

基于双重注意力机制的间质性肺病高分辨率 CT 图像分类方法^{*}

赵琪玉¹ 张俊华¹ 张剑青² 徐铭蔚²

(1. 云南大学信息学院 昆明 650500; 2. 昆明医科大学第一附属医院 昆明 650032)

摘要: 为了更精确地分类间质性疾病,提出了一种基于深度学习的分类网络,首先将多头自注意力机制模块 DenseNet-121 结合,使得模型能够同时关注多个重点区域。然后采用卷积注意力模块实现更高效的特征提取,提升网络的空间感知能力,从而增强分类性能。最后,添加改进的空间金字塔池化层将不同尺度的特征图拼接起来以捕获更丰富的空间信息。此外针对高分辨率 C 图像数据集类别不平衡问题,引入 Focal Loss 损失函数,使得模型在训练时更专注于难分类的样本,从而进一步增强模型的分类能力。所提方法在未经训练的数据集上进行测试,达到了 88.28% 的准确率。相较于原始 DenseNet-121 在准确率、召回率、精确率、F1 分数和 Kappa 系数提高了 4.65%、5.08%、5.82%、5.45% 和 6.38%。实验结果表明,该方法具有特征提取能力强和分类准确率高的特点。

关键词: 间质性肺病;深度学习;注意力机制;DenseNet-121;高分辨率 CT 图像

中图分类号: TP391 **文献标识码:** A **国家标准学科分类代码:** 520.6040

High resolution CT image classification method for interstitial lung disease based on dual attention mechanism

Zhao Qiyu¹ Zhang Junhua¹ Zhang Jianqing² Xu Mingwei²

(1. School of Information, Yunnan University, Kunming 650500, China;

2. The First Affiliated Hospital of Kunming Medical University, Kunming 650032, China)

Abstract: In order to classify interstitial diseases more accurately, this paper proposes a classification network based on deep learning Xi, which first combines the multi-head self-attention mechanism module with DenseNet-121, so that the model can focus on multiple key regions at the same time. Then, the convolutional attention module is used to achieve more efficient feature extraction and spatial perception capabilities, so as to improve the classification ability of the network. Finally, an improved spatial pyramid pooling layer is added to stitch together the feature maps of different scales to capture richer spatial information. In addition, aiming at the problem of category imbalance of high-resolution CT image datasets, the Focal Loss function is introduced, which makes the model focus more on the difficult samples during training, so as to further improve the classification ability of the model. The proposed method is tested on the untrained dataset in this paper, and the accuracy rate reaches 88.28%. Compared with the original DenseNet-121, the accuracy, recall, precision, F1 score and Kappa coefficient are increased by 4.65%, 5.08%, 5.82%, 5.45% and 6.38%. Experimental results show that the proposed method has the characteristics of strong feature extraction ability and high classification accuracy.

Keywords: interstitial lung disease; deep learning; attention mechanism; DenseNet-121; high resolution CT image

收稿日期:2024-01-08

^{*} 基金项目:国家自然科学基金(62063034,61841112)项目资助

0 引言

结缔组织疾病(connective tissue disease,CTD)是一组与自身免疫有关的异质性疾病,可导致自身免疫介导的器官损伤。间质性肺病(interstitial lung disease,ILD)是CTD患者常见的肺部并发症^[1]。ILD在CTD中除了有着较高的发病率,也和CTD患者的死亡密切相关。由于CTD相关ILD的诊断与鉴别方法复杂,病情评估需要多个科室共同参与治疗方案的选择。因此,充分了解CTD-ILD患者肺部影像学的特征有助于对CTD患者早期诊断,并指导进一步诊疗。目前,胸部薄层高分辨率CT(high resolution CT,HRCT)是主要的诊断方法,该方法具有灵敏度高,高分辨率,耗时短等特点。另外,胸部HRCT检测可以提供病人各个阶段的详细资料,可以清晰地显示肺组织的细微结构,从而能更好的判断病人的情况,提供相应的治疗方案。但是由于CTD-ILD的类型繁多,不同CTD的ILD在临床表现、肺部影像和病理特征上各有差异,呈现各自不同的发展与转归,给诊断和治疗带来了挑战。为应对这一难题,需要医生进行初步筛查,以便及时发现尽早治疗^[2-3]。而胸部HRCT图像的判断,高度依赖于医生的专业知识,因此易受医生经验、疲劳度等主观因素影响。综上所述,利用人工智能对CTD-ILD患者胸部HRCT图像进行分类,显得更具实际意义。

自ILD检测以来,基于数字图像处理和模式识别技术的研究取得了巨大进展。典型的过程为提取输入数据的特征,然后馈送到分类器进行分类。近年来人工智能已经在计算机视觉方面取得诸多成就,其中以卷积神经网络(convolutional neural network,CNN)为基础的医学图像分类与分割应用成熟^[4-6]。

对于ILD模式分类的问题,国内外相关研究表明,利用深度学习对间质性肺病HRCT进行识别分类是行之有效的。Gao等^[7]使用公开数据集,尝试了一种ILD模式分类的新方法,因为他们意识到手动识别自动肺计算机辅助诊断系统的感兴趣区域的困难,表明了利用深度学习的方法可以准确检测ILD。Koo等^[8]收集了1 085名患者的胸部CT图像,利用EfficientNet-B3网络进行特征提取,构建了一种深度学习算法对CT切片进行分类,其准确率可达88%。Walsh等^[9]收集了1 157名患者的HRCT图像,开发了一种算法将几个CT切片分类为不同的ILD模式,其测试集的准确率达到了76.4%,但该算法具有一定的局限性,需要将肺部的CT切片进行分割后重新采样,增加了数据处理的复杂性和工作量。Choe等^[10]搜集了288名患者的HRCT图像,设计了一个专为ILD模式的图像检索网络,该网络基于内容进行分类,其准确率达到了85.0%,然而,该网络的功能主要局限于作为一个搜索引擎使用,通过查询图像的视觉内容与存档中的内容进行匹配,以寻找外观相似的图像,其分类准确率并不高,这在一定程度上限制了其在疾病诊断等精确分类任务中的应

