

融合上下文注意力的两段式生成对抗网络的肺结节图像生成与分类

尹智贤^{1,4}, 夏克文¹, 张昭², 贺紫平³

1. 河北工业大学电子信息工程学院, 天津 300401; 2. 天津中医药大学第一附属医院/国家中医针灸临床研究中心, 天津 300193;
3. 长沙理工大学计算机与通信工程学院, 湖南 长沙 410114; 4. 天津中德应用技术大学软件与通信学院, 天津 300350

【摘要】提出一种融合上下文注意力的两段式生成对抗网络用于肺结节生成和分类。上下文注意力采用一种通道增强的多头上下文注意力机制,将通道注意力和多头上下文注意力结合,更好地处理特征图中的复杂语义关系,有效增强了模型的特征提取能力;两段式生成对抗网络框架用于实现肺结节在指定肺部区域的注入,该框架将生成任务分为两个阶段:第一阶段生成肺结节感兴趣区域图像,然后通过泊松融合模块与指定的肺实质进行融合,生成初始样本;第二阶段使用改进的CycleGAN模型对初始样本进行微调。同时,在判别器中引入跨层激励模块和辅助分类器实现对特征通道的再校正以及对肺结节的分类。在LIDC-IDRI数据集上进行实验验证,实验结果表明,所提方法在肺结节生成上的FID、IS和KID评分分别为115.153、2.619±0.095和0.062;在肺结节恶性度分类上准确率为70.23%,灵敏度、F1值和AUC分别为68.66%、68.92%和87.59%,表现出优于ADGAN等基于GAN的分类模型,以及VGG16等基准网络的性能。

【关键词】肺结节生成;上下文注意力;生成对抗网络;肺结节分类;CycleGAN

【中图分类号】R318;TP391

【文献标志码】A

【文章编号】1005-202X(2024)12-1517-15

Synthesis and classification of pulmonary nodules using two-stage-based generative adversarial network incorporating contextual transformer

YIN Zhixian^{1,4}, XIA Kewen¹, ZHANG Zhao², HE Ziping³

1. School of Electronics and Information Engineering, Hebei University of Technology, Tianjin 300401, China; 2. First Teaching Hospital of Tianjin University of Traditional Chinese Medicine/National Clinical Research Center for Chinese Medicine Acupuncture and Moxibustion, Tianjin 300193, China; 3. School of Computer and Communication Engineering, Changsha University of Science and Technology, Changsha 410114, China; 4. School of Software and Communication, Tianjin Sino-German University of Applied Sciences, Tianjin 300350, China

Abstract: A two-stage-based generative adversarial network incorporating contextual transformer is proposed for synthesis and multiclass classification of pulmonary nodules. Contextual transformer adopts a channel-enhanced multi-head contextual transformer mechanism which combines channel attention and multi-head contextual transformer to better deal with the complex semantic relationship in the feature space, thereby effectively enhancing the feature extraction capability of the model. A two-stage-based generative adversarial network framework is used to achieve the injection of pulmonary nodules in the designated lung area, and divide the synthesis task into two stages. In the first stage, pulmonary nodule regions of interest images are generated and then fused with designated lung parenchyma through a Poisson blending module to generate preliminary samples; in the second stage, an improved CycleGAN model is used to fine-tune the preliminary samples. Meanwhile, the skip layer excitation module and auxiliary classifier are introduced into the discriminator for realizing the re-correction of the feature channel and the classification of pulmonary nodules. Experiments on LIDC-IDRI dataset reveal that the proposed method has a FID, IS and KID of 115.153, 2.619±0.095 and 0.062 on pulmonary nodule synthesis, and achieves an accuracy, sensitivity, F1 value and AUC of 70.23%, 68.66%, 68.92% and 87.59% on pulmonary nodule malignancy classification, respectively, outperforming GAN-based classification models such as ADGAN, as well as benchmark networks such as VGG16.

Keywords: pulmonary nodule synthesis; contextual transformer; generative adversarial network; pulmonary nodule classification; CycleGAN

【收稿日期】2024-05-27

【基金项目】国家自然科学基金(42075129);河北省自然科学基金(E2021202179);河北省重点研发项目(21351803D);河北省关键技术与产品研发专项(SJMYF2022Y06)

【作者简介】尹智贤,博士研究生,研究方向:医学图像处理、智能信息处理,E-mail: zhixian.yin@hotmail.com

【通信作者】夏克文,博士,教授,博士研究生导师,研究方向:智能信息处理、无线通信与智能天线,E-mail: kwxia@hebut.edu.cn

前言

肺癌具有侵袭性强、早期发现困难等特点,是目前发病率和死亡率最高的恶性肿瘤之一。2022年美国癌症年度报告预测肺癌是癌症死亡的主要原因^[1]。数据显示原位肺癌、局限性肺癌和晚期肺癌的5年生存率分别为60%、33%和6%。因此,早期诊断和治疗对提高肺癌患者的生存率起着至关重要的作用。早期肺癌多以结节形式存在,因此准确诊断结节的良恶性已成为肺癌防治的重要措施^[2]。病理诊断被广泛认为是鉴别肺结节良恶性的基本依据。在临床实践中,对标本的组织病理学和分子生物学检查(如手术切除和针吸活检)是结节病理诊断的标准;然而,肺结节的病理诊断具有侵袭性且复杂耗时,很难成为一种常规检测手段。而计算机断层扫描(Computed Tomography, CT)作为一种无创检查方法,为肺结节的诊断提供有效的补充。传统的肺结节诊断方式主要是通过放射科医生逐片阅读CT图像对肺结节进行诊断。然而每位患者的扫描结果包含大量的图像,这无疑加重了医生的工作强度,且由于主观经验的影响,不同医师的诊断结果可能会存在一定差异,甚至会出现误诊、漏诊等情况。

如何快速、自动地对肺结节进行诊断一直是众多学者关注的焦点。近年来,随着深度学习的不断发展,出现了一系列基于数据驱动的肺结节诊断方法^[3-5]。这些方法可以从大量的数据中自动学习高质量的鉴别特征,从而更准确、有效地诊断结节症状。然而,对于包括肺部CT图像的医学图像,由于伦理问题、标注成本限制等诸多因素,导致高质量注释医学数据集的缺乏,这自然对现有的基于数据驱动的深度学习方法提出挑战。此外,肺结节的症状和医学语义特征的多样性也增加了深度学习模型的训练难度^[6]。因此,探索一种能够有效缓解数据匮乏的方法,对促进深度学习在医学领域的应用和发展具有重要意义。

数据增强(Data Augmentation, DA)是一种常用的缓解数据不足的方法^[7],其实质是在有限的数据集上,通过特定的规则和算法生成更多的数据。传统的图像数据增强方法,如随机裁剪、缩放、翻转、旋转、平移等,虽然可以在一定程度上增加数据量,但这些方法获得的图像特征仍然与原始数据相似,不能生成足够的新特征。而生成式对抗网络(Generative Adversative Networks, GANs)凭借良好的学习真实样本分布的能力,为数据增强注入了新的动力^[8]。

在肺结节合成中,基于GANs的方法有很多,一

般可分为:(1)以随机向量结合标签向量作为输入,直接生成所需的肺结节图像;(2)以原始结节图像的掩模图像结合肺结节的标签作为输入,生成肺结节图像。然而,前者容易出现结构失真或生成的图像语义信息不正确等问题。后者虽然可以有效避免上述问题,但也会带来新的问题。例如,当原始图像中较大的结节转化为较小的结节时,可能会出现周围组织的畸变。另外,有学者借鉴Auxiliary Classifier GAN、ACGAN^[9],将输入图像的真假判别问题转化为判别输入图像的真、假和肺结节分类问题,这与单纯利用GANs生成能力的肺结节诊断方法不同。但是由于需要考虑训练稳定性等问题,判别器的结构也会相对简单,导致分类精度较低。

针对上述问题,提出一种融合上下文注意力的两阶段生成对抗网络用于肺结节生成和恶性度分类。通过引入两阶段生成模型的思想,将生成肺结节的任务分为两个阶段:生成肺结节感兴趣区域(ROI)和在指定位置注入ROI。为了进一步提高模型的特征提取能力,构建一种通道增强的多头上下文注意力机制,同时,借鉴辅助分类器的思想,构造一个具有肺结节恶性程度分类功能的判别器,以期在生成高质量肺结节图像的同时实现良好的肺结节恶性度分类。本研究的主要内容如下:(1)提出一种通道增强的多头上下文注意力(Channel-Enhanced Multi-head Contextual Transformer, CEMCoT)模块,CEMCoT包含两个主要组件:高效通道注意力(Efficient Channel Attention, ECA)和多头上下文注意力(Multi-head Contextual Transformer, MCoT)模块。ECA通过一个简洁高效的网络,可以自适应地学习不同信道之间的相关性,并调整每个信道的权值。而MCoT支持长距离建模,融合上下文信息,以及同时学习和关注不同的语义信息。(2)提出一种基于两段式的肺结节生成对抗网络(Two-Stage-based Pulmonary Nodule Generative Adversarial Network, TS-PNGAN)。首先构造用于生成肺结节ROI的生成器 G_{ROI} ,然后将生成的肺结节ROI注入到肺部CT图像的指定位置。同时,构建基于CycleGAN的微调模块,使最终的融合图像更加真实。(3)对TS-PNGAN的生成器和判别器进行改造。对于生成器,将CEMCoT引入到生成器中,并构造改进的Attention U-Net对第一阶段的融合图像进行微调。在判别器方面,引入CEMCoT和跨层激励(Skip Layer Excitation, SLE)增强判别器的表示能力,同时引入辅助分类器,使判别器具有对肺结节进行分类的能力。

