型使用融合操作 α ,将编码器产生的每个单独表示组合成1个融合表示,称之为 Z_a 。融合步骤增强了模型对于任意缺失模态的鲁棒性,它将任意数量的潜在表示作为输入,并返回1个单一的融合潜在表示。融合表示应该对不同数量的输入具有鲁棒性,如果某些输入模式缺失,它应该容纳这些缺失的输入。具体来说,我们的目标是给定4个模态的任何子集,网络产生1个融合的潜在表示。使用Max函数将提取的特征进行融合,融合成最终的潜在表示。Max函数意味着在每个通道中,潜在表示的每个像素与原始输入模态息息相关。融合部分的公式如下:

$$Z_a = \text{Max}(Z_1, Z_2, Z_3, Z_4) \quad (1)$$

混合注意力模块主要包括通道注意力^[12]和空间注意力^[13]两个部分,使得网络能够自适应地选择重要的对象和区域,结构如图4a所示。输入的特征图F通过通道注意力得到1×1×C的通道注意力特征图。再将通道注意力的输出与原始的输入进行矩阵相乘

得到 F_c 。 F_c 通过空间注意力得到H×W×1的空间注意力特征图。再将空间注意力的输出与 F_c 相乘。整体的注意力公式如下:

$$F_c = M_c(F) \otimes F \quad (2)$$

$$F' = M_s(F_c) \otimes F_c \quad (3)$$

通道注意力可以确定网络需要关注的输入特征图的特征,其结构如图4b所示。整个过程为:将输入的特征图分别经过H×W的最大池化和平均池化,分别得到1个1×1×C的特征图。再将它们分别输入全连接神经网络(Multilayer Perceptron, MLP)^[14],然后再将MLP的输出结果相加进行混合,并将混合的特征图经过Sigmoid激活函数,其中MLP网络中激活函数均为ReLU。空间注意力可以确定网络需要关注输入特征图的空间位置信息,结构如图4c所示。过程与通道注意力类似,分别做2次池化得到2个H×W×1的特征图,并将这2个特征图进行拼接。将拼接结果经过卷积,卷积核大小为7×7,步长为1,并将原来的二维降为一维。最后将得到的H×W×1的特征图经