用。Kim等^[11]从两个机构收集了318名患者的HRCT图像,比较了利用CNN进行深层学习和通过支持向量机(support vector machine,SVM)的浅层学习,结果表明CNN分类器的准确性性能明显要比SVM分类器高,随着卷积层的增加,CNN的分类效果得到了提升,但其数据数量有限,可能导致模型对某些特征过分依赖,从而影响其性能。Huang等^[12]使用公开的Kylberg纹理数据集(kylberg texture dataset,KTD),以DenseNet作为骨架网络,设计了小内核DenseNet(SK-DenseNet),并提出了一种新的两阶段迁移学习策略,其结果表明作者提出的CNN取得了良好的基准测试性能,并超过了当时最先进的CNN,然而该方法需要依赖足够的纹理源数据和从辅助未标记的肺部CT数据中学习到的知识,才能有效地应用于目标领域。Hwang等^[13]收集了246名患者的HRCT切片,使用CNN设计了一个基于内容的图像检索系统,其准确率可达81.7%。Pawar等^[14]使用公开的ILD数据集,开发了一种用于ILD分类的深度学习网络的两阶段混合方法,其准确率可达84.12%,但该方法需进行分割之后再进行分类,操作繁琐且准确率并不是很高。Nishikiori等^[15]使用私有慢性纤维化间质性肺疾病(chronic fibrosing interstitial lung diseases,CF-ILD)影像数据集,利用DenseNet-121进行特征提取,并结合WILDCAT网络进行判别,构建了一种深度学习算法,使用四折叠交叉验证方法来处理数据集,其检测灵敏度为89.6%,表明了该算法的检测效果不逊于医生。该网络采用WILDCAT网络作为判别层,在一定程度上增加了网络的复杂性。虽然这有助于提升模型的判别能力,但也带来了过拟合的风险,并可能会增加在实际应用中的部署难度。

上述分类方法虽然可以取得一定的效果,但是也存在一定的局限性。在面对复杂的分类问题时,网络对于上下文特征提取能力尚显不足,分类效果有待提升。

针对上述问题,本文提出了一种基于DenseNet-121的深度学习模型。该模型在DenseNet-121的基础上融合了多头自注意力机制以构建像素间长距离的依赖关系,使得模型能够同时关注多个重点区域。然后使用卷积注意力模块以实现更高效的特征提取和增强空间感知能力,从而提升网络的分类能力。最后添加改进空间金字塔池化层将多个不同尺度的特征图拼接起来,可以提取到更丰富的多尺度特征信息。此外,为了解决样本分类不均衡的问题,引入Focal Loss函数来减少易分类样本的权重,使得模型在训练时更专注于难分类的样本,从而进一步提升模型的分类能力。

1 方法

1.1 间质性肺炎HRCT图像分类网络

本文提出的DA-DenseNet(dual attention DenseNet)包括用于提取HRCT切片特征的骨架网络;聚合深层次特征的多头自注意力(multi-head self-attention, MH-

SA)模块;聚合通道与空间特征的卷积注意力模块(convolutional block attention module,CBAM);捕获更丰富空间信息的改进空间金字塔池化层(spatial pyramid pooling,SPP)。

1)DA-DenseNet

本文设计了 DA-DenseNet,以 DenseNet-121^[16] 作为主体结构,主要包括密集模块、过渡模块、MHSA 模块、CBAM 和改进 SPP 模块,网络结构如图 1 所示。使用 MHSA 模块捕捉图像中的全局特征和关系,通过将输入

进行特征提取和关系建模,可以更好地理解图像中的各种特征和它们之间的依赖关系。通过引入 CBAM,图像分类网络能够更好地获取图像中的关键特征和位置信息,从而提高网络的分类准确性和性能,可以实现更高效特征提取和增强空间感知能力。添加改进后的空间金字塔池化层不但可以接收不同大小的输入,通过池化操作将输入映射到预先设定的空间金字塔网格上,而且还可以捕获到更丰富的空间信息。这样可以减小输入的维度,同时保留重要特征,从而降低模型的复杂性和计算量,从而提升网络的分类能力。

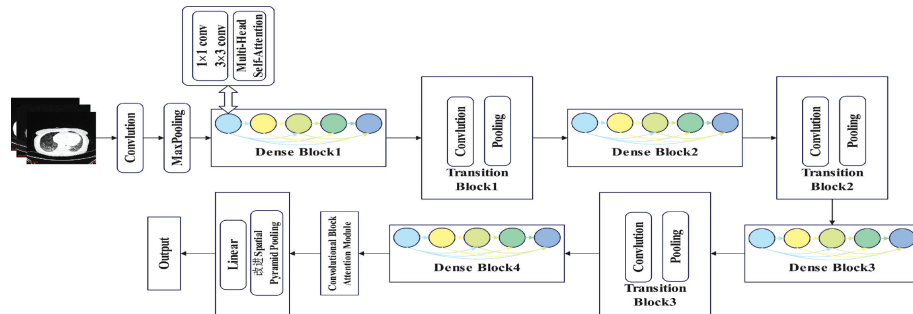


图 1 DA-DenseNet 的网络结构

Fig.1 Network structure of DA-DenseNet

DenseNet 是由 Huang 等^[16] 使用密集连接构建的,它的基本思路与 ResNet 一致,但具体的连接方式不同。ResNet 是通过元素级相加,而 DenseNet 每个层都会与前面所有层在会在通道维度上进行特征的拼接,实现了多次特征复用。DenseNet-121 由网络底端的卷积、4 个密集模块(dense block)和 3 个过渡模块(transition block)组成,

其中 Dense Block 重复了 6、12、24 和 16 次,并且和 Transition Block 以前馈方式交替串联组。Dense Block 将特征图级联来提高特征的正向传递效率;Transition Block 由 1×1 卷积和 2×2 平均池化层组成,将特征图大小和通道数减半,从而进一步降低模型复杂度。本文模型的具体结构如表 1 所示。

表 1 DA-DenseNet 的具体网络结构

Table 1 The specific network structure of DA-DenseNet

结构	DM-DenseNet	结构	DM-DenseNet
Input layer	7×7 Conv	Transition Block3	$\begin{bmatrix} 1 \times 1 \text{ Conv} \\ 2 \times 2 \text{ pool} \end{bmatrix}$
Pooling layer	3×3 pool	Dense Block4	$\begin{bmatrix} 1 \times 1 \text{ Conv} \\ 3 \times 3 \text{ Conv} \\ \text{MHSA} \end{bmatrix} \times 16$
Dense Block1	$\begin{bmatrix} 1 \times 1 \text{ Conv} \\ 3 \times 3 \text{ Conv} \\ \text{MHSA} \end{bmatrix} \times 6$	CBAM	—
Transition Block1	$\begin{bmatrix} 1 \times 1 \text{ Conv} \\ 2 \times 2 \text{ pool} \end{bmatrix}$	改进 SPP	$\begin{bmatrix} 5 \times 5, 9 \times 9, 13 \times 13 \text{ pool, stride} = 1 \\ 2 \times 2 \text{ pool, stride} = 2 \end{bmatrix}$
Dense Block2	$\begin{bmatrix} 1 \times 1 \text{ Conv} \\ 3 \times 3 \text{ Conv} \\ \text{MHSA} \end{bmatrix} \times 12$	Classifier	$\begin{bmatrix} 7 \times 7 \text{ pool} \\ \text{SoftMax} \end{bmatrix}$
Transition Block2	$\begin{bmatrix} 1 \times 1 \text{ Conv} \\ 2 \times 2 \text{ pool} \end{bmatrix}$		
Dense Block3	$\begin{bmatrix} 1 \times 1 \text{ Conv} \\ 3 \times 3 \text{ Conv} \\ \text{MHSA} \end{bmatrix} \times 24$		

