

中图法分类号: TP391.9 文献标识码: A 文章编号: 1006-8961(2025)07-2558-12

论文引用格式: Li T H, He Z Q, He L, Zhang W and Huang Z F. 2025. Fusion of image enhancement and self-attention mechanism for the generation of active pulmonary tuberculosis CT image data. Journal of Image and Graphics, 30(7):2558-2569(李天虎, 何志琴, 何力, 张蔚, 黄忠锋. 2025. 融合图像增强和自注意力机制的活动性肺结核CT影像数据生成. 中国图象图形学报, 30(7):2558-2569)[DOI:10.11834/jig.240518]

融合图像增强和自注意力机制的活动性 肺结核CT影像数据生成

李天虎¹, 何志琴^{1*}, 何力², 张蔚³, 黄忠锋³

1. 贵州大学电气工程学院, 贵阳 550025; 2. 广州中医药大学第四临床医学院, 广州 511400;
3. 贵阳市公共卫生救治中心, 贵阳 550000

摘要: **目的** CT(computed tomography)影像数据是基于深度学习模型辅助医生进行活动性肺结核预防和诊断的重要基础,然而,隐私保护和获取成本等因素制约了大规模数据集的制作。而且,现有数据生成技术在处理低质量和来源多样化数据时存在局限,难以有效提升生成影像的质量,导致细微病灶难以有效生成,且生成影像清晰度不佳。因此,提出一种融合图像增强和自注意力机制的数据生成方法,提升活动性肺结核影像数据的生成质量。**方法** 首先基于循环一致性对抗网络CycleGAN(cycle-consistent adversarial network)对低质量的源数据进行图像增强处理,然后再基于图像到图像的条件生成对抗网络Pix2Pix(image-to-image translation with conditional adversarial network)进行数据生成。在生成器设计上,采用U-Net架构,并融入含自注意力机制的残差模块,增强模型对图像中不同上、下文信息与细节特征的学习能力。同时,对于判别器,采用PatchGAN(patch-based generative adversarial network)架构,聚焦于分析图像的局部特征,并采用PReLU(parametric rectified linear unit)激活函数,其可自适应调整负斜率的特性有助于提升判别器的图像判别性能。**结果** 在自建和公开的混合数据集上与主流方法进行比较的实验结果显示,在未经过图像增强的数据生成实验中,本文方法的数据生成效果与性能第2的方法相比较,峰值信噪比(peak signal-to-noise ratio, PSNR)和结构相似性(structural similarity index measure, SSIM)分别提高1.38%和2.15%, FID(Fr chet inception distance)降低10.70%。在经过图像增强后的数据生成实验中,PSNR和SSIM分别提高0.52%和1.05%, FID降低2.01%。**结论** 本文提出的数据生成方法,通过融合图像增强技术和自注意力机制有效提升了生成数据的整体质量。

关键词: 数据生成;图像增强;活动性肺结核;自注意力机制;CT影像;CycleGAN;Pix2Pix

Fusion of image enhancement and self-attention mechanism for the generation of active pulmonary tuberculosis CT image data

Li Tianhu¹, He Zhiqin^{1*}, He Li², Zhang Wei³, Huang Zhongfeng³

1. School of Electrical Engineering, Guizhou University, Guiyang 550025, China; 2. The Fourth Clinical College, Guangzhou University of Chinese Medicine, Guangzhou 511400, China; 3. Guiyang Public Health Treatment Center, Guiyang 550000, China

Abstract: Objective Computed tomography (CT) image data provide an important basis for deep learning models to assist doctors in the prevention and diagnosis of active pulmonary tuberculosis. The accurate generation and processing of CT

收稿日期:2024-09-18;修回日期:2024-12-17;预印本日期:2024-12-24

* 通信作者:何志琴 zqhe@gzu.edu.cn

基金项目:贵州省科学技术基金项目(黔科合支撑[2023]一般096)

Supported by:Guizhou Science and Technology Foundation(Qian[2023]096)

image data are crucial for the diagnosis level of doctors. Compared with traditional diagnostic methods, deep learning models can automatically identify and process large-scale image data, improving the efficiency and accuracy of diagnosis. However, forming large-scale and high-quality datasets still faces many challenges due to privacy protection regulations and high acquisition costs. These factors limit the amount of data required to train a deep learning model, affecting the performance of the model. Existing data generation techniques exhibit limitations when dealing with data from low-quality and diverse sources. In particular, these techniques experience difficulty in achieving ideal results in the quality of generated images, especially in generating subtle lesions and ensuring image clarity. This limitation not only affects the ability of the model to identify detailed lesions, but also limits the ability of doctors to use the generated images for accurate diagnosis. Therefore, this study proposes a data generation method that combines image enhancement and self-attention mechanism to improve the generation quality of active pulmonary tuberculosis image data. This method not only performs well in improving image clarity but can also effectively enhance the expression ability of the generated data for subtle lesions.

Method First, we use a cycle-consistent adversarial network (CycleGAN) to enhance the original low-quality source data. CycleGAN uses two generators and two discriminators to ensure transformation consistency between the source domain and target domain through the cycle consistency loss function, improving the quality of image data in an unsupervised manner. Then, image-to-image translation with a conditional adversarial network (Pix2Pix) was used for data generation. Pix2Pix gives more attention to image-to-image mapping and makes the generated image as close as possible to the real image through the cooperative work of the generator and the conditional discriminator. This scheme provides special attention to the matching of data samples and their detail processing, enhancing the accuracy and fineness of image data. At the generator design level, this study uses the U-Net architecture and integrates the residual module with a self-attention mechanism. The U-Net architecture enables the model to effectively use the information of each feature layer through a symmetric encoder-decoder path, enhancing the model's ability to learn the features of images with different resolutions. Combined with the self-attention mechanism, the residual module significantly improves the model's ability to perceive and process image context information and detail features. For the discriminator part, this study adopts the PatchGAN architecture, which focuses on analyzing the local features of an image. PatchGAN can finely detect the details of an image by dividing the image into multiple small blocks and discriminating authenticity separately. This study also uses the parametric rectified linear unit activation function, which can adaptively adjust the negative slope, endowing the model with better nonlinear expression ability in the process of backpropagation to improve the image discrimination performance of the discriminator.

Result To verify the effectiveness of the proposed method, experiments are performed on the self-built and public mixed datasets, and the results are compared with those of mainstream methods. In the data generation experiment without image enhancement processing, the data generation effect of the proposed method is 1.38% and 2.15% higher than that of the second performing method in the peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) indicators, respectively, while Fréchet inception distance (FID) is reduced by 10.70%. These results show that the proposed method exhibits significant advantage in terms of image quality. In the data generation experiment after image enhancement processing, the PSNR and SSIM of the proposed method are increased by 0.52% and 1.05%, respectively, while FID is reduced by 2.01% compared with the other methods. These experimental results further prove the effectiveness of the proposed method in improving the quality of image data generation, particularly in the generation of subtle lesions and the improvement of image clarity.