1 相关工作

随着深度学习在医学图像处理领域的重要发展,与传统的肺结节分类方法相比^[10-11],基于数据驱动的方法,尤其是卷积神经网络(Convolutional Neural Networks, CNNs)显示出更好的分类性能^[12],将用户从繁琐的手工特征提取中解脱出来。一系列基于CNNs的模型被提出并成功地用于肺结节分类。Zhang等^[13]为捕获更多的高维语义信息引入残差网络,从不同尺度提取结节特征来提升模型分类精度。Xie等^[14]将在ImageNet上预训练的3个ResNet-50模型转移到肺部CT图像中,以分别表征肺结节的整体外观、体素值异质性和形状异质性,有效提升分类精度。在此基础上通过将三维结节分解为9个固定视图来学习三维肺结节特征,联合训练了9个基于ResNet-50的模型实现肺结节良恶性分类精度的提升^[3]。Al-Shabi等^[15]研究了兼顾图像局部和全局特征的神经网络模型(例如Densenet^[16]和Resnet^[17])在肺结节良恶性分类任务中的应用,并以残差网络为骨干网络,结合注意力机制提出一种能够使用全局特征提取器分析结节形状和大小,使用局部特征提取器分析结节密度和结构的新方法来预测结节恶性程度。以上方法虽然能够从多尺度、多视图等角度构建肺结节特征提取器以达到精确分类结节的目的,但并未考虑数据增强对模型性能的影响。且受制于伦理问题、时间成本等因素,难以获取大规模带标注的医学图像训练集,这仍旧是制约模型性能的关键因素。

近年来,GANs及其变体在计算机视觉领域中被广泛使用,其强大的生成能力为研究人员解决医学图像样本不足问题提供新思路^[9,18-21]。Chuquicusma等^[22]较早尝试使用深度卷积生成对抗网络(Deep Convolutional Generative Adversarial Networks, DCGAN)生成肺结节CT图像样本以挖掘肺结节高辨别性特征^[19]。Jin等^[23]为进一步提高生成肺结节的质量,解决训练样本分布不均衡的问题,开发一个3D条件生成对抗网络模型,在三维空间学习结节的形状和外观分布,并为使生成结节与背景上下文更好地融合,提出一种多掩码重建损失。Han等^[24]为生成真实、多样化的3D肺结节图像,在考虑结节所在位置、大小、衰减差异等信息的基础上构建3D多条件生成对抗网络,为生成指定肺结节,模型的生成器输入中融合结节的大小和衰减度等类别约束条件,并使用两个判别器分别从结节周围组织信息和结节大小、衰减两个方面判别输入图像的真假。Yang等^[25]为削弱监督任务对数据集的依赖,提出一种类感知对抗合成框架来合成CT图像中的肺结节。该框架包含一个由粗到细的生成器和两个类感知鉴别器,以随机潜变量和目标结节标签为条件,在给定相同

上下文的情况下生成不同的结节。Kowsari等^[26]尝试使用GAN生成多类别肺结节,首先使用GAN生成假的肺结节样本,再使用预训练的分类网络对生成样本进行分类、打标签,最终将标注好的假样本与真样本混合用以训练分类网络。Wang等^[6]为了将肺结节形状、背景、语义特征等作为学习目标,提出一种具有联合损失函数的多目标协同引导合成网络,以实现图像生成和语义特征学习的协同引导,利用前景和背景掩码引导结节形状和肺组织信息,利用肺窗和纵隔窗口引导毛刺和纹理的生成。Toda等^[27]为保证生成肺结节在形状上的多样性,在InfoGAN的基础上引入形状标签控制生成结节的形状,并构建辅助分类器进一步约束生成图像类别。Apostolopoulos等^[28]为精确表征检测到的孤立性肺结节,开发一种特征融合VGG19的分层CNNs以增强模型的特征提取能力,并为避免缺乏足够规模数据来训练,使用深度卷积生成对抗网络生成新的、逼真的结节表示来扩充数据集。Salama等^[29]针对Chset X-ray图像数据集小且不平衡问题,提出一个以卷积变分自动编码器为生成器的生成模型,模型能够生成不同大小和位置的肺部肿瘤图像以扩充训练集,并在重构的数据集上训练了一个基于ResNet-50的网络以实现肺结节良恶性分类。Roy等^[30]针对肺结节类别不均衡问题,提出一种属性驱动的生成对抗网络,设计了融合自注意力空间模块和自注意力属性模块的生成器网络以生成指定属性的样本,并为充分利用判别器性能,将最后的分类器改为了多分类形式,既能判断输入样本真假,也能判别输入样本良恶性及其等级。

尽管上述基于GANs的方法能够生成较高质量的肺结节图像,但仍存在一定的局限性。例如,有些方法不能生成不同恶性度的肺结节图像,也不能将结节准确地注入到肺部指定区域。而且大多数研究主要集中在利用GANs的生成能力,而较少关注判别器。此外,由于以上方法的生成器和判别器网络结构大多较为简单,导致模型的表征能力较弱,有待进一步提升。而本研究提出的TS-PNGAN则能有效缓解上述问题。

2 模型构建

2.1 总体结构设计

鉴于带标注的肺结节样本稀缺,本研究提出一种改进的两段式肺结节生成模型TS-PNGAN,其概述如图1所示。TS-PNGAN的核心思想是将肺结节的生成过程分为两个阶段(Stage I和Stage II),从而获得更准确、高效的生成结果。

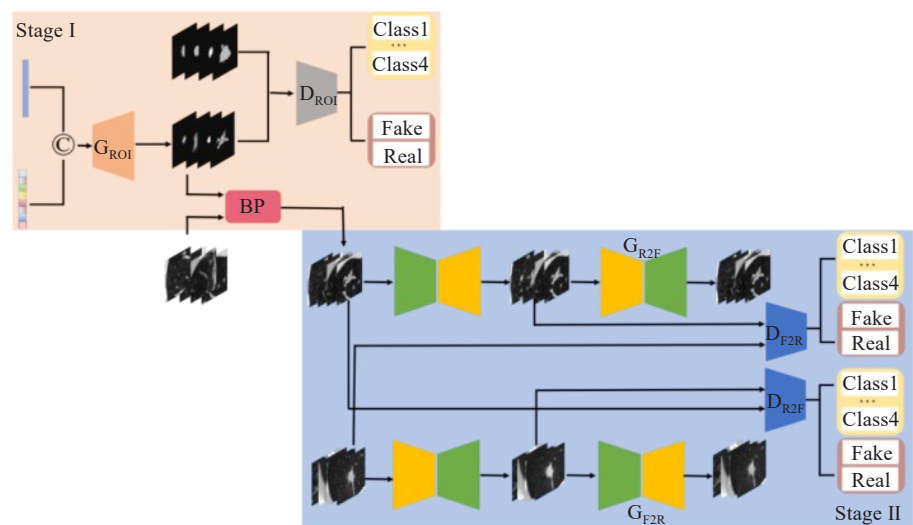


图 1 TS-PNGAN 整体结构
Figure 1 Overall structure of TS-PNGAN

Stage I 设计了一个用于生成肺结节 ROI 的生成器 G_{ROI} 。 G_{ROI} 的目标是通过学习和模拟真实肺结节的形态和纹理特征,生成只包含结节本身特征信息且无周围组织的肺结节图像。随后,将生成的 ROI 图像与从指定肺区域采样的 64×64 大小的切片进行融合。该步骤通过泊松融合 (Poisson Blending, PB) 模块实现,保证了肺结节 ROI 与背景图像之间的自然过渡,最终初步构建一幅包含周围组织信息的完整肺结节图像。选择 64×64 的尺寸是因为在医学图像处理中,细节信息对于准确诊断和分析至关重要,该尺寸通常能够保留足够的细节信息,使得医生能够观察到必要的细节特征,同时也能缓解由于图像过大而导致的计算负担。

Stage II 将初步获得的肺结节图像输入到基于 CycleGAN 的微调模块 (Fine Tuning Module, FTM) 中。CycleGAN 是一个具有强大风格转换能力的生成性对抗网络,支持跨不同领域的图像风格转换。TS-PNGAN 利用微调模块进一步调整图像细节和风格,使其更接近真实肺结节图像,从而将融合图像映射到真实的肺结节样本空间。

通过采用上述方法,TS-PNGAN 能在保留肺结节本身特征信息的同时,兼顾周围组织的上下文信息,最终生成高质量的肺结节图像。随后,详细介绍了所提出的 CEMCoT 模块、TS-PNGAN 模型的生成器和判别器以及对应的损失函数。

2.1.1 CEMCoT 模块 虽然 CNNs 在特征工程相关任务中表现出良好性能,但单独使用卷积建模图像中的远程依赖性存在一定不足,这会使传统的 CNNs 模型在处理复杂、大规模数据时面临计算效率低和特征提取能

力不足的挑战。而通过可扩展方式触发远程交互的注意力机制可以凭借其以高权重聚焦重要信息,以低权重忽略任务不相关信息特性在计算机视觉领域表现出良好性能。然而,自注意力机制主要依赖于独立成对的“查询-键”交互来计算注意力矩阵,一定程度上忽略了相邻键之间丰富的上下文信息和通道之间的关联关系。为此,本研究构建 CEMCoT 模块,其结构如图 2 所示。可以看出,在结构上,CEMCoT 将 ECA 模块和 MCoT 模块结合在一起,这种串联结构允许模型在两个维度上同时关注重要的特征信息,这能有效提高模型的代表能力。

由图 2 可以看出,CEMCoT 在通道维度上利用 ECA 模块来区分和强调不同通道。首先通过全局平均池化对输入的特征图进行压缩以达到全局上下文信息的融合,该步将大小为 $C\times W\times H$ 的特征图压缩为 $C\times 1\times 1$;然后通过一个自适应卷积核尺寸的一维卷积计算通道的权重以使模型更好地关注关键特征,并通过 Sigmoid 激活函数将权重映射在 0~1 之间,其中,卷积核 $k=\lfloor \log_2(C)/\gamma+b/\gamma \rfloor$, $b=1$, $\gamma=2$ 。通过这种方式,CEMCoT 为每个输入特征图生成一个权重向量,这些权重表示每个通道对模型决策的重要性。

在空间维度上,为增强模型表达能力,构建一种新的 MCoT 模块,其结构如图 2 紫色虚线矩形框中所示。单头 CoT 实现了在单一架构中统一键-值对之间的上下文挖掘和特征图上的自注意力,即可以在捕捉特征图全局信息的同时,兼顾 CNN 捕捉局部信息的能力。而多头 CoT 还具备了多尺度特征捕捉、增强特征提取等能力。由图 2 不难看出,MCot 即为单头 CoT 的重构形式。