2) 多头自注意力机制模块

间质性肺病图像具有复杂的纹理和形态,以及多样化的病变特征,这使得特征提取和分类任务变得非常具有挑战性。MHSA^[17]的提出可以有效地解决上述问题。MHSA通过关注输入序列中的不同位置,将输入序列中的每个位置关联起来,使输出与所有输入都有关^[18]。这有助于模型更全面地提取病变特征,为分类提供更丰富的信息。相较于其他注意力机制,MHSA的优势是通过多个注意力头的并行计算,能够从不同的角度和层次对输入数据进行特征提取和表示,增强了模型的表示能力,更好地处理复杂的间质性肺病分类问题。

(1) 自注意力机制

Zhao等^[19]和Li等^[20]分别提出了针对视频字幕的区域级注意力和帧级注意力。SENet^[21]通过全局平均池化层来理解不同通道的重要性,自动建立通道间的关系。Ding等^[22]设计了一个 patch attention 模块,以突出 feature map 的焦点区域。Panboonyuen等^[23]在GCN^[24]框架的每个阶段引入通道注意块,实现特征映射的分层优化。Su等^[25]关注小批图像中的相似对象,利用自注意力机制来编码它们之间的交互信息。

与上述方法不同的是,本文所用方法考虑了垂直和水平像素级的自注意力关注,更侧重于特征的内部相关性,其结构如图2所示。

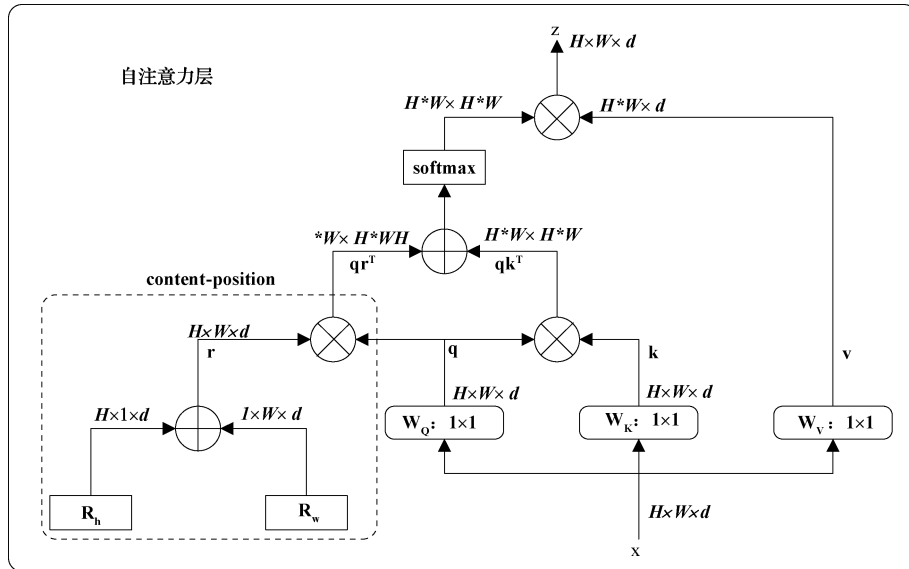


图2 自注意力的结构

Fig. 2 Structure of self-attention

注意力逻辑为 $qk^T + qr^T$, 其中 q, k, r, v 分别表示查询编码、键编码、位置编码和价值编码。 $\oplus$ 和 $\otimes$ 分别表示矩阵加法和矩阵乘法, 而 1×1 表示逐点卷积。给定大小为 $H \times W \times C$ 的输入特征图 X , 然后通过 (W_q, W_k, W_v) 将 X 转换为 $Q = XW_q, K = XW_k, V = XW_v$ 之后, 可以得到 K 和 Q 之间的局部相关性矩阵 $R \in R^{H \times W \times H \times W}$:

$$R = K \otimes Q \quad (1)$$

此外, content-position 这个部分是 MHSA 的位置编码部分, 与 Transformer^[26] 不一样的地方在于经过 reshape 操作后, 特征图会被转化为二维信息。与此相对应, 位置编码也是二维的, 而不是一维的。通过引入相对位置编码, 能够捕获特征图中垂直和水平方向上的相对信息:

$$\hat{R} = R + (R_h + R_w) \otimes Q \quad (2)$$

式中: R_h 和 R_w 分别代表高度和宽度的相对位置编码。通过将高度和宽度的位置编码分别初始化为可学习的参数向量 R_h 和 R_w , 并通过广播机制相加, 能够获取到一个更加

精确的相对位置编码。这种编码方式将原本 $H \times W \times d$ 的编码空间简化到了 $(H+W) \times d$, 大大减少了参数数量, 同时保留了位置信息的表达能力。接下来, 对每个通道维度上进行 Softmax 操作得到注意力矩阵 A :

$$A = \text{Softmax}(\hat{R}) \quad (3)$$

最后, 将 A 的每个空间位置的特征向量重塑为 d 个局部注意力矩阵, 得到输出特征图 $values$ 与局部注意力矩阵的聚合 Z 。

$$Z = A \otimes V \quad (4)$$

(2) 多头自注意力机制

自注意力机制就是通过某种运算来直接计算得到图像在编码过程中每个位置上的注意力权重; 然后再以权重和的形式来计算得到整个图像的隐含向量表示^[27]。而它的缺陷在于模型在对当前位置的信息进行编码时, 会过度的将注意力集中于自身的位置, 因此可以通过 MHSA 来解决这一问题。而 Transformer 结构的 MHSA 可以构建像素间长距离的依赖关系, 具有强大的全局信息提取能力

使得模型能够同时关注多个重点区域;而且它还为模型的注意力层提供了多个表示子空间。MHSA 将查询(Query)、键(Key)和值(Value)矩阵分别表示 Q 、 K 和 V , 这些矩阵的维度为 dk 且均是独立地随机初始化,然后将输入向量映射到不同子空间中,进而丰富了信息的特征表达。将权重矩阵输入到注意力函数中,注意力函数为:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{dk}}\right)V \quad (5)$$