Conclusion In general, the data generation method that combines image enhancement technology and self-attention mechanism proposed in this study effectively improves the overall generation quality of CT image data of active pulmonary tuberculosis. By processing low-quality source data and extracting the fine features of the generated data, the proposed method not only improves the generation ability of subtle lesions, but also significantly enhances the definition of the generated image, overcoming the limitations of the existing generation technology in dealing with complex medical image data.

Key words: data generation; image enhancement; active tuberculosis; self attention mechanism; CT images; CycleGAN; Pix2Pix

0 引言

肺部CT(computed tomography)影像是评估活动性肺结核等疾病的关键诊断工具(Jeong 和 Lee, 2008)。随着深度学习技术的快速发展,越来越多的深度学习模型开始广泛应用于医学影像分析中,以辅助医生进行更准确的诊断和病情评估(Oulefki 等, 2021)。然而,深度学习模型的训练需要依赖大量的影像数据(Adnan 等, 2022),在实际应用中,由于患者隐私保护和数据采集成本的限制,想要获取大规模的数据存在困难。因此,如何进行有效的数据增强,对于深度模型训练至关重要(Kim 等, 2022)。

目前,数据增强的方法包括传统方法和深度学习方法。传统方法包括几何变换、颜色变换和像素变换等。几何变换包括图像的翻转(水平翻转和垂直翻转)、旋转、平移、缩放以及裁剪等操作,通过这些方法改变图像中像素值的位置增加数据集的多样性(Kalaivani 等, 2023)。颜色变换通过对图像的亮度、对比度、饱和度和色调进行调整实现数据增强。像素变换通过添加随机噪声,用来模拟采集过程中可能存在的情况。然而,这些传统的方法应用在医学影像时,存在很大的限制。几何变换并不能改变数据的多样性(Suh 等, 2021),无法从根本上解决数据不平衡的问题。像素变换和颜色变换可能会影响医学影像中的临床特征,从而影响诊断。

近年来,随着深度学习的发展,特别是生成对抗网络(generative adversarial network, GAN)(Goodfellow 等, 2020)及其变体凭借其生成质量高、效果逼真的优势,已成为数据增强的主流方法。GAN的生成方法主要分为两大类:基于GAN、DCGAN(deep convolutional generative adversarial network)(Radford 等, 2015)、WGAN(Wasserstein generative adversarial network)(Arjovsky 等, 2017)、WGAN-GP(Gulrajani 等, 2017)等模型的无条件生成方法和基于cGAN(conditional generative adversarial network)(Mirza 和 Osindero, 2014)、Pix2Pix(image-to-image translation with conditional adversarial network)(Isola 等, 2017)、CycleGAN(Zhu 等, 2017)等模型的有条件生成方法。自GAN问世以来,在数据生成方面取得了巨大突破。GAN通过输入噪声,并通过生成器和判别器的

对抗训练,使生成的图像数据满足真实图像分布。然而,在训练过程中,可能会出现梯度消失的问题,因此出现了基于GAN改进的DCGAN、WGAN和WGAN-GP等模型。这些无条件的生成方法无法确定生成图像的方向,生成的图像是随机的。为了控制生成图像的方向,又有研究学者在已有的基础上提出有条件的数据生成方法。例如条件GAN(CGAN)通过将标签信息作为条件输入到网络中,使得网络能够在给定条件的约束下生成数据。然而,这些标签信息所包含的信息有限,无法完全描述图像的所有视觉特征和细节。此外,由于不同的主观意识对标签信息的理解不同,可能导致生成的图像存在差异。因此,以直观的图像作为输入条件可以解决这些问题,如Pix2Pix、CycleGAN等图像到图像的转换生成网络。对于医学影像,可以将病灶的分割掩膜(樊圣澜 等, 2023)作为输入条件,从而生成特定的病灶数据,解决各类病灶数据不平衡的问题。相比于类别标签或文本标签,这种方式提供了更为丰富和直观的视觉上下文信息,有助于生成更准确和真实的图像。但是,在临床实践中,为了平衡扫描时间和放射剂量,肺部常规使用低剂量CT扫描,因此很难获得高分辨率图像,而上述方法基于这些低质量的源数据生成图像,尽管解决了数据不足、不平衡的问题,但生成的图像质量仍然很低,导致生成的影像仍然存在细微病变的细节特征表达不明显的问题,这对于存在多种细微病灶类型的活动性肺结核CT影像是致命的,进而影响深度学习模型对病灶细节特征的分析 and 诊断性能。另外,用于训练的数据来源、采集参数可能不同,从而使生成的图像对比度等信息不统一,影响生成的图像质量。因此,本文在上述方法的基础上,提出一种基于GAN和图像增强预处理的数据增强方法,并用于活动性肺结核的CT影像数据增强,与其它方法相比,本文方法主要有以下贡献:1)为了解决数据来源不统一、质量的问题和提高生成图像的质量,本文基于CycleGAN模型对数据进行了图像增强预处理,为后续生成任务提供了更好的数据基础;2)基于Pix2Pix模型进行数据生成,采用掩膜到肺野再到完整CT的二阶段生成策略,提高对肺质中细节特征的生成能力;3)在模型的生成器中引入融合自注意力机制的残差模块,提高模型上、下采样时的特征捕捉能力。同时采用PatchGAN的判别器结构关注图像的局部区域,使用

负斜率可学习自适应的PReLU激活函数,从而提高模型的性能。4)通过定性和定量实验表明,本文方法对活动性肺结核细微病灶的生成效果更接近真实图像,同时整体的生成图像质量更好。

1 相关工作

1.1 图像增强

基于GAN的图像增强方法是当前图像增强的主流方法(姚韩敏等,2024),可分为超分辨率和域转换两类。Zhao等人(2022)基于DnCNN(denoising convolutional neural network)和SRGAN(super-resolution generative adversarial network)架构开发了一种新的Dn-SRGAN模型,通过引入去噪模块对CMR(cardio-vascular magnetic resonance)图像进行去噪。该方法不仅可以对原始低分辨率和有噪声的CMR图像进行超分辨率处理,还可以对原始高分辨率CMR图像进行超分辨率重建,解决了CMR图像中的伪影和噪声对GAN性能的影响。Wang等人(2023a)提出的一种多阶段混合注意网络的高效模型,通过能够增强空间信息差异的放大空间注意块、具有数据一致性层的自注意块以及同时关注空间和通道信息的自适应局部剩余注意块,执行从低分辨率(low-resolution, LR)欠采样测量中恢复高分辨率(high-resolution, HR)MR图像的多任务。