过 Sigmoid 激活函数。在混合注意力中作用于特征图的两种池化,可以对信息进行整合。平均池化可

以更多地保留图像的背景信息,最大池化可以更多地保留纹理信息。

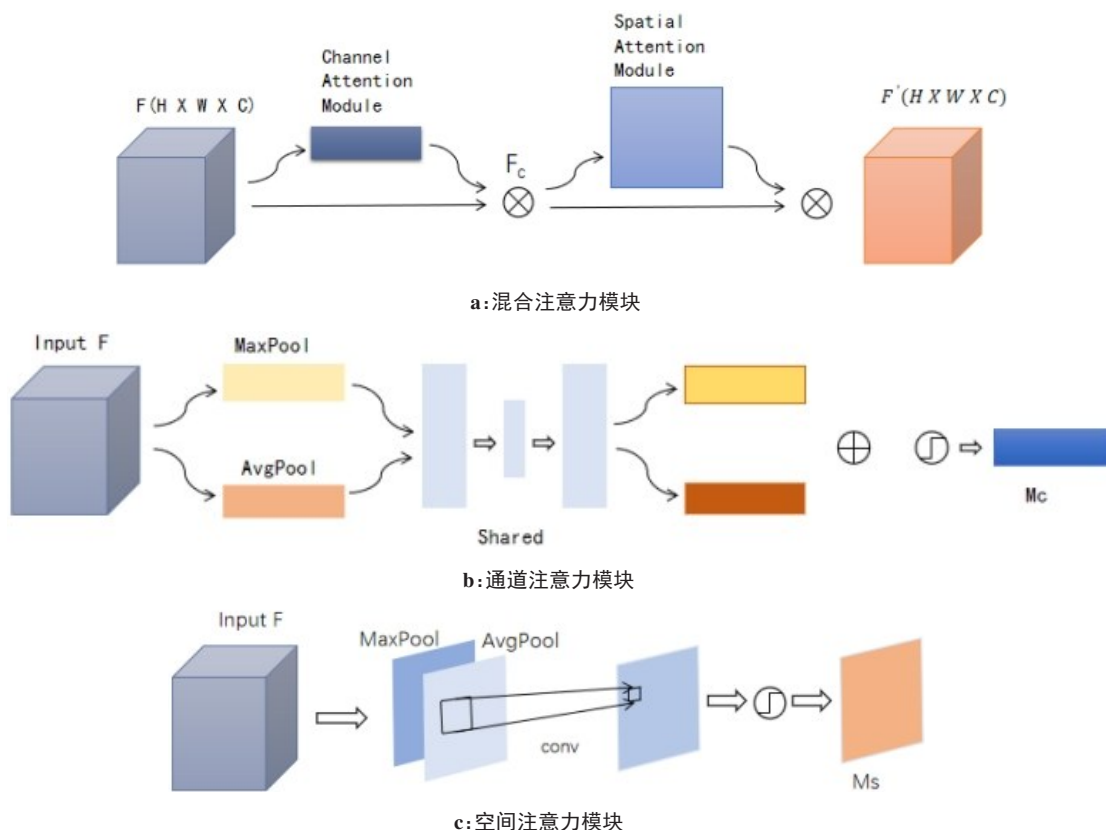


图4 混合注意力模块

Figure 4 Hybrid attention module

判别器可以使得生成的图像更加接近真实图像。在原生的 GAN 网络^[15]中有一些常见的问题,在训练中生成器与判别器需要进行对抗训练,二者在训练过程中学习率设置不当使得训练不平衡,判别器可能会训练得过好或者过差导致模式坍塌。因此我们借鉴了 Pix2pix 中提出的 PatchGAN^[16]。PatchGAN 是一个全卷积网络,特征图随着卷积的过程感受野不断扩大(图 5),最终输出 7×7 的特征图。特征图中的每个点对应原图的 1 个小块,因此 PatchGAN 可以更加细化地判断生成图像是否符合要求。在网络中包含 6 个卷积层,前 4 个卷积核为 4×4,步长为 2,Padding 为 1,后 2 个卷积层卷积核为 1×1,步长为 1,每个卷积层在参数更新时使用谱归一化^[17-18],使网络更快收敛。前 5 个卷积层后面均接了参数为 0.2 的 Leaky ReLU 激活函数。

1.3 损失函数

由于判别器只能判定生成的图像是否为此模态,而不能对所产生的图像与原始图像的相似性做约束,所以需要为生成器提供更多的信息。根据

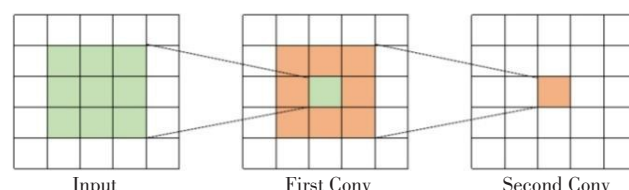


图5 感受野示意图

Figure 5 Illustrative diagram of the receptive field

Pix2pix 所述,常见的损失函数有均方损失 (Mean Square Error, MSE, 也称 L_2 损失) 和平均绝对损失 (Mean Absolute Error, MAE, 也称 L_1 损失)。在这里使用对抗损失和 L_1 损失约束生成器,其中对抗损失约束生成图像高频部分, L_1 损失约束低频部分。

对抗损失使用 Wasserstein 距离,根据 WGAN 所述,计算公式如式下:

$$L = E_{x \sim p_r} [f_w(x)] - E_{x \sim p_g} [f_w(x)] \quad (4)$$

其中, $E[\cdot]$ 代表期望, f_w 代表参数为 w 的判别器网络, $x \sim p_r$ 表示真实图像, $x \sim p_g$ 表示生成图像, L 为真实图像与生成图像之间的 Wasserstein 距离。其中,生成

器的损失有2个部分如式(5)所示,包含对抗损失和 L_1 损失:

$$L_G = -E_{x \sim p_g}[f_w(x)] + \lambda L_1$$

(5)

判别器损失如式(6)所示:

$$L_D = E_{x \sim p_r}[f_w(x)] - E_{x \sim p_g}[f_w(x)]$$

(6)

2 实验与结果

本研究的所有网络框架基于PyTorch网络框架(CUDA版本为10.2, Python版本为3.10.2),运行环境Anaconda3-2021.11-Linux-x86_64,实验平台为Tesla V100,显存为16 GB。