MHSA 是在自注意力机制的基础上发展起来的,是自注意力机制的变体,旨在增强模型的表达能力和泛化能力。它通过使用多个独立的注意力头,分别计算注意力权重,并将它们的结果进行拼接或加权求和,从而获得更丰富的表示,其结构如图3所示。

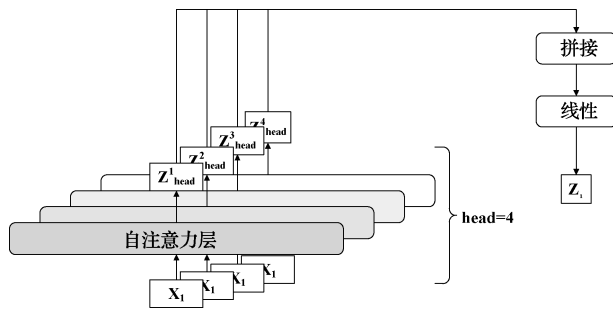


图3 多头自注意力的结构

Fig. 3 Structure of multi-head self-attention

可以得到 MHSA 的数学表达:

$$MulitHead(Q, K, V) = Concat(head_1, \dots, head_4)W_0 \quad (6)$$

式中: $head_1, \dots, head_4$ 表示输入经过 MHSA 后得到的输出; W_0 表示输出矩阵。通过多头 (本文 $head = 4$) 自注意力机制得到了 4 个输出 $z_1^{head}, z_2^{head}, z_3^{head}, z_4^{head}$, 为了获得与 x_1 对应的输出 z_1 , 在 MHSA 中将得到的 4 个输出进行拼接, 然后通过线性转换得到 z_1 , 对于图像中其他位置上的信息也是同样的处理方法。

在 MHSA 中, 输入数据首先经过多头自注意力机制的处理, 每个头独立计算注意力分数并对值进行加权求和, 得到 4 个不同的输出头。这些输出头随后沿着特征维度进行拼接, 形成一个融合了所有头信息的特征矩阵。紧接着, 该特征矩阵通过一个线性层进行转换, 以得到最终的输出矩阵。在整个过程中, 所有位置的输入数据都共享相同的网络参数, 确保模型能够学习跨位置的通用特征表示。要获取特定位置的输出, 只需在输出矩阵中定位该位置并提取相应特征向量。通过这种方式, MHSA 能够捕获输入数据中的复杂依赖关系, 并提升模型的表达能力和性能。

3) 卷积注意力模块

针对传统的 CNN 通常只关注数据的特征信息而忽略了上下文信息的缺陷, 提出了 CBAM^[28], 该模块能够利用卷积运算有效提取图像中的空间特征, 这对于医学影像分

析至关重要。它同时包含通道注意力机制和空间注意力机制, 不但使网络能够关注不同通道的特征图, 更加精细地处理各种医学信息, 而且能够关注图像中的关键区域, 提高病变区域的识别准确性。将 CBAM 引入该网络, 该模块可以通过全局和局部的上下文信息来综合考虑, 从而更好地理解图像内容, 从而提高网络的识别准确性和鲁棒性。

CBAM 是一种用于前馈卷积神经网络的简单而有效的注意力模块。整体结构如图4所示, 可以看到 CBAM 包含两个独立的子模块, 通道注意力模块(channel attention module, CAM)和空间注意力模块(spatial attention module, SAM), 分别在通道与空间上生成注意图。这样不只能够节约参数和计算力, 并且保证了其能够做为即插即用的模块集成到现有的网络架构中去^[29]。

首先, 将中间层的特征图作为输入, 通过 CAM 进行处理。该模块利用全局平均池化和全局最大池化操作提取通道信息, 并通过共享多层感知机学习通道间的依赖关系, 最终生成通道注意力权重图。这些权重与原始特征图进行逐通道的乘法操作, 实现通道注意力的加权。接着, 加权后的特征图进入 SAM 该模块通过通道池化操作生成空间描述子, 并利用卷积层进行特征融合, 得到空间注意力权重矩阵。最后, 这个权重矩阵与特征图进行逐元素的乘法操作, 实现空间注意力的加权。经过 CBAM 模块处理后, 得到的特征图同时融合了通道和空间注意力信息, 有助于提升网络的性能。

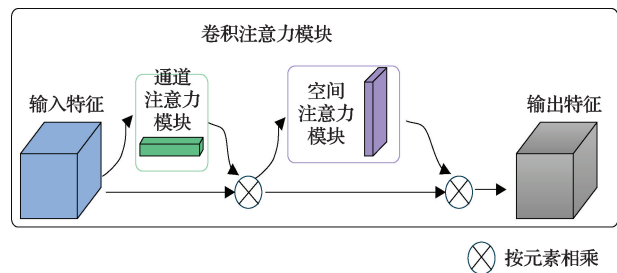


图4 卷积注意力模块的结构

Fig. 4 Structure of convolutional block attention module

CAM 如图5(a)所示, 假设输入特征图的通道数为 C , 分别采用最大池化层和平均池化层对输入特征图进行处理, 获得两个尺寸为 $1 \times 1 \times C$ 的特征向量, 可以分别记为 F_{avg}^c 和 F_{max}^c , 然后将它们分别送入一个共享的多层感知器 (multi-layer perceptron, MLP) 中进行计算。然后将计算结果对应相加并且经过 sigmoid 激活层得到权重系数 M_c 。最后, 将 M_c 和输入特征图 F 做 element-wise 乘法操作, 生成 Spatial attention 模块需要的输入特征。假设 MLP 两层的权重系数分别表示为 W_0 和 W_1 , 则:

$$M_c(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) = \sigma(W_1(W_0(F_{avg}^c)) + W_1(W_0(F_{max}^c))) \quad (7)$$

式中: σ 表示 sigmoid 函数。

SAM 利用特征间的空间关系生成空间注意力图,与通道注意力不同,空间注意力关注的是“哪里”是一个有信息的部分,这是对通道注意力的补充。其结构如图 5(b) 所示,在通道维度分别进行了平均值池化和最大值池化。AvgPool 的操作就是在通道上提取平均值,提取到的值记为 F'_{avg} ; MaxPool 的操作就是在通道上提取最大值,提取到的值记为 F'_{max} ; 接着将前面所提取到的特征图(通道数都为 1)合并得到一个 2 通道的特征图。将它们按照通道拼接在一起,使用一个 7×7 卷积核处理,而后经过 sigmoid 激活函数得到权重系数 M_s ,然后将输入特征图与权重相乘,即可得到注意力分配之后的特征图。表达式为:

$$M_s(F) = \sigma(f^{7 \times 7}([AvgPool(F); MaxPool(F)])) = \sigma(f^{7 \times 7}([F'_{avg}; F'_{max}])) \quad (8)$$