而基于模态转换的图像增强方法的研究中, Deng等人(2023)通过在CycleGAN的生成器中添加包含多样性分支模块的辅助链,以获得低分辨率的补充语义信息,结合CBCT(cone beam computed tomography)和CT各自的优点,利用改进的CycleGAN生成更高质量的CT。Wang等人(2023b)提出一种结构约束的CycleGAN,用于脑MR-CT合成,解决MR图像不能直接提供电子密度信息的局限性,而能够提供电子密度信息的CT既耗时又昂贵,且对患者可能有害的缺点。蒋泽宇等人(2023)利用CycleGAN实现了1.5 mm和5 mm两个域的肺部CT影像转换,以厚层肺部CT为源域,生成了更高质量的薄层肺部CT,验证了跨不同厚度肺部CT图像转换的可行性。

对于活动性肺结核的CT影像,从低剂量的CT影像中得到一定量的高分辨率图像是困难的,但可以相对容易地获得薄层的CT。另外,为了避免数据

来源不同导致可能存在的数据对比度不一致的问题,本文采用基于CycleGAN的域转换方法,将低剂量获得的CT域转换到高剂量获得的CT域,实现数据增强之前的图像增强预处理。

1.2 数据生成

近年来,越来越多基于GAN的数据增强算法在医学领域中展开应用,例如, Frid-Adar等人(2018)提出的使用生成对抗网络生成肝脏病变数据增强方法,扩大肝脏病变数据的规模和多样性,从而提高了肝脏病变的分类精度。Sandhiya等人(2024)提出的利用DCGAN对脑肿瘤MRI(magnetic resonance imaging)数据进行预处理的方法,通过生成更多的病变数据,从而提高对脑肿瘤进行分割和分类。Mendes等人(2023)为了解决与肺部相关的、高质量医疗数据集获取困难的问题,使用两个生成式对抗网络,从相应的位置和语义注释合成人工肺图像,包括Pix2Pix方法,从肺分割图生成肺图像;以及在生成过程中使用附加语义标签实现的条件生成对抗网络方法。Jiang等人(2021)针对COVID-19获取困难、缺乏标记的数据的问题,提出一种基于条件生成对抗网络的CT图像合成方法,该方法有效地生成高质量、逼真的COVID-19CT图像,并且有望用于包分割和分类的深度学习应用。

对于包含多种病灶的活动性肺结核CT影像数据,基于图像条件输入的图像生成方法能够生成更高质量的图像数据。因此,本文基于上述方法,利用病灶的分割掩膜数据作为输入条件,并基于Pix2Pix生成有效的CT数据。

2 网络结构

本文将整个流程分为图像增强和数据生成两个部分。如图1所示,在图像增强阶段,本文基于CycleGAN网络对数据进行预处理。以5 mm厚度层且图像对比度、亮度和来源不统一的肺部CT图像作为源域数据,以0.625 mm厚度层且图像属性相同的肺部CT图像作为目标域数据。通过训练生成器学习目标域的特征分布,使生成器生成类似目标域的特征分布,同时,利用生成器与判别器之间的对抗训练,使生成器生成高质量的目标域图像。如图2所示,在数据增强阶段,使用Pix2Pix网络完成数据的生成。为了使生成器更专注于肺野中的特

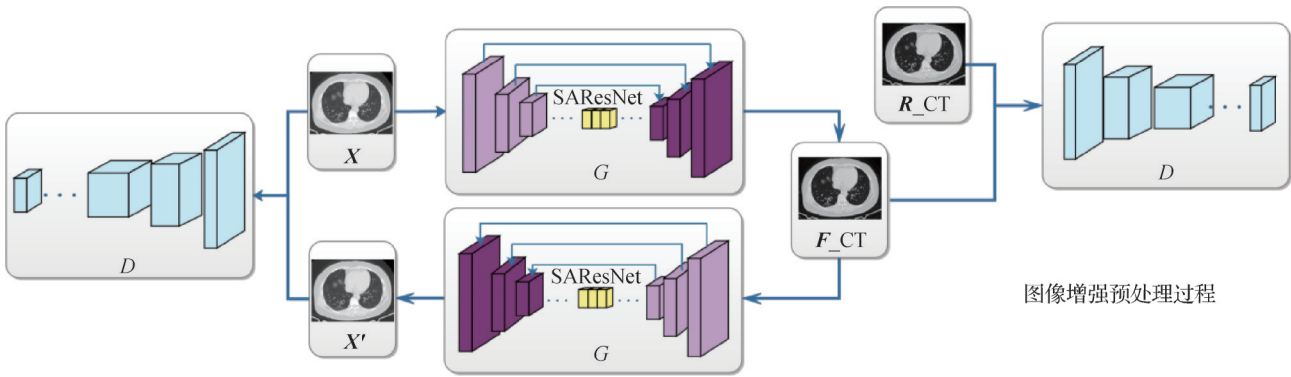


图1 图像增强网络结构
Fig. 1 Image enhancement network structure

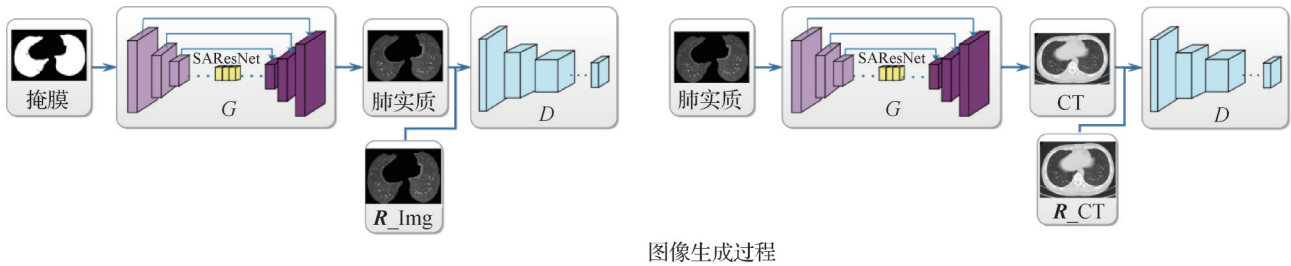


图2 数据生成网络结构
Fig. 2 Data generation network structure

征,本文通过两个阶段实现数据的增强。首先以肺野数据为第1阶段的输入,通过训练生成器生成只包含肺质的数据;再将第1阶段生成的只包含肺质的数据作为第2阶段的输入,最终生成完整的CT图像。

2.1 生成器

生成器通过学习不同域之间的映射关系实现域转换,本文利用U-Net(Ronneberger等,2015)的编码器架构并结合残差模块进行搭建,利用U-Net具有多尺度融合方式进行跨层连接和ResNet(residual network)(He等,2016)模块增强信息保留、提高特征表示能力的特性,提升了生成器生成图像的质量。编码器使用3个卷积层,每个卷积层之后使用实例归一化(instance normalization,IN)和激活函数ReLU(rectified linear unit)进行处理。解码器由3个反卷积层构成,并对特征向量进行重建,每个反卷积层后都使用了IN归一化和ReLU激活函数处理,最后通过Tanh函数处理后输出生成的图像。解码器由3个反卷积层构成,并对特征向量进行重建,每个反卷积层后同样使用IN归一化和ReLU激活函数处理,最后通过Tanh函数处理后输出生成的图像。