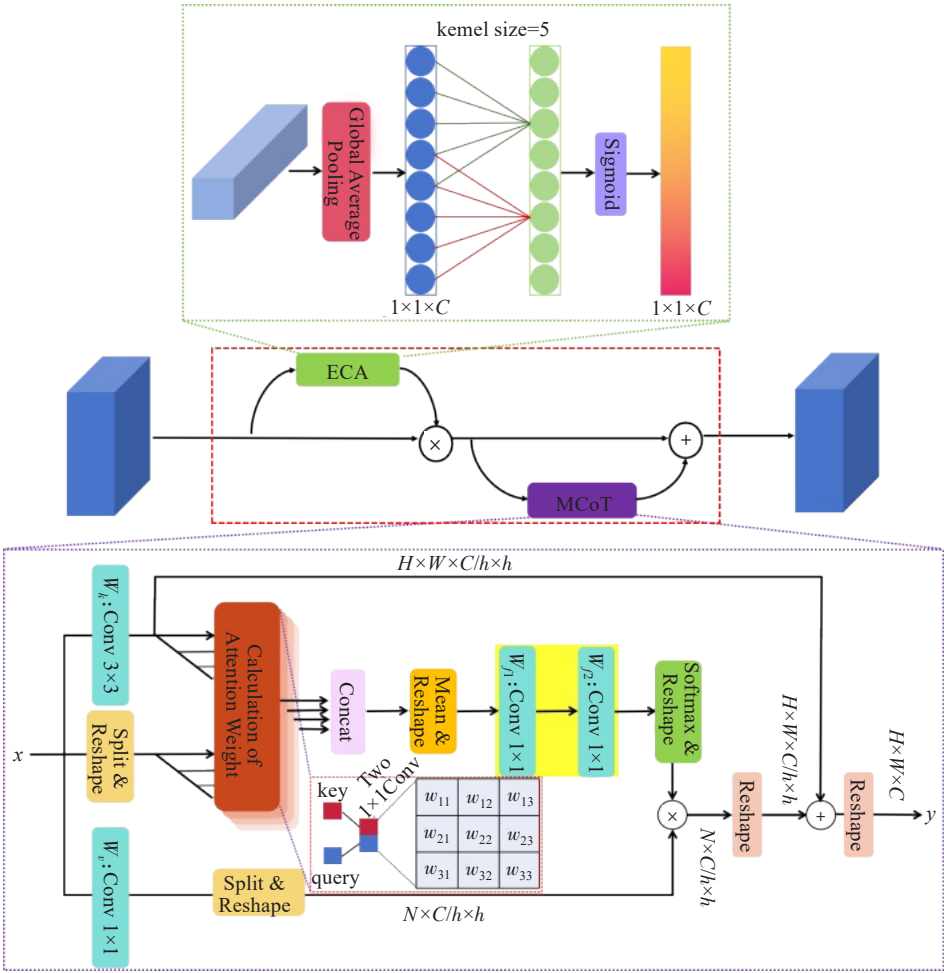


图2 CEMCoT网络结构

Figure 2 Network structure of CEMCoT

具体地,给定一个 $H \times W \times C$ 的特征图 x ,其中 H 、 W 、 C 分别代表特征图的高、宽和通道数。与自注意力机制相似,MCoT 同样是将输入特征图进行 queries、keys、values 编码,并通过构建注意力矩阵使得模型能够根据任务目标有效聚焦输出结果。MCoT 首先通过 3×3 卷积对输入特征图进行上下文编码构建静态键值 keys,卷积操作的引入能够充分挖掘特征图邻近局部信息,并反映出局部邻近键值之间上下文关系。查询 queries 并未做特殊处理,依旧为输入特征图 x 。之后将 keys 和 queries 进行切分构建多头注意力的输入 $H \times W \times C/h \times h$,其中, h 为注意力头的个数。然后将对应拆分后的各子 keys 和 queries 进行拼接,并通过两个连续的 1×1 卷积进行自我注意,生成动态上下文,此时得到的注意力矩阵的通道数为 $3 \times 3 \times C \times h$ 。以单头 CoT 为例注意力矩阵的计算如式(1)所示:

$$A' = \text{Soft max} \left(\text{Mean} \left(W_o \left(\text{ReLU} \left(W_\theta \left(\text{ReLU} \left(W_k x' \right) \odot x' \right) \right) \right) \right) \right) \quad (1)$$

其中,Mean表示对输入特征图按通道求平均,并对结果形状进行调整以便后续计算, W_k 为 3×3 卷积,目标即为获取特征图的邻近局部信息, W_θ 、 W_o 为 1×1 卷积操作, \odot 表示向量的连接操作, x' 表示单头注意下的输入, A' 表示单一注意力头下获得的注意力矩阵。

需要注意的是,本节构建的是多头上下文注意力模块,若要获得最终的注意力矩阵 A ,需将获得的各个子注意力矩阵进行拼接,然后在 Mean 和 Softmax 操作之间插入一个基于两个 1×1 卷积的模块以实现个注意力头之间的信息交互。

得到注意力矩阵后乘以输入值 values 得到动态上下文建模结果,也即获得图像的全局信息。最后将静态和动态上下文表示相融合得到最终输出 y ,其数学表示如式(2)所示:

$$y = (W_v x) \otimes A + \text{ReLU}(W_k x) \quad (2)$$

其中, y 为MCoT模块的输出, W_v 为 1×1 卷积, \otimes 表示两个向量对应位置元素相乘。需要指出的是,为适应多头注意力的计算,values 也需要进行拆分。

2.1.2 基于CEMCoT的 G_{ROI} GAN的生成器通常采用

反卷积或上采样加卷积的方式将输入的随机噪声映射到图像空间。虽然这两种方式能够有效挖掘数据特征,帮助生成器生成逼真、多样化的图片,但这两种方式侧重对特征图中局部邻域信息的处理,因此存在对图像中远程依赖性建模不足的问题。众多研究成果表明将注意力机制融入到生成对抗网络中可以有效提高模型的代表能力,大幅提高生成图像的质量。因此本节在构建TS-PNGAN第一阶段的 G_{ROI} 时引入CEMCoT,以期利用其在处理通道信息和挖掘上下文信息的优势, G_{ROI} 网络结构如图3所示。由图3可以看出,生成器本质上是一个解码器网络,通过若干次反卷积(Transposed Convolution)操作对小

分辨率的图像进行上采样得到符合目标任务的图像。为生成指定类别的肺结节图像,生成器输入为隐变量和标签向量的融合 z ,本文所涉及的隐变量为从均匀分布中提取的100维噪声向量,标签向量则通过对类别标签进行one-hot编码获得。网络主要由5个反卷积块组成,每个Block由反卷积、批归一化、激活函数以及CEMCoT模块中的几个组件依次排列组成。向量 z 通过一个全连接层投影至1024维的数据空间,并重塑为尺寸 $1024 \times 1 \times 1$ 的输入向量,然后通过一系列上采样操作进行分辨率扩张,最终通过一个卷积层将特征表示映射为输出图像。

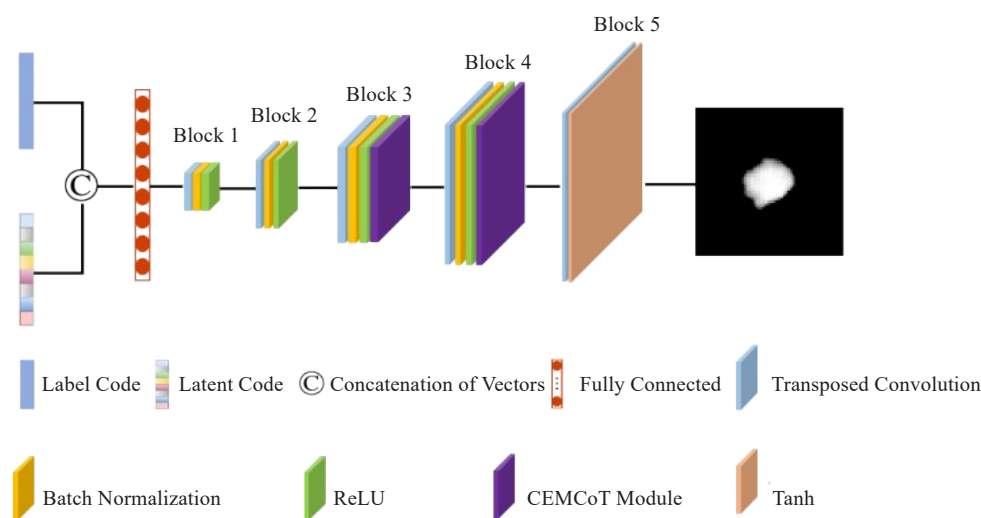


图3 基于CEMCoT的 G_{ROI}
Figure 3 G_{ROI} based on CEMCoT

2.1.3 CycleGAN的生成器结构 鉴于CycleGAN两个生成器相同,本节以生成器 G_{F2R} 为代表展开说明。 G_{F2R} 是通过深入挖掘真实肺结节样本特征对输入的融合图像进行微调,以实现将对应类别的合成样本映射到真实肺结节样本空间,最终生成高质量的肺结节样本, G_{F2R} 的网络结构如图4所示。其构建借鉴作者先前提出的改进CycleGAN^[31]的生成器结构。不同之处在于:首先, G_{F2R} 采用注重特征图上下文信息和通道特征信息的CEMCoT模块替代原模型中的部分 3×3 卷积块;其次,为生成指定类别的肺结节图像,生成器输入添加了标签向量作为条件约束。

由图4可以看出, G_{F2R} 本质上是一个编码器-解码器网络。将标签向量进行编码后与第一阶段生成的肺结节图像结合作为网络输入,经卷积、下采样等一系列操作后在不同分辨率上生成一系列特征图,并通过跳跃连接和注意力门机制与对应解码路径上的

特征图进行拼接从而得到新的包含更丰富信息的特征图,再经过一系列上采样和卷积等操作对特征图分辨率进行扩张,最终通过一个卷积操作得到 $1 \times 64 \times 64$ 的肺结节图像。值得注意的是,图4中描述的“卷积块”集成了卷积、instance normalization和ReLU,其中卷积核大小为 3×3 ,步长为1。“CEMCoT块”包括3个组件:CEMCoT、instance normalization和ReLU激活。采用max pooling作为下采样,反卷积作为上采样。

2.1.4 判别器网络结构 传统的基于CNNs架构的判别器目标是判别输入样本的真假,从而为模型训练提供判别损失以指导生成器训练。但这单一化了判别器功能,在一定程度上浪费了判别器的特征提取能力。为此,借鉴ACGAN引入辅助分类器构建一种面向多任务的判别器网络,本节以 D_{F2R} 为例做详细介绍,网络结构如图5所示。