式中: σ 表示 sigmoid 函数; $f^{7 \times 7}$ 表示一个滤波器大小为 7×7 的卷积运算。

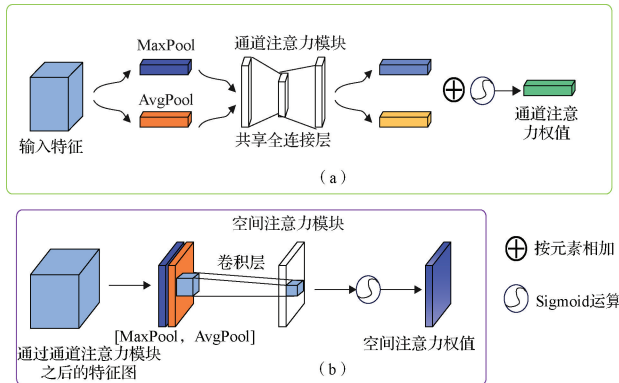


图 5 通道注意力模块与空间注意力模块的结构

Fig. 5 The structure of the channel attention module and the spatial attention module

4) 改进空间金字塔池化层

针对间质性肺病图像中的病变区域的大小不同以及形状各异的问题,采用 SPP 来提取空间多尺度信息。SPP 是何凯明等^[30]提出来的,其主要功能是将不同尺寸的输入转化为固定尺寸的输出。通过将输入图像划分为多个不同尺度的空间区域,并对每个区域进行池化操作,SPP 能够同时提取出局部和全局的特征,使得网络能够更好地适应不同大小和形状的物体或病变区域。原本的 SPP 使用了多尺度特征提取,通过在不同尺度的空间金字塔上执行池化操作,从而捕获到不同尺度的特征信息。然而通过不同尺度的池化操作将特征图划分为多个子区域,并取每个子区域的最大值或平均值作为输出。这种操作可能会丢失一些重要的细节信息,导致信息损失。因此本文使用更大尺度的 MaxPool 层与不同尺度的 MaxPool 层相结合,这使得模型能够更有效地从输入特征图中提取多尺度的特征。网络能够扩大感受野以捕获更广泛的特征信

息,同时减小特征图尺寸以减少计算量,从而优化图像复杂结构和模式的识别能力。而且通过计算隐藏通道数与池化层数量的乘积,动态地调整了拼接后的特征图的通道数,以适应多尺度特征图的拼接。改进后的 SPP 结构如图 6 所示。

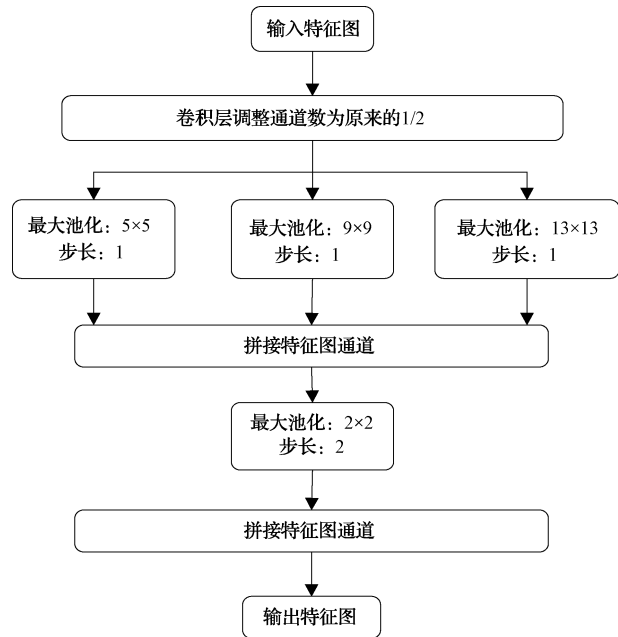


图 6 改进空间金字塔池化结构

Fig. 6 Improved spatial pyramid pooling structure

具体操作为最后一层卷积层和全连接层前加入改进的 SPP 结构,对通过最后一层卷积的多层特征图进行通道数的调整,将调整后的特征图按 5×5 、 9×9 以及 13×13 不同大小划分为不同尺寸的网格,然后对每个网格进行最大池化操作,将得到的结果进行拼接。接着将拼接后的特征图通过一个更大的 MaxPool 层进行池化操作,这个 MaxPool 层的核大小为 2,步长为 2。从而使得模型能够同时处理和利用不同尺度的特征信息。

1.2 数据集

本文数据集使用昆明医科大学提供的 HRCT 图像数据集(涵盖了昆明医科大学第一附属医院、云南省第一人民医院、昆明市第一人民医院甘美医院和昆明市延安医院等多家医院的数据),该数据集包括 3 种不同疾病的患者和一类正常对照组。有 4 类数据,寻常型间质性肺炎(usual interstitial pneumonia, UIP)、非特异性间质性肺炎(nonspecific interstitial pneumonia, NSIP)、机化性肺炎(organizing pneumonia, OP)和正常对照组(Normal)。该数据集包括 180 名患者的 26 918 张肺部图像。

本文根据患者数按 6:2:2 划分数据集,80%用于训练和验证,20%用于测试,同时对每类的切片数进行微调,以保证每类切片的比例也为 6:2:2,数据集划分及分布结果如表 2 所示。

表2 数据集划分(切片/患者)

Table 2 Data set division and distribution(slices/patients)

	UIP	NSIP	OP	Normal	Total
训练集	2 680/23	3 655/30	3 424/26	6 324/30	16 083/109
验证集	954/8	1 251/10	1 142/9	2 107/10	5 454/37
测试集	907/7	1 226/9	1 140/8	2 108/10	5 381/34
Total	4 541/38	6 132/49	5 706/43	10 539/50	26 918/180