2.2 融合自注意力机制的残差模块

为了更加突出肺部CT中的细微病灶,在传统残差模块的基础上,本文引入自注意力机制(Wang等,2023c),称为SAResNet(self-attention residual network),如图3所示,它接收到经过编码器得到的特征向量后,分别经过3个输入通道,分别为直接输出、经过卷积之后输出和经过自注意力机制处理后的输出。其中,自注意力机制处理先通过3个 1×1 的卷积层进行线性变换,得到3个新的特征矩阵 $Q、K、V$,然后计算 Q 与 K 的点乘得到能量矩阵,该矩阵经过softmax激活后得到注意力权重矩阵,再将自注意力权重矩阵和 V 进行矩阵乘法操作,得到自注意力机制处理的结果。最后,将分别经过3个通道处理后的特征进行融合,得到最终的结果,并输入到解码器中。权重计算为

$$A_i = (Q, K, V) = f_{\text{softmax}}\left(\frac{QK^T}{\sqrt{d_k}}\right) \times V \tag{1}$$

式中, $Q、K、V$ 表示通过线性变换得到的3个新的特征矩阵, d_k 表示键向量维度,缩放因子 $\sqrt{d_k}$ 用于缓解梯度爆炸问题。

2.3 判别器

判别器是对生成器生成的图像和真实图像进行

区分和评估。本文采用 PatchGAN (Chen 等, 2023) 架构构建判别器, 相对于传统的 GAN 模型的判别器, 利用 PatchGAN 判别器能更好地捕捉图像的细节信息, 利用卷积层提取图像特征, 并映射为矩阵, 通过求取每个特征矩阵的真实样本概率的均值作为判别器的最终输出。输入图像首先经过 4 个卷积层, 用于提取输入图像特征, 在最后一个卷积层之后, 利用通道过滤器使输出变为一个 1 通道的张量, 即得到输入图像的局部判别结果。在判别器中采用 IN 进行归一化处理。对于激活函数, 传统的 PatchGAN 使用负斜率固定的 LeakyReLU 激活函数, 而本文使用负斜率可学习自适应的 PReLU 激活函数, 从而提高模型的性能。

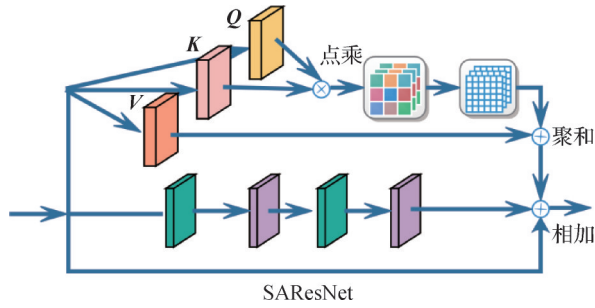


图3 融合自注意力机制的残差模块

Fig. 3 Residual module of integrated self-attention mechanism

2.4 损失函数

2.4.1 CycleGAN 的损失函数

1) 对抗损失。对抗损失是使生成器生成高质量图像的关键。对于生成器, 对抗损失是使生成器生成判别器无法区分的“假”图像。对于判别器, 对抗损失是使判别器始终能够区分生成的“假”图像。在 CycleGAN 中, 有两个生成器和两个判别器, 生成器 G_{xy} 的目标是学习 $X \rightarrow Y$ 的映射 $G(x)$, 将源域 X 空间中的样本转化成目标域 Y 空间中的样本, 并使判别器 D_y 无法区分。生成器 G_{yx} 则学习 $Y \rightarrow X$ 的映射 $F(y)$, 将目标域 Y 空间中的样本转化成源域 X 空间中的样本, 并使判别器 D_x 无法区分。根据交叉熵损失, 可以构造下面的损失函数。

对于生成器 G_{xy} 和判别器 D_y , 利用交叉熵损失构建的对抗损失为

$$L_{ad_G_{xy}, D_y}(G_{xy}, D_y, X, Y) = E_{y \sim p_{data(y)}} [\ln D_y(y)] + E_{x \sim p_{data(x)}} [\ln (1 - D_y(G_{xy}(x)))] \quad (2)$$

式中, $E_{y \sim p_{data(y)}}$ 和 $E_{x \sim p_{data(x)}}$ 分别表示满足 Y 空间中的样本

分布的随机变量 y 和满足 X 空间中的样本分布的随机变量 x 的期望值。 $D_y(y)$ 表示对 y 进行判别。 $D_y(G_{xy}(x))$ 表示对生成的“假”图像进行判别。

对于生成器 G_{yx} 和判别器 D_x , 对抗损失与式(2)相似, 具体为

$$L_{ad_G_{yx}, D_x}(G_{yx}, D_x, Y, X) = E_{x \sim p_{data(x)}} [\ln D_x(x)] + E_{y \sim p_{data(y)}} [\ln (1 - D_x(G_{yx}(y)))] \quad (3)$$

2) 循环一致性损失。循环一致性损失的目标在于确保使用两个映射函数进行转换后的样本能够返回到初始域, 并且与原始输入的差异最小, 从而保证输入和输出之间的一致性。即对于源域 X 中的样本图像, 要使经过转换后的域之间的差异最小, 同理, 对于 Y 域中的样本图像, 也应保证将经过转换后的域之间的差异最小, 定义循环一致性损失函数, 具体为

$$L_{cyc_loss}(G_{xy}, G_{yx}) = E_{x \sim p_{data(x)}} [\|G_{yx}(G_{xy}(x)) - x\|_1] + E_{y \sim p_{data(y)}} [\|G_{xy}(G_{yx}(y)) - y\|_1] \quad (4)$$

式中, $\|G_{yx}(G_{xy}(x)) - x\|_1$ 和 $\|G_{xy}(G_{yx}(y)) - y\|_1$ 分别表示利用 L1 范数计算 x' 与 x , y' 与 y 之间的差值。通过循环一致性损失的约束, 生成器能够更好地进行跨域图像转换, 并生成更加准确和真实的转换结果。

3) 身份损失。为了将 X 域的图像输入生成器 G_{yx} , Y 域的图像输入生成器 G_{xy} 后, 得到的仍尽可能还是它本身, 而不做转换, 构建身份损失函数, 具体为

$$L_{id}(G_{xy}, G_{yx}) = E_{x \sim p_{data(x)}} [\|G_{yx}(x) - x\|_1] + E_{y \sim p_{data(y)}} [\|G_{xy}(y) - y\|_1] \quad (5)$$