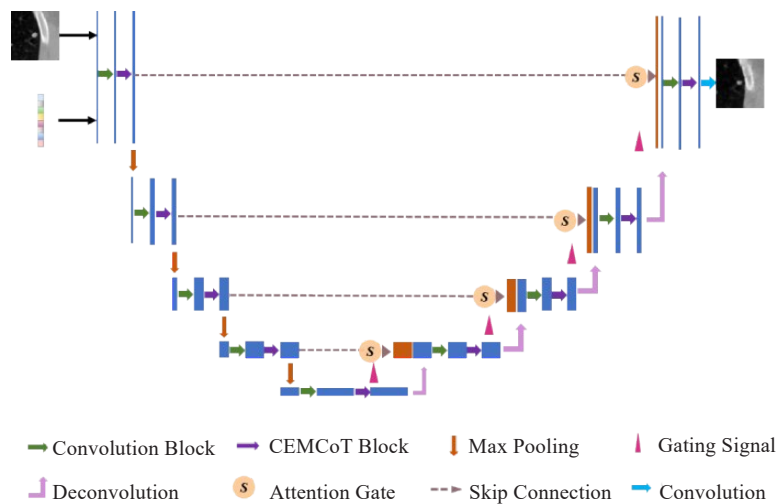


图4 G_{F2R} 网络结构
Figure 4 Network structure of G_{F2R}

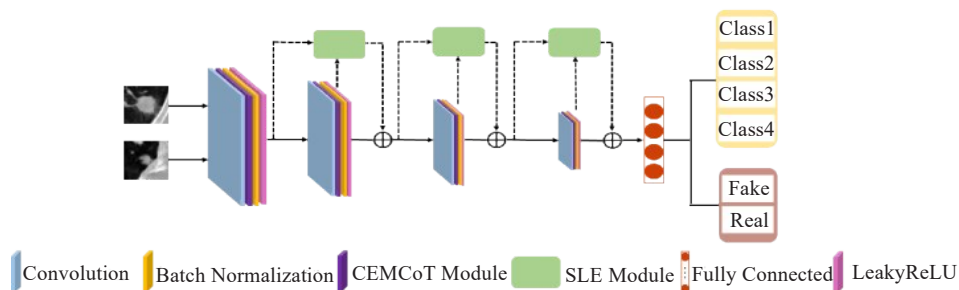


图5 D_{F2R} 网络结构
Figure 5 Network structure of D_{F2R}

从图5可以看出,本节采用的判别器结构与传统判别器有所不同。它包括4个特征提取模块和1个多任务分类器,用于区分输入样本的真实性和分类输入样本的标签。卷积的核大小为 3×3 ;步幅为2;LeakyReLU的斜率为0.1。此外,为进一步提高判别器网络的表征能力,引入CEMCoT和跨层激励模块,

前者能够有效挖掘输入特征在通道上的重要信息以及在空间中的上下文信息,后者则通过自适应学习的方式重新校准各通道的重要程度,SLE模块如图6所示。需要说明的是,鉴于Stage I阶段生成器结构较为简单,在构建Stage I的判别器时也不建议过于复杂,本节在构建Stage I判别器时去除了SLE模块。

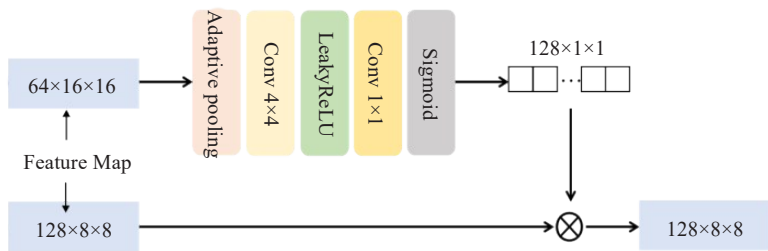


图6 SLE网络结构
Figure 6 Network structure of SLE

2.2 TS-PNGAN 目标函数构造

针对TS-PNGAN生成器、判别器的任务目标,重新构建网络的损失函数。首先描述肺结节ROI生成

网络 G_{ROI} 的损失函数,然后阐明微调阶段CycleGAN生成器的损失函数构成,最后描述训练判别器时所需的损失函数。

2.2.1 G_{ROI} 目标函数 不同于常见的生成损失, G_{ROI} 在对抗损失的基础上引入辅助分类损失来指导生成器生成指定类别的样本。此外, 为了使训练过程更加稳定, 对抗损失采用最小二乘损失, 最终生成网络损失函数 L_G 如式(3)所示:

$$L_G = E_{z \sim p_z} [D_{ROI}(G(z|c)) - 1]^2 + E_{x \sim X_{fake_ROI}} [\log D_{ROI}(C=c|x)] \quad (3)$$

其中, z 为潜变量, c 为样本类别, X_{fake_ROI} 为生成的肺结节 ROI 样本集合。 $E_{z \sim p_z} [D_{ROI}(G(z|c)) - 1]^2$ 为最小二乘损失, $E_{x \sim X_{fake_ROI}} [\log D_{ROI}(C=c|x)]$ 为辅助分类损失表示生成样本被正确分类的概率。

通过最优化式(3), 就能得到一个较高性能的生成器。但为了能够使 G_{ROI} 生成的肺结节更具多样性, 有效缓解训练过程中的模式崩溃现象, 引入模式寻找损失 L_{ms} , 如式(4)所示:

$$L_{ms} = \frac{\text{Dis}(G(z_1|c), G(z_2|c))}{\text{Dis}(z_1, z_2)} \quad (4)$$

其中, $\text{Dis}(\cdot)$ 为距离度量函数, 本论文采用 L_1 距离。

需要注意的是, 模型需要通过最小化损失来优化, 因此式(4)需要做修改如下:

$$L_{ms_new} = \frac{1}{L_{ms} + \varepsilon} \quad (5)$$

其中, ε 为调节参数, 防止出现 L_{ms} 为 0 的情况, 取值 1×10^{-5} 。

综上, 生成器 G_{ROI} 的损失如式(6)所示:

$$L_{G_{ROI}} = E_{z \sim p_z} [D_{ROI}(G(z|c)) - 1]^2 + E_{x \sim X_{fake}} [\log D_{ROI}(C=c|x)] + \lambda L_{ms_new} \quad (6)$$

2.2.2 CycleGAN 生成器目标函数 本节以 G_{F2R} 为代表详述 CycleGAN 生成器目标函数。生成器需要生成指定类别的样本, 因此需在原对抗损失的基础上添加辅助分类损失, 如式(7)所示:

$$L_{adv} = E_{x \sim p_{X_{fake}}} [D_{F2R}(G_{F2R}(h|c)) - 1]^2 + E_{y \sim Y_{fake}} [\log D_{F2R}(C=c|y)] \quad (7)$$

其中, X_{fake} 为肺结节 ROI 与指定的肺部组织融合后的样本空间, Y_{fake} 是生成器 G_{F2R} 将 X_{fake} 转化到真实肺结节样本空间的结果。

为了确保 CycleGAN 的性能及其在训练过程中的稳定性, 还需要引入循环一致性损失和身份损失。这两个损失函数的表达式分别如式(8)、式(9)所示:

$$L_{cycle} = E_{x_{fake} \sim p_{X_{fake}}} \left[\left\| G_{R2F}(G_{F2R}(x_{fake})) - x_{fake} \right\|_1 \right] + E_{x_{real} \sim p_{X_{real}}} \left[\left\| G_{F2R}(G_{R2F}(x_{real})) - x_{real} \right\|_1 \right] \quad (8)$$

$$L_{Id} = E_{x_{fake} \sim p_{X_{fake}}} \left\| G_{F2R}(x) - x_{fake} \right\|_1 +$$

$$E_{x_{real} \sim p_{X_{real}}} \left\| G_{R2F}(x_{real}) - x_{real} \right\|_1 \quad (9)$$

最后构建的 G_{F2R} 损失函数如式(10)所示:

$$L_{G_{F2R}} = L_{adv} + L_{Cycle} + L_{Id} \quad (10)$$

2.2.3 判别器目标函数 本节所构建的3组判别器损失函数相同, 包含两个部分, 最小二乘损失和辅助分类损失, 以判别器 D_{F2R} 为例做详细表述如式(11)所示:

$$L_{D_{F2R}} = E_{x \sim X_{fake}} [(D_{F2R}(G_{F2R}(x)))^2] + E_{x \sim X_{real}} [(D_{F2R}(x) - 1)^2] + L_{C_{F2R}} \quad (11)$$

其中, X_{fake} 和 X_{real} 分别为 Stage I 得到生成样本和真实肺结节样本, $L_{C_{F2R}}$ 为辅助分类损失如式(12)所示:

$$L_{C_{F2R}} = E_{x \sim X_{real}} [\log D_{F2R}(C=c|x)] + E_{y \sim G_{F2R}(X_{fake})} [\log D_{F2R}(C=c|y)] \quad (12)$$

3 数据集与实验设置

3.1 数据集及预处理

本研究使用的数据集是公共数据集 LIDC-IDRI^[32], 它由胸部医学文件(如 CT 和 X 射线图像)和相应的病变诊断注释组成, 共包含 1018 个实例。对于每个实例下的图像, 由 4 位经验丰富的胸部放射科医生进行两阶段诊断标注。肺结节的恶性程度由放射科医师按 1~5 分(Class 1~5)进行分级, 评分越高, 恶性程度越高。需要指出的是, 由于 Class 3 的恶性程度不确定, 因此 Class 3 的样本被丢弃。此外, 由于患者的 CT 图像由多个放射科医生注释, 同一个结节可能有几个不同的恶性评分。因此, 笔者对同一结节的恶性度评分求平均。另外, 由于肺结节存在于三维形态中, 本研究仅提取一幅具有代表性的横断面图像。最后, 从数据集中选取 1559 个案例, 其中 1400 个用于培训, 159 个用于测试。具体样本分布如表 1 所示。

表 1 样本的数据分布

Table 1 Data distribution of samples

| 恶性等级 | Class 1 | Class 2 | Class 4 | Class 5 | 总计 |
|------|---------|---------|---------|---------|------|
| 原始样本 | 321 | 600 | 325 | 313 | 1559 |
| 训练集 | 288 | 541 | 293 | 278 | 1400 |
| 测试集 | 33 | 59 | 32 | 35 | 159 |