2.1 实验数据

选择BraTs2018数据集作为实验数据,该数据集共有285个病例,每个病例有完整的 T_1 、 T_{1ce} 、 T_2 、Flair共4个模态,且每个模态的图像大小均为 $240 \times 240 \times 155$ 。按8:2拆分训练集与测试集,并将所有切片做归一化。所有网络的优化过程使用Adam优化器,学习率为0.001,平滑常数分别为0.5和0.99。每次训练批次大小为8,每个轮次随机选取所有训练切片的10%,减少运行时间。生成器与判别器交替训练,频次为1:5。在生成器损失中 L_1 损失与GAN损失的比例为500:1。

2.2 评价指标

为了定量评价合成性能,使用峰值信噪比(Peak Signal-to-Noise Ratio, PSNR)和结构相似性(Structural Similarity Index Measurement, SSIM)作为指标。PSNR是基于图像像素的图像质量评价指标。计算公式如下:

$$MSE = \frac{1}{N} \|y(x) - G(x)\|_2^2$$

(7)

$$PSNR = 10 \log_{10} \frac{\max^2(y(x), G(x))}{MSE}$$

(8)

$$SSIM = \frac{(2\mu_{y(x)}\mu_{G(x)} + c_1)(2\sigma_{y(x)G(x)} + c_2)}{(\mu_{y(x)}^2 + \mu_{G(x)}^2 + c_1)(\sigma_{y(x)}^2 + \sigma_{G(x)}^2 + c_2)}$$

(9)

其中 $y(x)$ 为真实图像, $G(x)$ 为合成图像, N 为除边界外的有效点个数, $\max^2(y(x), G(x))$ 表示 $y(x), G(x)$ 中最大灰度值的平方, μ 表示均值, δ 表示方差, $\sigma_{y(x)G(x)}$ 表示 $y(x)$ 和 $G(x)$ 图像的协方差。期望更高的PSNR和SSIM,因为其代表更高的图像质量。

2.3 实验结果与分析

为了定量分析本文算法模型的表现,本文算法与其他方法进行了比较。这些方法可以总结如下:(1)Pix2pix,该方法通过保持像素强度相似性来合成整个图像。(2)cycleGAN^[19],该方法训练不需要成对数据。本实验中使用成对的数据合成从存在模态到不存在模态的医学图像。由于Pix2pix和

cycleGAN应对所有缺失情况需要训练14种模型,时间成本比较高,实验对提出的模型进行了3个任务的评估。在BraTs2018数据集上,使用 T_{1ce} 合成 T_1 、 T_2 、Flair模态(任务1), T_1 、 T_{1ce} 合成 T_2 、Flair(任务2)和 T_{1ce} 、 T_1 、 T_2 合成Flair(任务3)。测试数据集的多项指标,见表1~表6。2个指标(PSNR和SSIM)除了任务2中PSNR略低于Pix2pix,其他均优于比较方法,这表明本文模型可以有效融合不同模态的互补信息,有利于合成效果。

表1 任务1的PSNR对比

Table 1 Comparison of PSNR for task 1

算法模型	T_1	T_2	Flair
Pix2pix	24.09	20.09	18.39
cycleGAN	23.15	20.48	23.05
本文算法	24.72	25.20	23.47

表2 任务1的SSIM对比

Table 2 Comparison of SSIM for task 1

算法模型	T_1	T_2	Flair
Pix2pix	0.994 5	0.987 1	0.995 1
cycleGAN	0.974 6	0.971 4	0.985 3
本文算法	0.994 8	0.987 7	0.995 7

表3 任务2的PSNR对比

Table 3 Comparison of PSNR for task 2

算法模型	T_2	Flair
Pix2pix	24.09	23.30
cycleGAN	18.41	20.84
本文算法	23.75	25.36

表4 任务2的SSIM对比

Table 4 Comparison of SSIM for task 2

算法模型	T_2	Flair
Pix2pix	0.986 7	0.995 5
cycleGAN	0.961 6	0.985 2
本文算法	0.987 8	0.996 2