2.4.2 Pix2Pix 的损失函数

1) 对抗损失。在 Pix2Pix 中, 只有一个生成器和一个判别器, 与 CycleGAN 同理, 生成器 G 的目标是学习源域 X 和目标域 Y 之间的特征分布, 生成使判别器 D 无法区分的“假”图像。因此, 可构建对抗损失为

$$L_{CGan}(G, D) = E_{x, y} [\lg D(x, y)] + E_{x, z} [\lg (1 - D(x, G(x, z)))] \quad (6)$$

式中, x, z, y 分别表示输入图像、随机噪声和目标图像。 $G(x, z)$ 是生成器 G 基于输入图像 x 作为输入条件和随机噪声 z 生成的图像, $D(x, y)$ 和 $D(x, G(x, z))$ 分别表示输入的图像接近真实图像的概率和与输入的图像接近生成图像的概率。

2)像素损失。除了对抗损失,还引入了L1损失使生成图像在像素级别上接近真实图像,这有助于提高生成图像的质量。L1损失计算的是生成图像与目标图像之间像素差异的绝对值,具体为

$$L_{L1} = E_{x,y,z} [\|y - G(x,z)\|_1] \tag{7}$$

式中, $\|y - G(x,z)\|_1$ 表示基于输入条件和随机噪声生成的图像与真实图像之间的L1范数,即它们的逐像素差的绝对值。

3 实验与分析

3.1 数据集

本文实验的数据集由项目合作医院提供,包含3 200幅活动性肺结核CT影像,以及对应的病灶掩膜。由于难以获得公开的活动性肺结核数据,考虑到新冠病毒在病灶类型上与活动性肺结核相似,因此,为了验证本文方法的鲁棒性和泛化能力,同时采用公开的COVID-19数据集进行训练。公开的COVID-19(Maftouni等,2021)数据集包含2 729幅CT图像及其对应的病变掩膜。

3.2 实验环境配置

实验的运行环境为64位Windows操作系统,

PyTorch为1.12.0的框架;硬件环境为NVIDIA GeForce RTX4090(24 GB),12th Gen Intel(R) Core(TM) i9-12900K 3.20 GHz。

实验的相关参数设定如下,训练200个epoch,采用Adam优化器,其中参数设定为:一阶矩估计的指数衰减率 $\beta_1=0.9$,二阶矩估计的指数衰减率 $\beta_2=0.999$,初始学习率 $\alpha=0.002$,前100个epoch学习率不变,后100个epoch采用线性衰减的更新策略递减,直至为0。

3.3 定性比较

针对数据中存在的不同来源、采集时不同参数的设置,从而导致的图像数据对比度、亮度不统一,以及低剂量的采集方式的导致的图像不清晰的问题,本文基于CycleGAN模型将数据集中的数据进行增强和统一,并对比了几种主流的方法。

如图4所示,利用MUNIT(multimodal unsupervised image-to-image translation)(Huang等,2018)模型得到的结果出现了内容失真和颜色失真。传统的CycleGAN模型虽然没有出现内容失真,但同样出现颜色失真的问题。而DC-CycleGAN和DCC-GAN生成的效果相比前两种方法更好,没有出现内容失真和颜色失真的情况,DC-CycleGAN在对比度上表现良好,但肺质中的细节特征不清晰。在DCC-GAN的

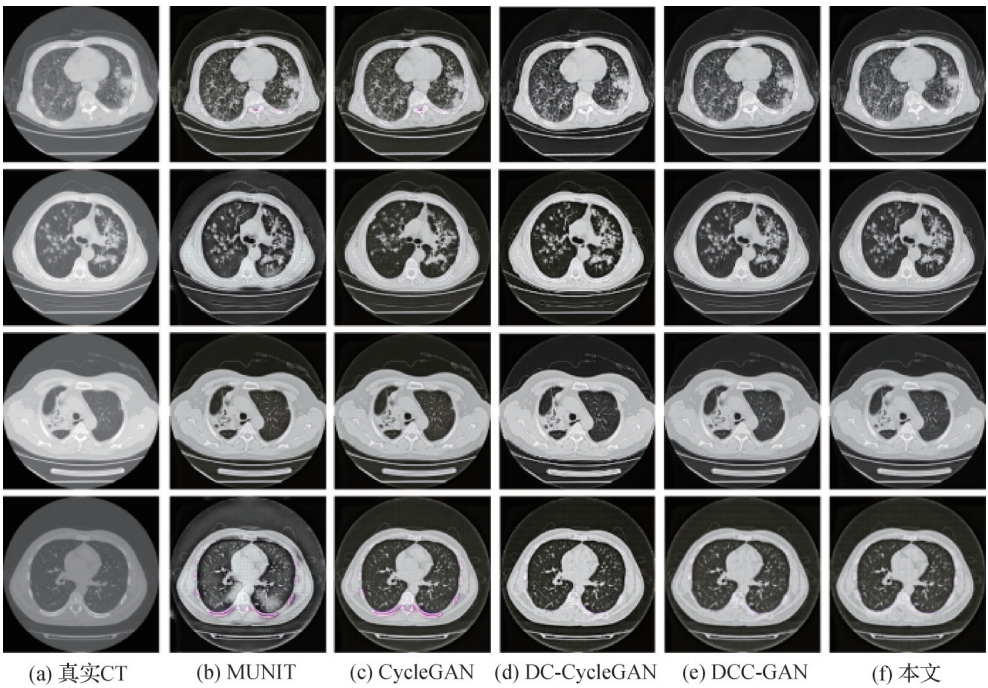


图4 图像增强处理对比结果图

Fig. 4 Comparison results of image enhancement processing
((a) real CT; (b) MUNIT; (c) CycleGAN; (d) DC-CycleGAN; (e) DCC-GAN; (f) ours)

结果中,虽然对比度没有很出色,但肺质中的细节特征表达效果更好。而本文方法相对于其他方法,在生成转换效果整体上表现更出色。

在进行图像增强预处理后,基于得到的处理数据,进行活动性肺结核CT影像的生成。本文采用两阶段的生成方法,实现了从掩膜到肺野,再从肺野到CT的生成过程,通过这种生成方法,提升生成的质量。如图5所示,利用本文方法生成肺野,并对比了Mendes 等人(2023)从肺位置注释生成肺图像的MTI-GAN生成方法、Jiang等人(2021)通过融合全局生成器和多分辨率鉴别的GL-MR-GAN生成方法。从图中可以看到,利用MTI-GAN的生成结果中,对于生成的活动性肺结节和实变并伴随的小空洞病灶

生成效果不佳,并且会生成多余的空洞病灶。相比于MTI-GAN, GL-MR-GAN在生成的小结节和小空洞病灶方面表现更好。但这两种方法与本文方法相比,本文的生成效果更出色,生成的结节和肺实变中的小空洞更接近真实图像。

在完成第1阶段的生成任务之后,本文将第1阶段生成的肺野作为输入,实现完整肺部CT的生成。如图6所示,本文的数据生成方法,不仅在视觉上和病理特征上都更接近于真实的影像,而且经过图像增强处理后的数据再生成,所生成的图像整体质量更高,在清晰度和对比度上表现出色,这意味着本文生成的影像能够更好地反映出肺部结构的细微差别,这对于辅助医生进行诊断和治疗至关重要。