1.3 Focal Loss 函数

针对多分类任务中因类别样本数量不均衡而导致模型性能下降的问题,本文选用 Focal Loss 作为损失函数来处理这个问题。该函数最初是由何恺明等^[31]提出来的,其核心思想是基于交叉熵损失函数的改进。在目标检测等计算机视觉任务中,由于正负样本不平衡的问题,传统的交叉熵损失函数往往无法有效地对模型进行训练。而 Focal Loss 的设计初衷就是为了解决这一问题。它通过调整损失的权重,使得模型在训练过程中能够更加关注那些难以分类的样本,从而提高了模型的泛化能力。其计算公式如下:

$$Lce(p_i) = -\log(p_i) \quad (9)$$

式中: p_i 为预测概率大小, Focal Loss 在交叉熵损失函数的基础上引入了权重系数 α , 旨在平衡正负样本的权重, 使用调制因子 $(1 - p_i)^\gamma$ ($\gamma > 0$) 调控易分类样本和难分类样本的比重, 综合上述两者就得到了 focal loss, 其计算式如下:

$$L_{fl} = -\alpha(1 - p_i)^\gamma \log(p_i) \quad (10)$$

式中: p_i 反应了与类别的接近程度; p_i 越大说明分类越准确。这个损失函数不仅考虑了“容易分辨”, 还考虑了“正负样本”的问题。因此本文采用该方法来解决样本分类不均衡的问题。

2 实验结果与分析

2.1 评价指标

由于本文是多分类任务, 只选取准确率作为评价指标会显得过于单一, 因此选取准确率(accuracy)、精确率(precision)、召回率(recall)、F1 分数、混淆矩阵(confusion matrix)和 Kappa 系数作为评价指标衡量模型性能。其中 Kappa 系数适用于衡量数据不均衡的情况下模型的偏向性, Kappa 系数越大代表模型的预测结果和标签的一致性越高, 模型的偏向性也就越小, 其稳健性也会有所提高, 定义如下。

准确率: 对于给定的测试数据集, 分类器正确分类的样本数与总样本数之比。

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (11)$$

精确率: 指分类正确的正样本个数占分类器判定为正样本的样本个数的比例。

$$Precision = \frac{TP}{TP + FP} \quad (12)$$

召回率是指分类正确的正样本个数占真正的正样本个数的比例。

$$Recall = \frac{TP}{TP + FN} \quad (13)$$

F1-Score 值就是精确率和召回率的调和平均值。

$$F1_Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (14)$$

Kappa 系数是一个用于一致性检验的指标, 也可以用于衡量分类的效果。

$$Kappa = \frac{p_0 - p_e}{1 - p_e} \quad (15)$$

$$p_e = \frac{\sum_{k=1}^n (\sum_{i=1}^n x_{ik} \times \sum_{j=1}^n x_{kj})}{(\sum_{j=1}^n \sum_{i=1}^n x_{ij})^2} \quad (16)$$

式中: TP(true positive)表示把标签正预测为正; FP(false positive)为标签负预测为正; TN(true negative)为标签负预测为负; FN(false negative)表示把标签正预测为负; p_0 代表的是准确率; n 代表分类的类别; p_e 代表混淆矩阵中所有类别分别对应的实际与预测数量的乘积的总和; 除以样本总数的平方; x_{ij} 表示的是混淆矩阵中的元素。

2.2 实验平台与模型训练

本文实验均在 Windows10 系统和 Spyder 编辑器下搭建, 在 Intel(R) Core(TM) i9-9900K CPU、NVIDIA GeForce RTX2080Ti GPU 和 32 G 内存平台进行加速计算, 使用 Pytorch1.12.1 搭建网络模型及编程语言 Python 3.8 实现。本文实验使用 Adam 优化器, 用于计算和更新优化模型训练的网络参数, 并将学习率与学习率衰减(learning rate decay)设置为 0.001, 模型的批量大小(batch size)设置为 16, 输入图像尺寸设置为 224×224 , 一共训练 50 个轮次。

2.3 基础网络骨架对比

为选择合适的骨架网络, 本文对 RegNet、EfficientNetV2、ConvNet、Shufflenet 和 DenseNet121 五种基础网络进行 HRCT 图像分类实验并记录其评价指标, 它们的网络性能如表 3 所示。

表3 基础网络性能

Table 3 Performance of the baseline network (%)

网络模型	Accuracy	Recall	Precision	F1	Kappa
RegNet	83.59	72.09	79.55	75.64	76.67
EfficientNetV2	82.50	77.54	80.36	78.92	75.58
ConvNet	82.13	78.49	78.23	78.36	75.23
Shufflenet	80.99	64.75	79.43	71.34	72.59
DenseNet121	83.63	80.67	79.62	80.14	77.38

对比表 3 结果可知, DenseNet121 网络在切片级别上

的综合评价指标表现最佳。RegNet 注重网络设计的规则性和简洁性,但可能在特征重用方面稍逊于 DenseNet121。EfficientNetV2 追求高效的性能和资源利用率,但在特定任务上可能不如 DenseNet121 适应性强。ConvNet 作为经典的网络结构,其性能稳定但可能缺乏 DenseNet121 那样的创新特性。Shufflenet 则专注于轻量级和移动端应用,可能在计算资源受限的情况下更具优势,但在性能上可能无法与 DenseNet121 相匹敌。因此选取 DenseNet121 作为骨架网络。

2.4 消融实验与注意力机制对比实验

为了验证本文所提 DA-DenseNet 有效性,本文对 ILD 的 HRCT 图像分类中的各个模块和训练策略进行消融实验对比;而后将常用的注意力机制与本文所提出的双

重注意力机制进行对比,并利用梯度加权类激活映射 (gradient-weighted class activation mapping, Grad-CAM) 绘制模型的热力图,进一步验证本文方法的有效性。

1)消融实验

DA-DenseNet 由骨架网络、MHSA 模块、CBAM、改进 SPP 层和针对样本分类不均衡的训练策略五部分组成。通过实验选取了 DenseNet121 作为骨架网络,对其他 2 个部分进行消融实验对比,以验证所提方法的有效性。消融实验所使用数据均使用了相同的数据处理操作,以保证实验的公正性,结果如表 4 所示。其中包含了评估分类性能的常用指标,如准确度、召回率、精确度、F1 分数和 Kappa 系数。这些指标从不同的角度评估了模型的性能,有助于全面理解每个组件对模型性能的影响。

表 4 消融实验的结果
Table 4 Results of ablation experiments (%)

主干	MHSA	CBAM	改进 SPP 层	策略	Accuracy	Recall	Precision	F1	Kappa
✓					83.63	80.67	79.62	80.14	77.38
✓	✓				85.93	83.19	82.87	83.03	80.51
✓	✓	✓			86.32	83.46	83.64	83.55	81.06
✓	✓	✓	✓		87.12	84.19	84.45	84.32	82.15
✓	✓	✓	✓	✓	88.28	85.75	85.44	85.59	83.76