为完成模型训练, 除真实肺结节样本数据集外还需构建真实肺结节 ROI 数据集, 以及指定位置的肺实质图像数据集, 图 7 展示了这 3 个数据集的代表性数据, 其中第 1~3 行的图像分别采样于真实肺结节图像、真实肺结节 ROI 图像和指定位置肺实质图像数据集, 图像尺寸均为 64×64 。其中, 第 1、2 列为良性结节, 恶性度等级

分别为 Class 1、Class 2。第 3、4 列为恶性结节, 恶性度等级分别为 Class 4、Class 5。肺实质图像数据集为从 LIDC-IDRI 其他 CT 图像上随机采样的 20 000 张 64×64

的肺部组织图像。由图 7 可以发现, 良性结节表面较为光滑、边缘表现较为清晰, 而恶性结节则表现出表面不光滑、边缘模糊、有毛刺等特征。

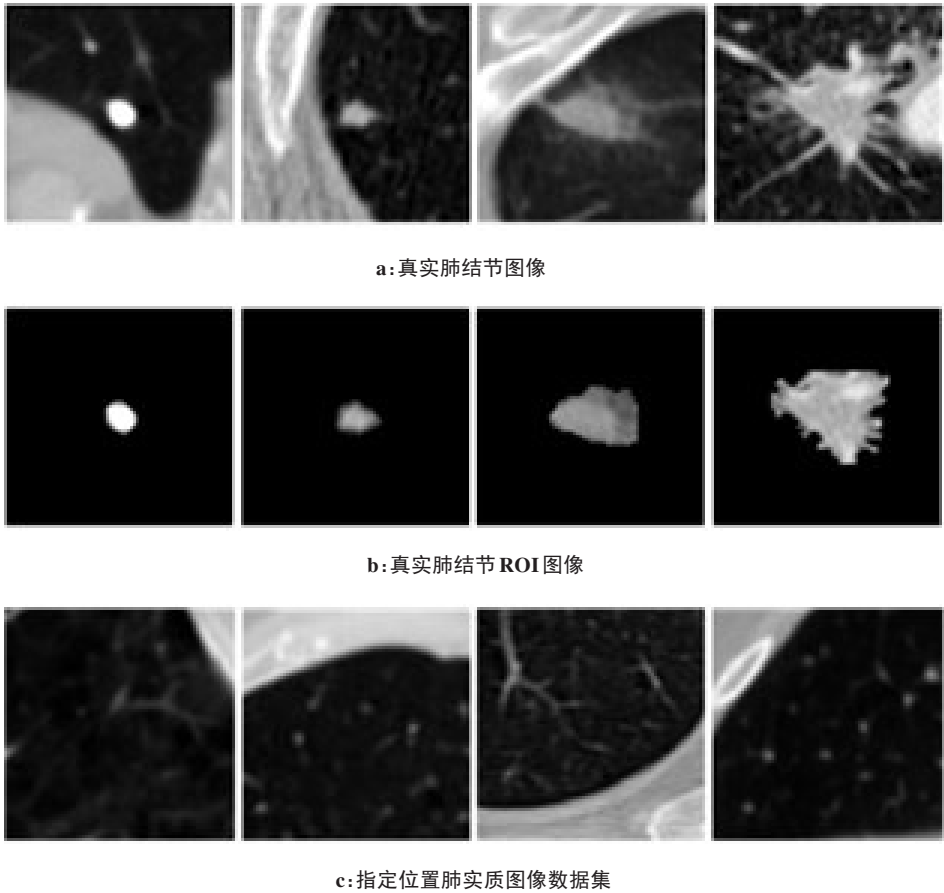


图7 3个样本集合的代表性样例

Figure 7 Representative samples from 3 datasets

第1~4列分别代表恶性度等级为 Class 1、Class 2、Class 4、Class 5

3.2 实验参数设置

实验仿真所用配置为 Intel Core i7-9700K 处理器, NVIDIA RTX 2070S 显卡, 显卡内存为 8 G。在开源框架 Pytorch 上实现网络模型并进行训练。模型训练过程采用 Adam 优化算法进行梯度更新, 每次输入网络的样本数量为 14, 迭代次数设置为 400。Adam 超参数的设置为 $\beta_1=0.5, \beta_2=0.999$, 学习率为 0.000 2。

4 结果与分析

为了验证 TS-PNGAN 的肺结节图像生成效果, 组织实验与 ACGAN^[9]、SAGAN^[21]、ADGAN^[30]、LSGAN^[33]、FastGAN^[34]、F&BGAN^[35] 进行对比分析, 其中 F&BGAN 和 ADGAN 同样采用了两段式的生成框架。为考察 TS-PNGAN 在肺结节分类中的有效性, 组织实验与 ACGAN^[9]、FF-VGG19^[28]、

ADGAN^[30]、ACGAN-Xu^[36], 以及一些现有的骨干网络 如 DenseNet^[16]、ResNet-50^[17]、VGG16^[37] 和 Inception V3^[38] 进行对比。为确保对比结果的公平性, 各方法均在本文构建的训练集下进行重新训练, 并在图像生成性能测试时保证各模型的输入隐变量相同。

4.1 肺结节生成实验结果与分析

为直观地展现各模型的肺结节图像生成效果, 首先选择了各模型具有代表性的生成结果进行视觉感官上的比较分析, 结果如图 8 所示。图 8 由上至下分别为 ACGAN、LSGAN、SAGAN、FastGAN、F&BGAN、ADGAN 和 TS-PNGAN 的生成结果。

由图 8 不难看出, 各模型均能生成较高质量的肺结节图像。仔细观察图 8a、b 中红色矩形区域以及红色箭头所指的区域可以看出, ACGAN、LSGAN 的生

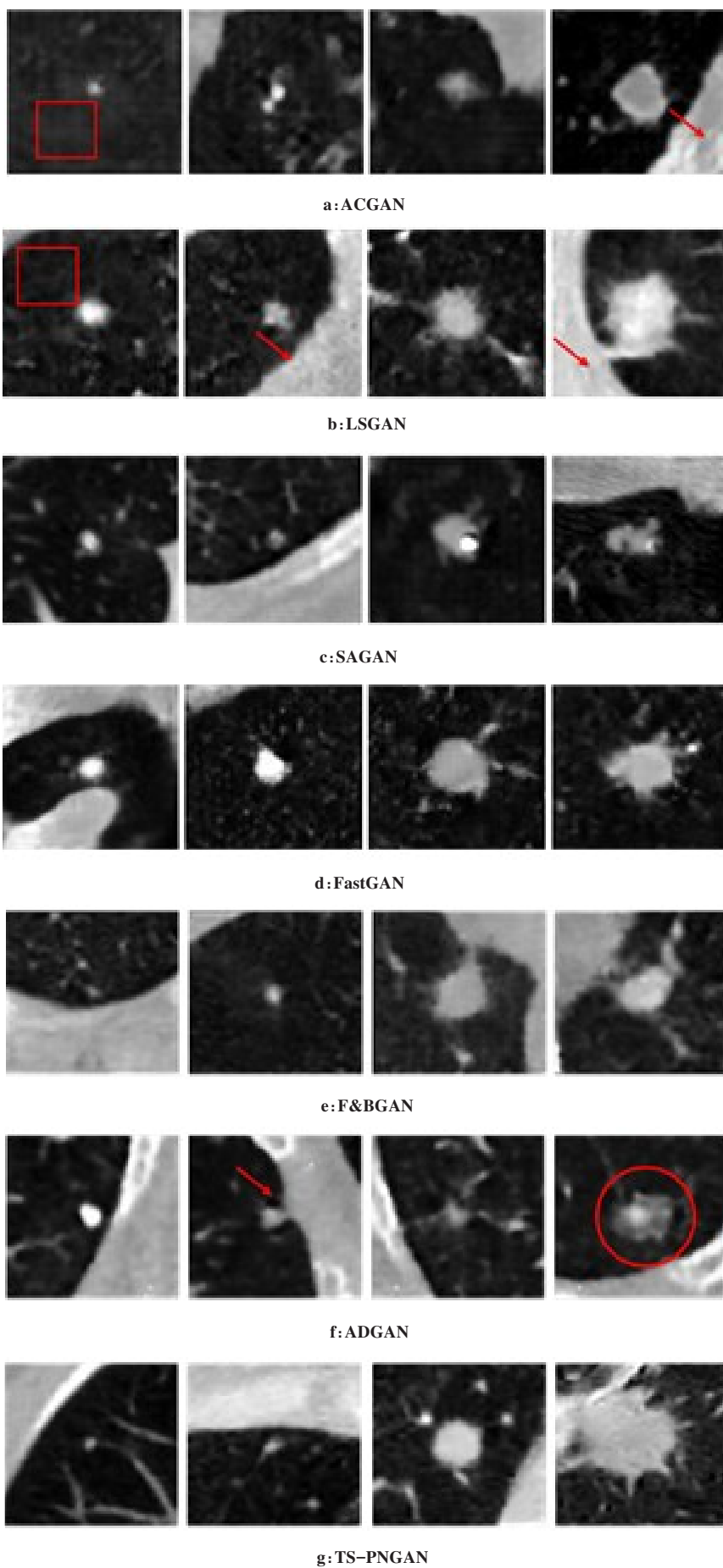


图8 各模型生成的肺结节样本

Figure 8 Pulmonary nodules generated by each model

第1~4列分别代表恶性度等级为 Class 1、Class 2、Class 4、Class 5

成结果在图像细节上稍显不足,表现为肺实质区域细节缺失、胸廓区域纹理及边缘较为粗糙。而SAGAN和FastGAN的表现稍好,这是因为注意力机制的引入以及网络结构的特殊设计有效提升模型的表征能力。这说明依靠单生成器生成图像的GAN模型可以通过引入注意力机制、改善网络结构来提升模型性能。