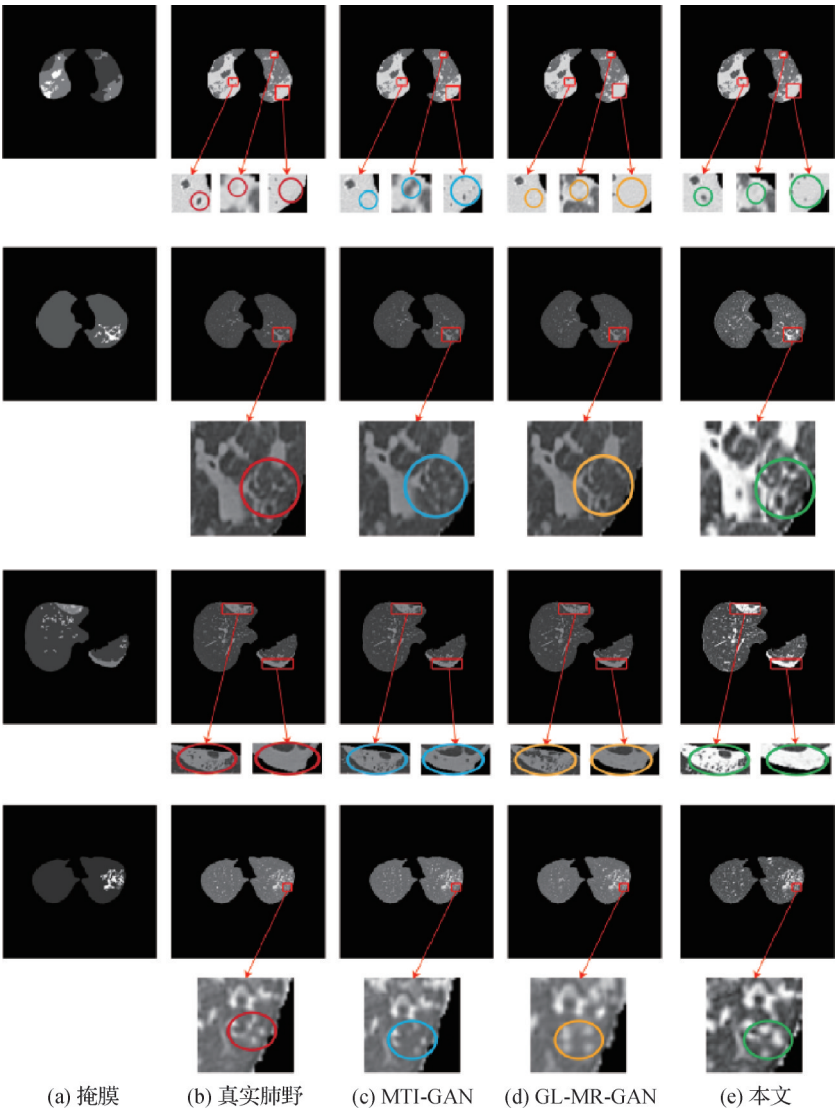


图5 生成的肺野结果对比

Fig. 5 Comparison of lung field results generated((a) mask; (b) real lung field; (c) MTI-GAN; (d) GL-MR-GAN; (e) ours)

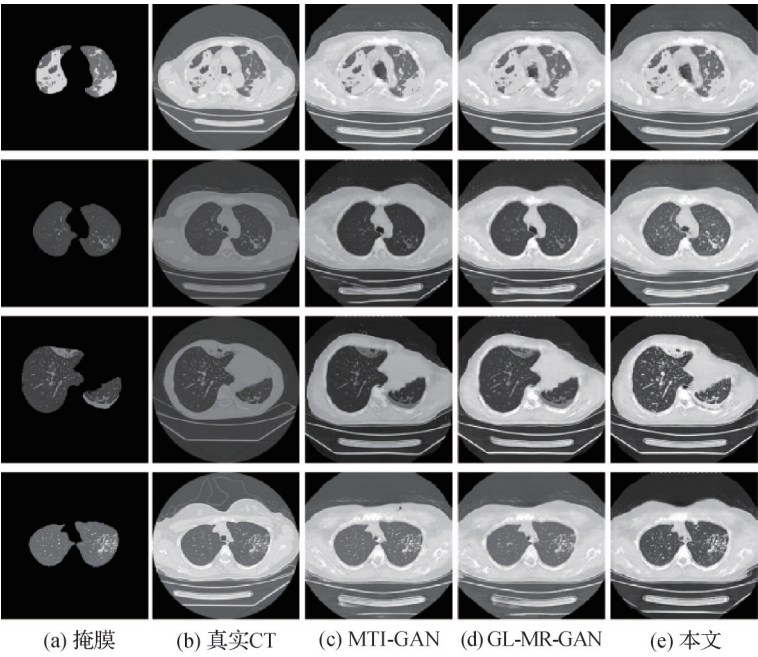


图6 生成的完整CT结果对比图

Fig. 6 Comparison map of complete CT results generated((a) mask;(b) real CT;(c) MTI-GAN;(d) GL-MR-GAN;(e) ours)

3.4 定量比较

为了从客观角度进一步验证实验结果,本文进行了定量实验。对于图像增强的实验结果,采用平均绝对误差(mean absolute error, MAE)、峰值信噪比(peak signal-to-noise ratio, PSNR)和结构相似指数度量(structural similarity index measure, SSIM)3种指标对实验结果进行评估。对于数据增强的实验结果,采用光谱起始距离(Fréchet inception distance, FID)、PSNR 和 SSIM 3 种指标对实验结果进行评估。其中,MAE可以用于衡量生成图像与目标图像之间的差异,将生成图像与目标图像进行比较,计算 MAE 值,以评估生成图像的质量,数值越小,意味着预测值与实际值之间的平均差异越小,模型的性能越好;FID是一种通过计算 Inception 网络中的特征表示的差异评估图像质量的方法,数值越低,表示生成的图像越接近真实图像,模型性能越好;PSNR 通过计算原始图像与处理后图像的最大像素值平方和均方误差来衡量图像质量,数值越高,表示图像的质量越好,失真越小;SSIM 通过考虑图像的亮度、对比度和结构信息,以更全面地反映图像的视觉质量,数值越大,表示生成的图像质量越好。

为了进一步验证本文方法的有效性,本文分别在自建数据集、公开数据集以及自建和公开混合的

数据集上进行实验,并对比了主流的一些方法,包括图像增强预处理和数据增强两个阶段。

表 1 验证了基于 CycleGAN 的图像增强有效性。分别在自建数据集、公开数据集以及自建和公开的混合数据集上进行了实验,并对比了主流的图像增强方法。实验结果表明,无论是在自建数据集、公开数据集或混合数据集上,本文方法的增强效果更好,同时也验证了混合数据的可靠性。