从表 4 可以看出,将 MHSA 模块融入 DenseNet121 网络当中,切片的准确率提高了 2.3%,而后将 CBAM 融入网络,提高了 0.39%,然后加入改进 SPP 层,其准确率提高了 0.8%,最后引入 Focal Loss 损失函数,提高了 1.16%。结果表明,本文的方法在准确率、召回率、精确

率、F1 分数和 Kappa 系数上都取得了最佳结果。

2)注意力机制对比

为了验证本文研究加入的 MHSA 模块和 CBAM 的有效性,本文选取 SE 和 ECA,并将它们融入网络进行实验,结果如表 5 所示。

表 5 不同注意力机制的性能
Table 5 Performance of different attention mechanisms (%)

Backbone	Attention	Accuracy	Recall	Precision	F1	Kappa
Densenet121	None	84.34	80.32	80.95	80.63	78.23
	SE	84.95	81.18	81.75	81.46	79.10
	ECA	85.84	83.14	82.97	83.05	80.41
	本文模型	88.28	85.75	85.44	85.59	83.76

将对比实验的混淆矩阵如图 7 所示,可以看出,模型对测试集的分类结果,其中行是真实标签,列是预测标签,混淆矩阵能直观地观察出算法是否将不同的类别混淆在一起。

本文使用 Grad-CAM 绘制出模型在决策的热力图进行网络的可视化,以了解 HRCT 切片的哪些区域让模型作出决策,热力图如图 8 所示。从图 8 可以看出,原始 DenseNet-121 仅关注到肺部的部分病变区域,对整体肺部的关注不够充分;而插入 SE 模块的网络在关注肺部区域的同时也学习到其他不相关部分,使得提升幅度不大,ECA 模块关注到的病变区域比较少,这就

留下了改进的空间;而融合了 MHSA 模块与 CBAM 之后,整个网络关注集中于肺部的病灶区域,而很少关注其余部分,说明融合 MHSA 模块与 CBAM 后的网络更加关注 HRCT 切片的关键病灶区域,聚合的特征图得到了充分的细化。

2.5 与其他 ILD 的图像分类方法对比

表 6 为本文模型以及一些现有研究的 ILD 分类准确性对比。根据现有文献,使用 HRCT 图像数据集的最佳分类准确率为 85.00%,CT 图像数据集为 88.0%。在从 HRCT 数据集中识别 ILD 时,本文方法在测试准确性方面优于以前的其他研究,得分为 88.28%。

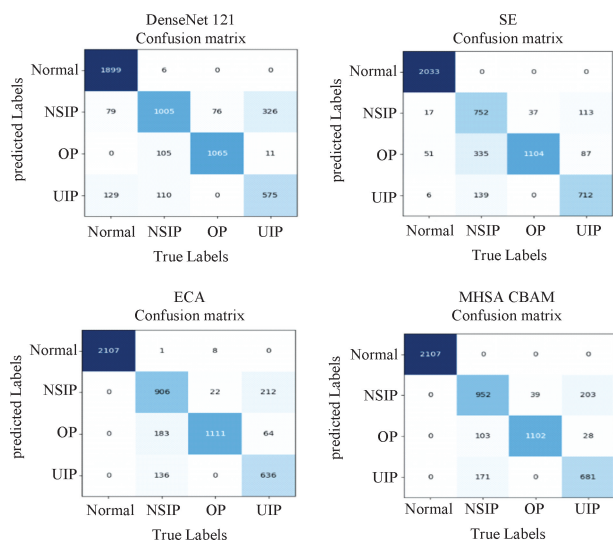


图7 不同注意力机制在切片级别的混淆矩阵
Fig. 7 Confusion matrix for different attention mechanisms at the slice level

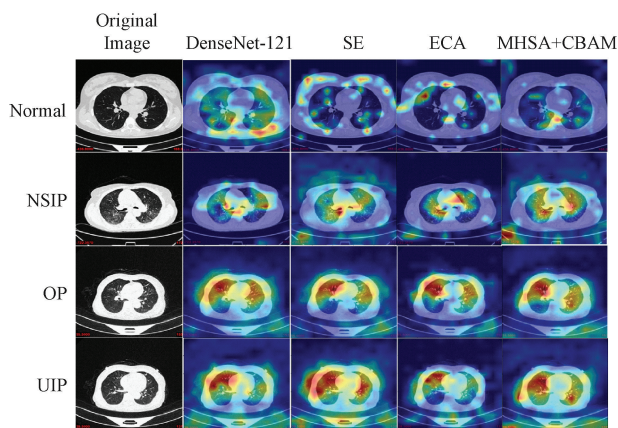


图8 不同注意力机制的热力图
Fig. 8 Heatmap of different attention mechanisms

表 6 不同方法对比			
Table 6 Comparison of different methods			
模型	Image	Model	Accuracy/%
文献[9]	HRCT	CNN	76.40
文献[13]	HRCT	CNN	81.70
文献[14]	HRCT	CNN	84.12
文献[10]	HRCT	CNN	85.00
文献[8]	CT	CNN	88.00
本文	HRCT	CNN	88.28

3 结 论

本文的主要研究目标是开发一种深度学习模型,该模型能够通过对 HRCT 图像的分析,实现对 ILD 类别的准确分类。对比当前已有的对于 ILD 的 HRCT 切片分类方

法,本文所提方法较原始 DenseNet-121 在准确率、召回率、精确率、F1 分数和 Kappa 系数提高了 4.65%、5.08%、5.82%、5.45%和 6.38%。结果表明,对比其他方法,本文提出的基于双重注意力机制的间质性肺病高分辨率 CT 图像分类方法充分提取了切片的病灶特征,在未经训练的数据上实现了 88.28%的测试准确率。在相较其他方法精度更高,并且具有良好的泛化性。未来的工作,将在不降低网络模型性能的情况下减少模型参数的方向展开研究,进一步提升网络模型的综合性能。

参 考 文 献

[1] WIJSENBEEK M, SUZUKI A, MAHER T M. Interstitial lung diseases [J]. Lancet, 2022, 400 (10354):769-786.

[2] 梁艳霞,樊芙蓉,王冰,等. 结缔组织病相关间质性肺病中医病因病机及治疗研究进展[J]. 中国医药导报, 2022,19(15):33-36.

LIANG Y X, FAN M R, WANG B, et al. Research progress on the etiology, pathogenesis and treatment of connective tissue disease-related interstitial lung disease in traditional Chinese medicine [J]. China Medicine Herald, 2022,19(15):33-36.

[3] 白祥梅,席磊,杨柳,等. 自身免疫特征间质性肺病研究进展[J]. 临床医学进展, 2023, 13(3): 3752-3758.