对比图8a、b和e可以发现,虽未构建新的网络结构或引入注意力机制,但基于两段式生成架构的F&BGAN在图像的细节生成上相对表现更好。这是因为F&BGAN对生成任务进行了拆解,利用递进的方式进行图像生成。相比之下ADGAN和TS-PNGAN的生成结果明显具备更高的视觉质量,不仅在结节自身特征上与图7所展示的真实样本更为接近,其肺实质和胸廓区域上的细节表现也更为自然。这是因为这两种方法均采用真实的肺部图像作为生成图像的背景图,这在很大程度上简化了第二阶段的任务,因此最终生成图像的细节信息会更加真实、自然。但ADGAN与TS-PNGAN的不同之处在于,前者的背景图是在原肺结节图像上进行的操作,后者是在其他肺部切片上获得的。因此ADGAN实质上是对原肺结节图像的再加工,这就使得该方法出现如图8f红色箭头所指示的弊端。由于ADGAN需要在原图上制作背景图,因此要对肺结节ROI区域进行掩码制作,当第二阶段生成的肺结节ROI区域与掩码形状有显著差异时,就会出现如红色箭头所示的割裂感或明显的人工编辑的痕迹。但TS-PNGAN是从其他肺部组织处截取的背景图,并采用PB的方式将生成的肺结节ROI与其进行初步融合后再输入到微调模块得到最终结果,因此能够很好地避免这种情况的发生。

为定量地比较各生成模型的性能,进一步计算各模型生成结果与测试集之间的Fréchet Inception Distance (FID)、Inception Score (IS),以及 Kernel Inception Distance(KID)得分,结果如表2所示。其中,FID通过计算真实样本和生成样本在特征空间分布的Fréchet距离来度量两组样本之间的差异,KID通过计算真实样本和生成样本的特征空间之间的最大均值差异的平方来度量两组样本之间的差异。FID和KID的值越低表示生成结果与测试集数据分布越相似。而IS通过计算生成样本的类概率分布与信息熵的方式实现两组样本的相似性度量,IS值越高表示生成结果的多样性越好。由表2可以看出,相比LSGAN,SAGAN表现更优。这是因为自注意力机制的引入能够有效提升模型的表征能力,从而使模型能够更好地学习真实样本特征以生成更接近真

实样本的数据。ACGAN则得益于辅助分类损失的引入,辅助分类损失可视为模型的类别先验信息,先验信息的引入使得模型将生成任务与类别信息有机结合,从而在一定程度上提升了模型的性能。FastGAN借助优异的网络构建模式,通过跨层激励模块和自监督判别器正则化的引入显著提升了模型表征能力。F&BGAN利用两段式的生成框架,将任务拆解为两个较容易实现的子模块从而提高生成图像的质量。而ADGAN为生成更佳“真实”的样本,充分利用真实的肺结节图像,构建一种新的基于两段式的生成模型。而本文所提方法TS-PNGAN在3个评价指标上均表现出较强竞争力,并在FID评分上取得最佳结果。之所以能够表现出优异性能,是因为TS-PNGAN能够借助两段式结构将任务进行拆解,然后针对性地设计各个组件,例如:为生成更为丰富多样的前景信息(即肺结节ROI图像)引入CEMCoT模块以及 L_{ms} 损失函数;为使生成结果表现得更真实、自然,在Stage II之前构建PB模块,并在Stage II阶段构建基于CycleGAN的融合CEMCoT和注意力门的U-Net结构对融合图像进行微调。也正是由于各组件的有机结合,使得模型能够更有效、更精确地对图像进行表征,从而提高生成图像的质量。

表2 各模型生成性能定量评价
Table 2 Quantitative evaluation of each model in pulmonary nodule synthesis

| 模型 | FID | IS | KID |
|----------|---------|-------------|-------|
| ACGAN | 170.179 | 2.207±0.112 | 0.136 |
| LSGAN | 185.213 | 1.998±0.077 | 0.159 |
| SAGAN | 168.015 | 2.585±0.123 | 0.128 |
| FastGAN | 164.105 | 2.653±0.099 | 0.121 |
| F&BGAN | 169.449 | 2.610±0.169 | 0.127 |
| ADGAN | 123.943 | 2.604±0.159 | 0.061 |
| TS-PNGAN | 115.153 | 2.619±0.095 | 0.062 |

此外,为进一步评估各生成模型的性能,为每个生成模型组织了视觉图灵测试,实验以随机顺序向两位放射科医生提供随机选择的40个真实肺结节图像和40个生成肺结节图像,放射科医生被要求评估所有肺结节图像并将其分为两类:真实图像和生成图像。两名放射科医生对彼此的评估情况互不知情,并且在实验前不会看到真实或生成样本。为了定量分析各模型的生成性能,计算了图灵测试的正确率(Accuracy, ACC)、真阴性率(True Negative Rate, TNR)、真阳性率(True

Positive Rate, TPR)、假阴性率(False Negative Rate, FNR)和假阳性率(False Positive Rate, FPR)指标,结果如表3所示,其中真实图像定义为阴性(Negative),生成图像定义为阳性(Positive)。

表3 各模型在肺结节生成中的视觉图灵测试评价(%)
Table 3 Visual Turing test evaluation of each model in pulmonary nodule synthesis (%)

| 模型 | 医生 | ACC | TPR | FPR | TNR | FNR |
|----------|----|------|------|------|------|------|
| ACGAN | 1 | 75.0 | 67.5 | 17.5 | 82.5 | 32.5 |
| | 2 | 66.3 | 62.5 | 30.0 | 70.0 | 37.5 |
| LSGAN | 1 | 78.8 | 72.5 | 15.0 | 85.0 | 27.5 |
| | 2 | 68.8 | 60.0 | 32.5 | 67.5 | 40.0 |
| SAGAN | 1 | 61.3 | 52.5 | 30.0 | 70.0 | 47.5 |
| | 2 | 62.5 | 57.5 | 32.5 | 67.5 | 42.5 |
| FastGAN | 1 | 60.0 | 47.5 | 27.5 | 72.5 | 52.5 |
| | 2 | 61.3 | 52.5 | 30.0 | 70.0 | 47.5 |
| F&BGAN | 1 | 71.3 | 67.5 | 25.0 | 75.0 | 32.5 |
| | 2 | 65.0 | 62.5 | 32.5 | 67.5 | 37.5 |
| ADGAN | 1 | 52.5 | 47.5 | 42.5 | 57.5 | 52.5 |
| | 2 | 55.0 | 45.0 | 35.0 | 65.0 | 55.0 |
| TS-PNGAN | 1 | 51.3 | 42.5 | 40.0 | 60.0 | 57.5 |
| | 2 | 53.8 | 45.0 | 37.5 | 62.5 | 55.0 |

表3显示,对于ACGAN、LSGAN、SAGAN、FastGAN、F&BGAN、ADGAN和TS-PNGAN,放射科医生正确分类真实图像和生成图像的平均准确率分别为70.70%、73.80%、61.90%、60.70%、68.15%、53.75%和52.55%。与其他模型相比,放射科医生在评估TS-PNGAN时,其正确识别真实图像百分比以及正确识别生成图像的百分比均有明显下降,表明所提方法比对照组模型生成的结节在视觉上更加真实。观察放射科医生正确识别的生成图像发现,这些生成图像存在一些显著特征:胸廓处的组织结构缺乏一定的密度变化;生成图像的肺实质区域缺乏细节特征、人工痕迹较明显。这在对照模型的生成结果中更为常见,以上情形也能从图8中观察到。

4.2 肺结节恶性度分类实验结果与分析

TS-PNGAN在实现肺结节图像生成的同时,也能完成肺结节的恶性度等级分类。为验证TS-PNGAN的分类性能,在相同实验条件下组织实验,与FF-VGG19、ACGAN、ACGAN-Xu、ADGAN 4种基于GAN的分类模型,以及4种基准分类模型ResNet-50、VGG16、DenseNet、Inception V3的分类性能进行对比

分析。其中,FF-VGG19是首先利用GAN进行数据增广,然后构建新的基于VGG19的分类网络实现肺结节分类;ACGAN、ACGAN-Xu和ADGAN则是将判别器重构为面向多任务的网络结构。需要指出的是,为适配本章所需的肺结节恶性度等级4分类,本节对以上模型做了一定修改。

各模型训练采用五折交叉验证,取平均值作为最终结果,采用ACC、灵敏度(Sensitivity, SEN)、F1值以及曲线下面积(Area Under Curve, AUC)值对各模型性能进行定量评估,结果如表4所示。通过表4可以看出,TS-PNGAN在ACC、SEN和F1指标上表现最优,与对照组模型相比,其能更准确地识别这4种类型的肺结节,体现出模型的准确性、可靠性。对比ACGAN和ACGAN-Xu可以发现,虽然两者模型结构基本相同,但后者的性能更佳,这是因为ACGAN-Xu的训练集是由28×28的肺结节样本构成,这便于模型聚焦肺结节的关键信息,降低网络的训练难度。此外,比较ACGAN、ACGAN-Xu、ADGAN和TS-PNGAN与其他4种基准分类模型的网络结构可以发现,虽然这些基于GAN模型的判别器结构相对更为简单,但最终却取得更优的分类结果。这在一定程度上得益于GAN模型的训练策略,这4种模型得益于生成器提供的对抗性训练以及生成样本对判别器提供数据正则化。

表4 不同模型肺结节恶性度分类性能对比(%)
Table 4 Performance comparison among different models in pulmonary nodule malignance classification (%)

| 模型 | ACC | SEN | F1 值 | AUC |
|--------------|-------|-------|-------|-------|
| ResNet-50 | 61.01 | 59.14 | 58.83 | 81.84 |
| VGG16 | 60.88 | 60.29 | 58.79 | 81.16 |
| DenseNet | 61.22 | 59.20 | 58.27 | 81.50 |
| Inception V3 | 63.36 | 61.89 | 61.27 | 84.47 |
| FF-VGG19 | 66.98 | 62.52 | 64.11 | 84.92 |
| ACGAN | 62.39 | 62.27 | 61.30 | 85.19 |
| ACGAN-Xu | 66.29 | 66.79 | 65.43 | 87.44 |
| ADGAN | 68.76 | 67.56 | 67.54 | 87.78 |
| TS-PNGAN | 70.23 | 68.66 | 68.92 | 87.59 |

为更直观地分析各模型的分类性能,分别选取各模型第五折的训练结果为代表绘制了肺结节恶性度等级分类结果的混淆矩阵,如图9所示。通过图9可以发现,模型在处理相近类别的样本时易出现错分的情况,结合错分结节的图像不难看出,类别相近的肺结节样本具有相似的图像特征,这会在模型挖