为了能够在同等条件下,分别验证本文所提出的图像增强方法对于生成任务带来的效果和所改进的生成方法性能,本文基于未经过图像增强处理和经过图像增强处理后的数据集,分别进行了数据生成效果对比实验。如表 2 所示,在基于未经过图像增强处理的数据集上,在两个生成阶段上分别验证了生成的效果。通过实验表明,无论是在掩膜生成肺野还是肺野生成完整 CT 的阶段,本文的数据生成方法在 MAE、PSNR 和 SSIM 3 个指标上均有提升,验证了本文数据生成方法在同等条件下的数据生成能力。

如表 3 所示,在经过图像增强处理后的数据上进行实验,本文方法和其他主流方法在 3 个指标上的表现更好,证明了经过处理后的数据对于生成任务的可靠性和有效性。

表 1 不同数据集上的图像增强预处理结果
Table 1 Image enhancement preprocessing results on different datasets

方法	自建数据集			公开数据集			混合数据集		
	MAE	PSNR/dB	SSIM	MAE	PSNR/dB	SSIM	MAE	PSNR/dB	SSIM
UNIT	0.084 31	24.103 51	0.691 87	0.087 94	24.071 28	0.671 37	0.085 41	24.094 31	0.683 20
CycleGAN	0.061 52	24.249 51	0.779 46	0.069 24	24.165 48	0.745 24	0.067 49	24.21917	0.764 96
DC-CycleGAN	0.035 69	24.798 40	0.848 11	0.038 56	24.548 10	0.824 79	0.036 98	24.668 11	0.837 92
DCC-GAN	0.035 01	24.802 49	0.839 84	0.036 73	24.674 81	0.814 61	0.035 25	24.704 56	0.828 51
本文	0.032 93	25.031 26	0.868 99	0.034 10	25.001 48	0.846 29	0.033 07	25.014 85	0.859 49

注:加粗字体表示各列最优结果。

表 2 基于未经过图像增强预处理的混合数据集生成结果
Table 2 Generation results based on the mixed dataset without image enhancement preprocessing

方法	肺野			CT		
	FID	PSNR/dB	SSIM	FID	PSNR/dB	SSIM
MTI-GAN	0.045 77	25.362 66	0.848 33	0.048 81	25.285 50	0.831 55
GL-MR-GAN	0.049 88	24.215 92	0.805 09	0.051 18	24.096 73	0.784 20
本文	0.041 59	25.693 15	0.868 23	0.043 59	25.634 80	0.849 40

注:加粗字体表示各列最优结果。

表 3 基于经过图像增强预处理的混合数据集生成结果
Table 3 Generation results based on the mixed dataset pre-processed with image enhancement

方法	肺野			CT		
	FID	PSNR/dB	SSIM	FID	PSNR/dB	SSIM
MTI-GAN	0.040 22	25.741 87	0.876 43	0.043 20	25.648 11	0.865 51
GL-MR-GAN	0.047 45	24.561 47	0.825 19	0.050 31	24.348 09	0.813 18
本文	0.039 77	25.948 15	0.881 89	0.041 33	25.781 66	0.874 56

注:加粗字体表示各列最优结果。

4 结 论

本文提出一种融合图像增强技术和自注意力机制的数据生成方法。通过利用 CycleGAN 先对有限的源数据进行增强处理后,再利用 Pix2Pix 进行数据生成的方法解决了现有的数据生成方法应用于活动性肺结核 CT 影像数据生成时存在的生成质量低的问题。另外,通过改进图像增强网络和数据生成网络,进一步提高 CT 影像中细微病灶的生成质量。实验结果表明,本文方法能够有效地生成复杂的肺部 CT 影像病灶,在细微病灶的准确生成方面取得了显著提升。在公开数据集上和自建数据集上,在 PSNR、SSIM、FID 和 MAE 等指标上,均优于对比方

法。原因在于,本文针对不同数据源、不同采集质量的源数据进行了增强处理,从而提高了生成数据的质量;另外,通过自注意力机制的嵌入,提升了生成的能力,从而进一步提升了生成性能。

然而,本研究的方法在生成质量上仍有改进空间,尚未实现在保证生成质量的同时,通过单一阶段从掩膜直接生成 CT 影像的目标。针对这些挑战,未来的工作将聚焦于进一步优化网络架构和算法,以提高生成质量并简化生成流程,从而实现更高效的数据生成策略。

参考文献(References)

Adnan M, Kalra S, Cresswell J C, Taylor G W and Tizhoosh H R. 2022.