表5 任务3的PSNR对比

Table 5 Comparison of PSNR for task 3

算法模型	Flair
Pix2pix	25.32
cycleGAN	18.90
本文算法	29.70

表6 任务3的SSIM对比
Table 6 Comparison of SSIM for task 3

算法模型	Flair
Pix2pix	0.996 0
cycleGAN	0.987 2
本文算法	0.996 6

3 结 语

本研究提出一种基于生成对抗网络的多模态MR缺失模态合成网络,解决了临床上MR图像模态缺失等问题,帮助医生做出更加准确的临床诊断。在数据集的选择上,选用了多模态脑肿瘤分割挑战(BraTs2018)数据集。在网络算法上,本研究利用了各模态的特定属性,同时使用了多个模态之间的相关性;此外,只需要1次训练(即1个网络模型)就可以应对所有缺失情况,与其他方法相比更为灵活;最后对每个存在的模态单独进行特征提取,对提取的特征进行特征融合,并引入CBAM模块对每个通道进行强调。通过在同等条件下,与Pix2pix和cycleGAN算法进行比较,结果显示本文模型具有较好的优越性。

【参考文献】

- [1] Fan J, Cao X, Wang Q, et al. Adversarial learning for mono- or multi-modal registration[J]. Med Image Anal, 2019, 58: 101545.
- [2] Ma C, Luo G, Wang K. Concatenated and connected random forests with multiscale patch driven active contour model for automated brain tumor segmentation of MR images[J]. IEEE Trans Med Imaging, 2018, 37(8): 1943-1954.
- [3] Zhang Z, Yang L, Zheng Y. Translating and segmenting multimodal medical volumes with cycle- and shape-consistency generative adversarial network[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2018: 9242-51.10.1109/cvpr.2018.00963
- [4] Zhou T, Liu M, Fu H, et al. Deep multi-modal latent representation learning for automated dementia diagnosis [M]. Medical Image Computing and Computer Assisted Intervention-MICCAI, 2019: 629-638.
- [5] Zhou T, Thung KH, Zhang Y, et al. Inter-modality dependence induced data recovery for MCI conversion prediction [M]. Medical Image Computing and Computer Assisted Intervention-MICCAI, 2019: 186-195.
- [6] Wang Y, Ma G, An L, et al. Semisupervised triple dictionary learning for standard-dose PET image prediction using low-dose PET and multimodal MRI[J]. IEEE Trans Biomed Eng, 2017, 64(3): 569-579.
- [7] Li R, Zhang W, Suk HI, et al. Deep learning based imaging data completion for improved brain disease diagnosis [J]. Med Image Comput Comput Assist Interv, 2014, 17(Pt 3): 305-312.
- [8] Yu BZ, Wang L. 3D cGAN based cross-modality MR image synthesis for brain tumor segmentation [C]. 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI2018), 2018.
- [9] Zhou Q, Zou H. A layer-wise fusion network incorporating self-supervised learning for multimodal MR image synthesis[J]. Front Genet, 2022, 13: 937042.
- [10] Woo S, Park J, Lee JY, et al. CBAM: convolutional block attention module [M]. Computer Vision-ECCV, 2018: 3-19.
- [11] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation [M]. Medical Image Computing and Computer-Assisted Intervention-MICCAI, 2015: 234-241.
- [12] Haut JM, Fernandez-Beltran R, Paoletti ME, et al. Remote sensing image superresolution using deep residual channel attention [J]. IEEE Trans Geosci Remote Sens, 2019, 57(11): 9277-9289.
- [13] Du W, Wang Y, Qiao Y. Recurrent spatial-temporal attention network for action recognition in videos [J]. IEEE Trans Image Process, 2018, 27(3): 1347-1360.
- [14] Yan H, Jiang Y, Zheng J, et al. A multilayer perceptron-based medical decision support system for heart disease diagnosis [J]. Expert Syst Appl, 2006, 30(2): 272-281.
- [15] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks [J]. Commun ACM, 2020, 63(11): 139-144.
- [16] Isola P, Zhu JY, Zhou T, et al. Image-to-image translation with conditional adversarial networks [C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017. DOI: 10.1109/CVPR.2017.632
- [17] Gulrajani I, Ahmed F, Arjovsky M, et al. Improved training of wasserstein gans [C]//Advances in Neural Information Processing Systems, 2017: 5767-5777.
- [18] Arjovsky M, Bottou L. Towards principled methods for training generative adversarial networks [EB/OL]. <https://arxiv.org/pdf/1701.04862.pdf>.
- [19] Zhu JY, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks [C]. 2017 IEEE International Conference on Computer Vision. DOI: 10.1109/iccv.2017.244.

(编辑:黄开颜)