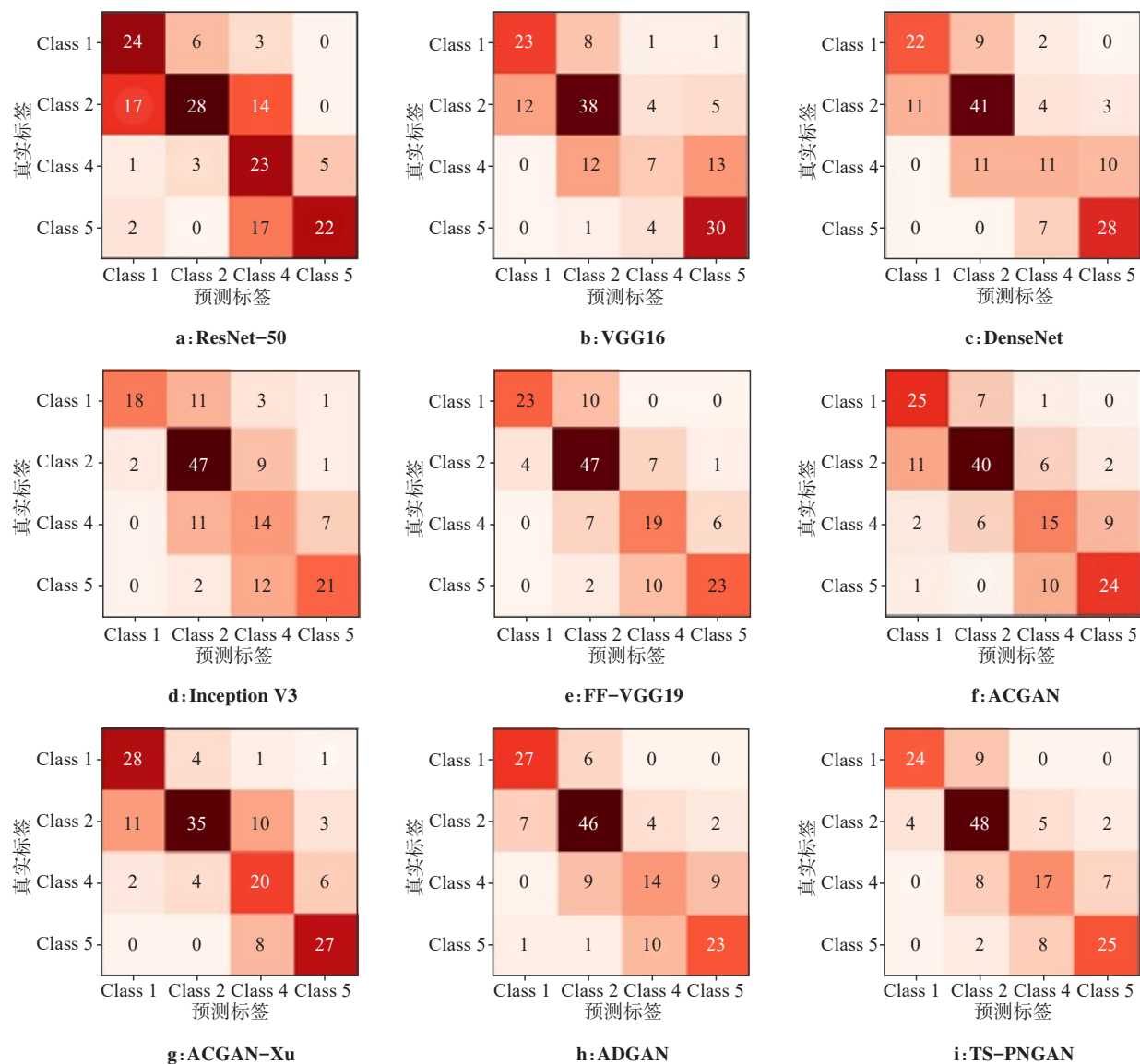


图9 肺结节分类混淆矩阵

Figure 9 Confusion matrixes of pulmonary nodule classification

掘这类图像的判别性特征时带来阻碍,从而一定程度上影响模型分类性能,这也是多分类任务面临的关键问题。例如,相较于Class 1的肺结节样本,恶性度等级为Class 2的肺结节被错分的概率更大。这是因为该类结节的尺寸一般较小,与Class 1的肺结节表现出相似特征,同时也有部分Class 2的肺结节样本处于肺叶边缘与胸膜黏连、呈现出不规则形状、有磨玻璃表现等特征,表现出与Class 4相似的特征。相较于Class 5类别的样本,Class 4肺结节的误检率更高,这也是因为Class 4类别的部分样本尺寸相对较小,会表现出与Class 2肺结节样本相似的特性。这就给模型的训练带来一定阻碍,考验了模型对相近类别肺结节关键特征的提取能力。综合考虑模型对4种样本的分类结果不难看出,TS-PNGAN表现出更优的性能。

4.3 消融实验

为验证 TS-PNGAN 中模式寻找损失、CEMCoT 模块和 PB 模块的有效性,本节对以上 3 个组件进行消融实验。需要指出的是,鉴于模型训练过程中模式寻找损失 L_{ms} 仅作用于生成器 G_{ROI} ,关于 L_{ms} 的对比实验按如下展开:实验在 G_{ROI} 的基础上通过是否引入 L_{ms} 来观察生成肺结节 ROI 区域的多样性。关于 CEMCoT 模块在图像生成任务中有效性的验证则是设计两组实验:一是通过比较有无 CEMCoT 模块下 G_{ROI} 的生成效果;二是考察微调模块有无 CEMCoT 时的性能。最后,验证 PB 模块在前、背景融合上的有效性。

表 5 中 G_{ROI} 为不引入 CEMCoT 模块的生成器,可以看出当添加了 CEMCoT 模块后模型在 FID、IS 和 KID 3 个指标上均有显著提升,表明 CEMCoT 模块在图像生

成任务中的有效性。 L_{ms} 的引入虽然在FID和KID上并未起到明显作用,但在多样性也就是IS指标上有所提升,表明 L_{ms} 在提升结节多样性上的有效性。本节还考察了有无CEMCoT时微调模块FTM的性能,表5中FTM-CEMCoT表示微调模块中不含CEMCoT,FTM+CEMCoT表示添加了CEMCoT。可以看出,随着CEMCoT模块的引入,生成的肺结节图像在3个评价指标上均有小幅度提升,这是因为CEMCoT能充分挖掘特征图的通道信息,并利用输入键之间的上下文信息指导注意力矩阵的构建以增强网络的视觉表征能力。此外,还比较了模型在使用CoT和CEMCoT时的性能,不难发现相较于CoT机制,CEMCoT的引入使得模型性能有了进一步提升,充分表明了CEMCoT在肺结节图像生成中的优越性,证明本文所提的改进策略的有效性。

为验证PB模块的有效性,设计了两组实验:分别在训练过程中采用PB和Cut-and-Paste融合方式进行肺结节ROI与指定位置肺组织图像的融合。为更好对比两

种模式下生成图像的不同,在测试阶段采用相同的输入以及相同的肺部组织图像。图10展示这两种融合方式下TS-PNGAN生成的代表性样例,其中图10a、b分别为使用Cut-and-Paste和PB方式得到的结果。图中红色箭头所指示的区域即为两组结果之间的差异之处,其中各个图片右下角的小图为箭头所指位置的放大图。对比图10a、b的第1列,可以看出当使用Cut-and-Paste方式融合前景图像即肺结节ROI图像和背景图像也即指定位置肺部组织图像时,由于Cut-and-Paste简单粗暴的融合方式使得两者之间可能会出现截断感,即使通过微调模块也可能难以缓解这种现象,这在图10a的其他图像中也能有所体现。相反,PB的方式能够使得前景图像和背景图像的融合更平滑也更为自然,在一定程度上缓解Cut-and-Paste带来的截断感,从而有效提升生成图像的质量。

5 结 论

为了缓解标注肺结节样本不足,提升肺结节恶性度分类精度等问题,提出一种融合上下文注意力的两段式生成对抗网络TS-PNGAN用于肺结节图像生成和恶性度等级分类。该模型能够从数据增广的角度出发,通过一个基于两段式的生成对抗网络框架生成“假样本”来缓解训练样本不足的问题。为了更好地挖掘图像特征以生成高质量的肺结节图像,提出一种CEMCoT机制充分挖掘特征图的通道信息和上下文信息以增强模型的表示能力。同时,为引导模型在生成肺结节时保持一定的多样性,构建了融合模式寻找损失的生成器网络损失函数。并为更真实、自然地在指定位置注入结节,引入PB模块,实现肺结节ROI图像与背景图像

表 5 消融实验定量评价
Table 5 Quantitative evaluation of ablation study

| 模型 | FID | IS | KID |
|-----------------------------|---------|-------------------|-------|
| G_{ROI} | 51.001 | 1.715 ± 0.050 | 0.160 |
| $G_{ROI} + CEMCoT$ | 43.242 | 1.981 ± 0.076 | 0.128 |
| $G_{ROI} + CEMCoT + L_{ms}$ | 43.080 | 2.108 ± 0.086 | 0.125 |
| FTM - CEMCoT | 129.333 | 2.558 ± 0.151 | 0.078 |
| FTM+CoT | 120.052 | 2.621 ± 0.119 | 0.064 |
| FTM+CEMCoT | 115.153 | 2.619 ± 0.095 | 0.062 |