BAI X M, XI L, YANG L, et al. Research progress on interstitial lung disease with autoimmune characteristics [J]. Advances in Clinical Medicine, 2023, 13(3): 3752-3758.

[4] 李子茂,于舒,郑禄,等. 基于融合注意力机制的小样本遥感场景分类方法[J]. 国外电子测量技术, 2023, 42(7):59-67.

LI Z M, YU SH, ZHENG L, et al. Small-sample remote sensing scene classification method based on fused attention mechanism [J]. Foreign Electronic Measurement Technology, 2023, 42(7):59-67.

[5] 杨昆,杨月婷,路宇飞,等. 多光谱视网膜成像动静脉自动分割方法[J]. 电子测量技术, 2023, 46(10):84-91.

YANG K, YANG Y T, LU Y F, et al. Automatic arteriovenous segmentation method for multispectral retinal imaging [J]. Electronic Measurement Technology, 2023, 46(10):84-91.

[6] 林桢哲,王桂棠,陈建强,等. 基于残差网络深度学习的肺部 CT 图像结节良恶性分类模型[J]. 仪器仪表学报, 2020, 41(3):248-256.

LIN ZH ZH, WANG G T, CHEN J Q, et al. Benign and malignant classification model of lung CT image nodules based on deep learning of residual network[J]. Chinese Journal of Scientific Instrument,

- 2020,41(3):248-256.
- [7] GAO M, BAGCI U, LU L, et al. Holistic classification of CT attenuation patterns for interstitial lung diseases via deep convolutional neural networks [J]. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, 2018, 6(1): 1-6.
 - [8] KOO C W, WILLIAMS J M, LIU G, et al. Quantitative CT and machine learning classification of fibrotic interstitial lung diseases [J]. European Radiology, 2022,32(12):8152-8161.
 - [9] WALSH S L F, CALANDRIELLO L, SILVA M, et al. Deep learning for classifying fibrotic lung disease on high-resolution computed tomography: A case-cohort study [J]. Lancet Respiratory Medicine, 2018, 6(11): 837-845.
 - [10] CHOE J, HWANG H J, SEO J B, et al. Content-based image retrieval by using deep learning for interstitial lung disease diagnosis with chest CT[J]. Radiology, 2022,302(1):187-197.
 - [11] KIM G B, JUNG K H, LEE Y, et al. Comparison of shallow and deep learning methods on classifying the regional pattern of diffuse lung disease [J]. Journal of Digital Imaging, 2018, 31(4): 415-424.
 - [12] HUANG S, LEE F, MIAO R, et al. A deep convolutional neural network architecture for interstitial lung disease pattern classification [J]. Med Biol Eng Comput, 2020, 58(4): 725-37.
 - [13] HWANG H J, SEO J B, LEE S M, et al. Content-based image retrieval of chest ct with convolutional neural network for diffuse interstitial lung disease: Performance assessment in three major idiopathic interstitial pneumonias [J]. Korean Journal of Radiology, 2021, 22(2): 281-90.
 - [14] PAWAR SP, TALBAR SN. Two-stage hybrid approach of deep learning networks for interstitial lung disease classification [J]. Biomed Research International, 2022,DOI: 10.1155/2022/7340902.
 - [15] NISHIKIORI H, KURONUMA K, HIROTA K, et al. Deep-learning algorithm to detect fibrosing interstitial lung disease on chest radiographs [J]. European Respiratory Journal, 2023, 61 (2): 2102269.
 - [16] HUANG G, LIU Z, MAATEN L V D, et al. Densely connected convolutional networks [C]. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.
 - [17] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [J]. ArXiv preprint ArXiv: 1706.03762, 2017.
 - [18] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words; Transformers for image recognition at scale [J]. ArXiv preprint ArXiv:2010.11929,2020.
 - [19] ZHAO B, LI X, LU X. CAM-RNN: Co-attention model based RNN for video captioning [J]. IEEE Transaction on Image Processing, 2019, 28 (11): 5552-5565.
 - [20] LI X, ZHAO B, LU X. MAM-RNN: Multi-level attention model based RNN for video captioning[C]. Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, 2017.
 - [21] HU J, SHEN L, ALBANIE S, et al. Squeeze-and-excitation networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(8): 2011-2023.
 - [22] DING L, TANG H, BRUZZONE L. LANet: Local attention embedding to improve the semantic segmentation of remote sensing images [J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 59(1): 426-35.
 - [23] PANBOONYUEN T, JITKAJORNWANICH K, LAWAWIROJWONG S, et al. Semantic segmentation on remotely-sensed images using enhanced global convolutional network with channel attention and domain specific transfer learning [J]. Remote Sensing, 2019, DOI: 10.20944/preprints201812.0090.v2.
 - [24] PENG C, ZHANG X, YU G, et al. Large kernel matters: Improve semantic segmentation by global convolutional network[C]. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.
 - [25] SU Y, WU Y, WANG M, et al. Semantic segmentation of high resolution remote sensing image based on batch-attention mechanism[C]. Proceedings of the IGARSS 2019 — 2019 IEEE International Geoscience and Remote Sensing Symposium, 2019.
 - [26] LIU Z, LIN Y, CAO Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows [J]. ArXiv preprint arXiv: 2103.14030, 2021.
 - [27] SRINIVAS A, LIN T Y, PARMAR N, et al. Bottleneck transformers for visual recognition[C]. Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021.

- [28] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional Block Attention Module [J]. ArXiv preprint arXiv:1807.06521, 2018.
- [29] 彭道刚,潘俊臻,王丹豪,等. 基于改进 YOLO v5 的电厂管道油液泄漏检测[J]. 电子测量与仪器学报, 2022,36(12):200-209.
- PENG D G, PAN J ZH, WANG D H, et al. Oil leakage detection of power plant pipeline based on improved YOLO v5 [J]. Journal of Electronic Measurement and Instrumentation, 2022, 36(12): 200-209.
- [30] HE K M, ZHANG X Y, REN S Q, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2015, 37(9): 1904-1916.
- [31] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal

loss for dense object detection [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2020, 42(2): 318-327.

作者简介

赵琪玉, 硕士研究生, 主要研究方向为医学图像处理。

E-mail: 1982124430@qq.com

张俊华(通信作者), 教授, 博士生导师, 主要研究方向为计算机视觉和医学图像处理。

E-mail: jhzhang@ynu.edu.cn

张剑青, 教授, 博士生导师, 主要研究方向为间质性肺病、肺癌和肺部感染性疾病。

E-mail: ydyzjq@163.com

徐铭蔚, 硕士研究生, 主要研究方向为临床医学。

E-mail: 147521202@qq.com