- Federated learning and differential privacy for medical image analysis. *Scientific Reports*, 12(1): #1953 [DOI: 10.1038/s41598-022-05539-7]
- Arjovsky M, Chintala S and Bottou L. 2017. Wasserstein generative adversarial networks//Proceedings of the 34th International Conference on Machine Learning. Sydney, Australia: JMLR.org: 214-223 [DOI: 10.48550/arXiv.1701.07875]
- Chen G, Zhang G P, Yang Z G and Liu W Y. 2023. Multi-scale patch-GAN with edge detection for image inpainting. *Applied Intelligence*, 53(4): 3917-3932 [DOI: 10.1007/s10489-022-03577-2]
- Deng L W, Ji Y F, Huang S J, Yang X and Wang J. 2023. Synthetic CT generation from CBCT using double-chain-CycleGAN. *Computers in Biology and Medicine*, 161: #106889 [DOI: 10.1016/j.compbiomed.2023.106889]
- Fan S L, Bai Z Y, Lu Q J and Zhou X. 2023. A transformer network based CT image segmentation for COVID-19-derived lung disease. *Journal of Image and Graphics*, 28(10): 3203-3213 (樊圣澜, 柏正尧, 陆倩杰, 周雪. 2023. 基于Transformer网络的COVID-19肺部CT图像分割. *中国图象图形学报*, 28(10): 3203-3213) [DOI: 10.11834/jig.220865]
- Frid-Adar M, Diamant I, Klang E, Amitai M, Goldberger J and Greenspan H. 2018. GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. *Neurocomputing*, 321: 321-331 [DOI: 10.1016/j.neucom.2018.09.013]
- Goodfellow I, Pouget-Abadie J, Mirza M, Xu M, Warde-Farley D, Ozair S, Courville A and Bengio Y. 2020. Generative adversarial networks. *Communications of the ACM*, 63(11): 139-144 [DOI: 10.1145/3422622]
- Gulrajani I, Ahmed F, Arjovsky M, Dumoulin V and Courville A. 2017. Improved training of wasserstein GANs//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, USA: Curran Associates Inc.: 5769-5779 [DOI: 10.48550/arXiv.1704.00028]
- He K M, Zhang X Y, Ren S Q and Sun J. 2016. Deep residual learning for image recognition//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 770-778 [DOI: 10.1109/CVPR.2016.90]
- Huang X, Liu M Y, Belongie S and Kautz J. 2018. Multimodal unsupervised image-to-image translation//Proceedings of the 15th European Conference on Computer Vision. Munich, Germany: Springer: 179-196 [DOI: 10.1007/978-3-030-01219-9_11]
- Isola P, Zhu J Y, Zhou T H and Efros A A. 2017. Image-to-image translation with conditional adversarial networks//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 5967-5976 [DOI: 10.1109/CVPR.2017.632]
- Jeong Y J and Lee K S. 2008. Pulmonary tuberculosis: up-to-date imaging and management. *American Journal of Roentgenology*, 191(3): 834-844 [DOI: 10.2214/AJR.07.3896]
- Jiang Y F, Chen H, Loew M and Ko H. 2021. COVID-19 CT image synthesis with a conditional generative adversarial network. *IEEE Journal of Biomedical and Health Informatics*, 25(2): 441-452 [DOI: 10.1109/JBHI.2020.3042523]
- Jiang Z Y, Han R, Liu X H and Wang G Y. 2023. Research on medical image generation method based on deep learning. *Chinese Medical Equipment Journal*, 44(2): 1-4 (蒋泽宇, 韩荣, 刘晓鸿, 王光宇. 2023. 基于深度学习的医学影像高效生成方法研究. *医疗卫生装备*, 44(2): 1-4) [DOI: 10.19745/j.1003-8868.2023021]
- Kalaivani S, Asha N and Gayathri A. 2023. Geometric transformations-based medical image augmentation//Solanki A and Naved M, eds. *GANs for Data Augmentation in Healthcare*. Cham, Germany: Springer: 133-141 [DOI: 10.1007/978-3-031-43205-7_8]
- Kim H E, Cosa-Linan A, Santhanam N, Jannesari M, Maros M E and Ganslandt T. 2022. Transfer learning for medical image classification: a literature review. *BMC Medical Imaging*, 22(1): #69 [DOI: 10.1186/s12880-022-00793-7]
- Maftouni M, Law A C C, Shen B, Grado Z K, Zhou Y Z and Yazdi N A. 2021. A robust ensemble-deep learning model for COVID-19 diagnosis based on an integrated CT scan images database//Proceedings of 2021 IISE Annual Conference. [s.l.]: Institute of Industrial and Systems Engineers: 632-637 [DOI: 10.1016/j.bspc.2023.105285]
- Mendes J, Pereira T, Silva F, Frade J, Morgado J, Freitas C, Negrão E, de Lima B F, da Silva M C, Madureira A J, Ramos I, Costa J L, Hespanhol V, Cunha A and Oliveira H P. 2023. Lung CT image synthesis using GANs. *Expert Systems with Applications*, 215: #119350 [DOI: 10.1016/j.eswa.2022.119350]
- Mirza M and Osindero S. 2014. Conditional generative adversarial nets [EB/OL]. [2024-09-18]. <https://arxiv.org/pdf/1411.1784.pdf>
- Oulefki A, Agaian S, Trongtirakul T and Laouar A K. 2021. Automatic COVID-19 lung infected region segmentation and measurement using CT-scans images. *Pattern Recognition*, 114: #107747 [DOI: 10.1016/j.patcog.2020.107747]
- Radford A, Metz L and Chintala S. 2015. Unsupervised representation learning with deep convolutional generative adversarial networks [EB/OL]. [2024-09-18]. <https://arxiv.org/pdf/1511.06434.pdf>
- Ronneberger O, Fischer P and Brox T. 2015. U-Net: convolutional networks for biomedical image segmentation//Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention. Munich, Germany: Springer: 234-241 [DOI: 10.1007/978-3-319-24574-4_28]
- Sandhiya B, Raja S K S, Shruthi K and Kamala S P R. 2024. Brain tumour segmentation and classification with reconstructed MRI using DCGAN. *Biomedical Signal Processing and Control*, 92: #106005 [DOI: 10.1016/j.bspc.2024.106005]
- Suh S, Lee H, Lukowicz P and Lee Y O. 2021. CEGAN: classification enhancement generative adversarial networks for unraveling data imbalance problems. *Neural Networks*, 133: 69-86 [DOI: 10.1016/j.neunet.2020.10.004]

- Wang J Y, Wu Q M J and Pourpanah F. 2023a. Dc-cycleGAN: bidirectional CT-to-MR synthesis from unpaired data. *Computerized Medical Imaging and Graphics*, 108: #102249 [DOI: 10.1016/j.compmedimag.2023.102249]
- Wang W L, Shen H X, Chen J C and Xing F S. 2023b. MHAN: multi-stage hybrid attention network for MRI reconstruction and super-resolution. *Computers in Biology and Medicine*, 163: #107181 [DOI: 10.1016/j.combiomed.2023.107181]
- Wang Y X, Xie Y Y, Ji X M, Liu Z and Liu X L. 2023c. RacPixGAN: an enhanced sketch-to-face synthesis GAN based on residual modules, multi-head self-attention mechanisms, and CLIP loss//Proceedings of the 4th International Conference on Electronic Communication and Artificial Intelligence (ICECAI). Guangzhou, China: IEEE: 336-342 [DOI: 10.1109/ICECAI58670.2023.10176715]
- Yao H M, Zhou Y Y, Guo J F, Qin J M, Li X X and Dong S Q. 2024. Esophageal endoscopic image enhancement method without reference samples. *Journal of Image and Graphics*, 29(11): 3487-3500 (姚韩敏, 周颖玥, 郭俊菲, 秦佳敏, 李小霞, 董舒琦. 2024. 无参考样本下的食管内镜图像增强技术. *中国图象图形学报*, 29(11): 3487-3500) [DOI: 10.11834/jig.230865]
- Zhao M, Wei Y and Wong K K L. 2022. A generative adversarial network technique for high-quality super-resolution reconstruction of cardiac magnetic resonance images. *Magnetic Resonance Imaging*, 85: 153-160 [DOI: 10.1016/j.mri.2021.10.033]
- Zhu J Y, Park T, Isola P and Efros A A. 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE: 2242-2251 [DOI: 10.1109/ICCV.2017.244]

作者简介

李天虎,男,硕士研究生,主要研究方向为医学影像分析和深度学习。E-mail: 770185886@qq.com

何志琴,通信作者,女,硕士,教授,主要研究方向为深度学习、机器视觉和电力电子运动控制技术。

E-mail: zqhe@gzu.edu.cn

何力,女,硕士,主治医师,主要研究方向为中医内科学。

E-mail: Lilypower@126.com

张蔚,女,主治医师,主要研究方向为呼吸系统疾病的病因学。E-mail: 1911604825@qq.com

黄忠锋,女,主治医师,主要研究方向为呼吸系统疾病的病因学。E-mail: 1017899870@qq.com