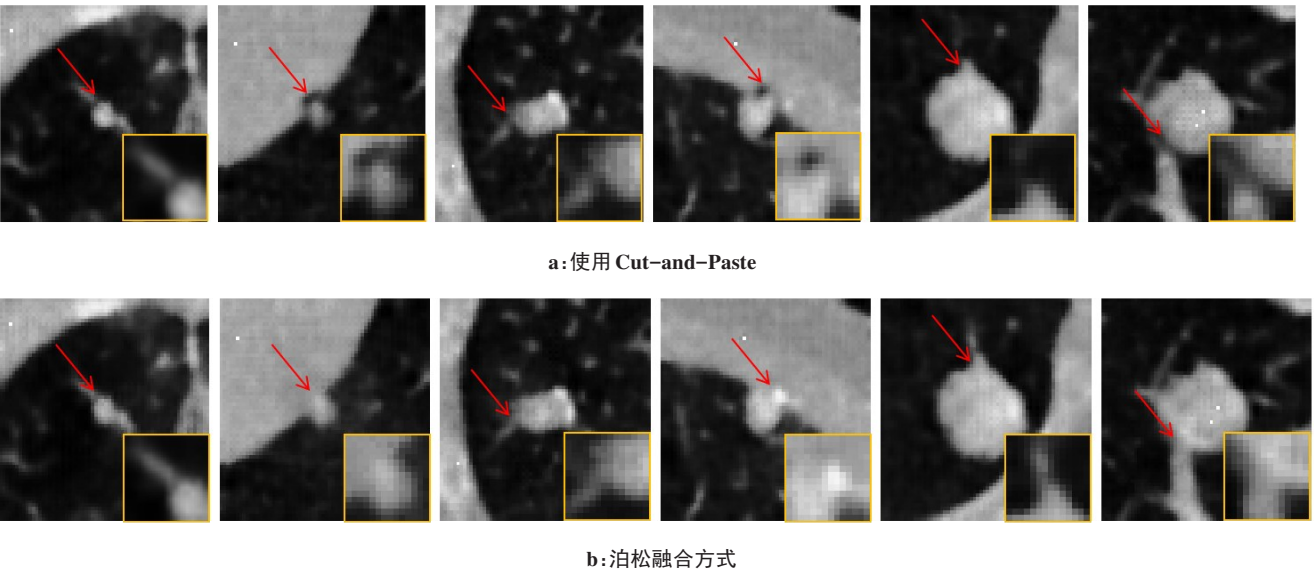


图 10 不同融合模式下 TS-PNGAN 的生成结果
Figure 10 Synthesis results using TS-PNGAN under different fusion modes

即指定位置肺部组织的融合。此外,重构模型的判别器网络,通过引入辅助分类器实现判别器网络既能判断结节真假也能够实现结节的恶性度等级分类。通过在公开数据集LIDC-IDRI进行实验,结果表明本文构建的TS-PNGAN能够生成高质量的肺结节图像,并在FID、IS、KID评价指标上表现出明显优势,同时在肺结节分类上能够在分类ACC、SEN等指标上表现出优于ADGAN等基于GAN的分类模型,以及VGG16等基准网络的性能。

【参考文献】

- [1] Siegel RL, Miller KD, Fuchs HE, et al. Cancer statistics, 2022[J]. CA Cancer J Clin, 2022, 72(1): 7-33.
- [2] Hoffman PC, Mauer AM, Vokes EE. Lung cancer[J]. Lancet, 2000, 355(9202): 479-485.
- [3] Xie YT, Xia Y, Zhang JP, et al. Knowledge-based collaborative deep learning for benign-malignant lung nodule classification on chest CT[J]. IEEE Trans Med Imaging, 2019, 38(4): 991-1004.
- [4] Mobiny A, Yuan PY, Cicalese PA, et al. Memory-augmented capsule network for adaptable lung nodule classification[J]. IEEE Trans Med Imaging, 2021, 40(10): 2869-2879.
- [5] Sahu P, Yu DT, Dasari M, et al. A lightweight multi-section CNN for lung nodule classification and malignancy estimation[J]. IEEE J Biomed Health Inform, 2019, 23(3): 960-968.
- [6] Wang QL, Zhang XH, Zhang W, et al. Realistic lung nodule synthesis with multi-target co-guided adversarial mechanism[J]. IEEE Trans Med Imaging, 2021, 40(9): 2343-2353.
- [7] Wang YL, Huang G, Song SJ, et al. Regularizing deep networks with semantic data augmentation[J]. IEEE Trans Pattern Anal Mach Intell, 2022, 44(7): 3733-3748.
- [8] Goodfellow IJ, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[C]//Proceedings of the 27th International Conference on Neural Information Processing Systems. Cambridge, MA, USA: MIT Press, 2014: 2672-2680.
- [9] Odena A, Olah C, Shlens J. Conditional image synthesis with auxiliary classifier GANs[C]//Proceedings of the 34th International Conference on Machine Learning. Chia Laguna Resort, Sardinia, Italy: PMLR, 2017: 2642-2651.
- [10] Farag AA, Ali A, Elshazly S, et al. Feature fusion for lung nodule classification[J]. Int J Comput Assist Radiol Surg, 2017, 12(10): 1809-1818.
- [11] Li XX, Li B, Tian LF, et al. Automatic benign and malignant classification of pulmonary nodules in thoracic computed tomography based on RF algorithm[J]. IET Image Process, 2018, 12(7): 1253-1264.
- [12] Gu JX, Wang ZH, Kuen J, et al. Recent advances in convolutional neural networks[J]. Pattern Recognit, 2018, 77: 354-377.
- [13] Zhang GK, Zhu DD, Liu X, et al. Multi-scale pulmonary nodule classification with deep feature fusion via residual network[J]. J Ambient Intell Humaniz Comput, 2023, 14(11): 14829-14840.
- [14] Xie YT, Xia Y, Zhang JP, et al. Transferable multi-model ensemble for benign-malignant lung nodule classification on chest CT[C]//Medical Image Computing and Computer Assisted Intervention-MICCAI 2017. Cham: Springer International Publishing, 2017: 656-664.
- [15] Al-Shabi M, Lan BL, Chan WY, et al. Lung nodule classification using deep local-global networks[J]. Int J Comput Assist Radiol Surg, 2019, 14(10): 1815-1819.
- [16] Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2017: 2261-2269.
- [17] He KM, Zhang XY, Ren SQ, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2016: 770-778.
- [18] Denton E, Chintala S, Szlam A, et al. Deep generative image models using a Laplacian pyramid of adversarial networks[C]//Proceedings of the 28th International Conference on Neural Information Processing Systems-Volume 1. Cambridge, MA, USA: MIT Press, 2015: 1486-1494.
- [19] Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks[EB/OL]. (2016-01-07)[2024-04-10]. <https://arxiv.org/abs/1511.06434>.
- [20] Gulrajani I, Ahmed F, Arjovsky M, et al. Improved training of wasserstein GANs[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. Red Hook, NY, USA: Curran Associates Inc., 2017: 5769-5779.
- [21] Zhang H, Goodfellow I, Metaxas D, et al. Self-attention generative adversarial networks[C]//Proceedings of the 36th International Conference on Machine Learning. Chia Laguna Resort, Sardinia, Italy: PMLR, 2019: 7354-7363.
- [22] Chuquicusma MJ, Hussein S, Burt J, et al. How to fool radiologists with generative adversarial networks? A visual Turing test for lung cancer diagnosis[C]//2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). Piscataway, NJ, USA: IEEE, 2018: 240-244.
- [23] Jin DK, Xu ZY, Tang YB, et al. CT-realistic lung nodule simulation from 3D conditional generative adversarial networks for robust lung segmentation[C]//Medical Image Computing and Computer Assisted Intervention-MICCAI 2018. Cham: Springer International Publishing, 2018: 732-740.
- [24] Han C, Kitamura Y, Kudo A, et al. Synthesizing diverse lung nodules wherever massively: 3D multi-conditional GAN-based CT image augmentation for object detection[C]//2019 International Conference on 3D Vision (3DV). Piscataway, NJ, USA: IEEE, 2019: 729-737.
- [25] Yang J, Liu SQ, Grbic S, et al. Class-aware adversarial lung nodule synthesis in CT images[C]//2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019). Piscataway, NJ, USA: IEEE, 2019: 1348-1352.
- [26] Kowsari Y, Mahdavi Chabok SJ, Moattar MH. Classification of pulmonary images by using generative adversarial networks[C]//2020 8th Iranian Joint Congress on Fuzzy and Intelligent Systems (CFIS). Piscataway, NJ, USA: IEEE, 2020: 133-137.
- [27] Toda R, Teramoto A, Tsujimoto M, et al. Synthetic CT image generation of shape-controlled lung cancer using semi-conditional InfoGAN and its applicability for type classification[J]. Int J Comput Assist Radiol Surg, 2021, 16(2): 241-251.
- [28] Apostolopoulos ID, Papathanasiou ND, Panayiotakis GS. Classification of lung nodule malignancy in computed tomography imaging utilising generative adversarial networks and semi-supervised transfer learning[J]. Biocybern Biomed Eng, 2021, 41(4): 1243-1257.
- [29] Salama WM, Shokry A, Aly MH. A generalized framework for lung cancer classification based on deep generative models[J]. Multimed Tools Appl, 2022, 81(23): 32705-32722.
- [30] Roy R, Mazumdar S, Chowdhury AS. ADGAN: attribute-driven generative adversarial network for synthesis and multiclass classification of pulmonary nodules[J]. IEEE Trans Neural Netw Learn Syst, 2024, 35(2): 2484-2495.
- [31] Yin ZX, Xia KW, Wang SJ, et al. Unpaired low-dose CT denoising via an improved cycle-consistent adversarial network with attention ensemble[J]. Vis Comput, 2023, 39(10): 4423-4444.
- [32] Armato SG, McLennan G, Bidaut L, et al. The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans[J]. Med Phys, 2011, 38(2): 915-931.
- [33] Mao XD, Li Q, Xie HR, et al. Least squares generative adversarial networks[C]//2017 IEEE International Conference on Computer Vision (ICCV). Piscataway, NJ, USA: IEEE, 2017: 2813-2821.
- [34] Liu BC, Zhu YZ, Song KP, et al. Towards faster and stabilized GAN training for high-fidelity few-shot image synthesis[C]//9th International Conference on Learning Representations, ICLR 2021. New York, USA: ICLR, 2021: 1-22.
- [35] Zhao DF, Zhu DD, Lu JW, et al. Synthetic medical images using F&BGAN for improved lung nodules classification by multi-scale VGG16[J]. Symmetry, 2018, 10(10): 519.
- [36] 徐久强,洪丽萍,朱宏博,等.一种用于肺结节恶性度分类的生成对抗网络[J]. 东北大学学报(自然科学版), 2018, 39(11): 1556-1561.
- [37] Xu JQ, Hong LP, Zhu HB, et al. Generative adversarial networks for the classification of Lung nodules malignant[J]. Journal of Northeastern University (Natural Science), 2018, 39(11): 1556-1561.
- [38] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[EB/OL]. (2015-04-10)[2024-04-15]. <https://arxiv.org/abs/1409.1556>.
- [39] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2016: 2818-2826.

(编辑:薛泽玲)