

# Dual-Path Vision Transformer 用于急性 缺血性脑卒中辅助诊断



张桃红<sup>1</sup>, 郭学强<sup>1</sup>, 郑瀚<sup>2\*</sup>, 罗继昌<sup>3</sup>, 王韬<sup>3</sup>, 焦力群<sup>3</sup>, 唐安莹<sup>1</sup>

(1. 北京科技大学 计算机与通信工程学院, 北京 100083; 2. 河池学院 广西高校人工智能与信息处理重点实验室, 广西 546300;  
3. 首都医科大学 宣武医院, 北京 100053)

**摘要** 急性缺血性脑卒中是由于脑组织血液供应障碍导致的脑功能障碍, 数字减影脑血管造影 (DSA) 是诊断脑血管疾病的金标准。基于患者的正面和侧面 DSA 图像, 对急性缺血性脑卒中的治疗效果进行分级评估, 构建基于 Vision Transformer 的双路径图像分类智能模型 DPVF。为了提高辅助诊断速度, 基于 EdgeViT 的轻量化设计思想进行了模型的构建; 为了使模型保持轻量化的同时具有较高的精度, 提出空间-通道自注意力模块, 促进 Transformer 模型捕获更全面的特征信息, 提高模型的表达能力; 此外, 对于 DPVF 的两分支的特征融合, 构建交叉注意力模块对两分支输出进行交叉融合, 促使模型提取更丰富的特征, 从而提高模型表现。实验结果显示 DPVF 在测试集上的准确率达 98.5%, 满足实际需求。

**关键词** 急性缺血性脑卒中; 视觉 Transformer; 双分支网络; 特征融合

中图分类号 TP391.41; TN911.73 文献标志码 A DOI 10.12178/1001-0548.2023081

## Dual-Path Vision Transformer for Auxiliary Diagnosis of Acute Ischemic Stroke

ZHANG Taohong<sup>1</sup>, GUO Xueqiang<sup>1</sup>, ZHENG Han<sup>2\*</sup>, LUO Jichang<sup>3</sup>, WANG Tao<sup>3</sup>,  
JIAO Liqun<sup>3</sup>, and TANG Anying<sup>1</sup>

(1. School of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing 100083, China;  
2. Key Laboratory of AI and Information Processing, Hechi University, Guangxi 546300, China;  
3. Xuanwu Hospital, Capital Medical University, Beijing 100053, China)

**Abstract** Acute ischemic stroke is one of the fatal brain dysfunction diseases caused by the interruption of blood supply to the brain tissue. Digital Subtract Angiography (DSA) is the gold standard for diagnosing such cerebrovascular diseases. Based on the frontal and lateral DSA images of the patients, a dual-path image classification intelligent model, Dual-Path Vision Transformer (DPVF), is constructed in this paper to evaluate the treatment effectiveness of acute ischemic stroke in a graded manner. In order to improve the speed of auxiliary diagnosis, the model is constructed based on the lightweight design idea of EdgeViT. And in order to make the model have high accuracy, the spatial-channel self-attention module is proposed to promote the transformer model to capture more comprehensive feature information and improve the model representation. In addition, for the feature fusion of two branches of DPVF, a cross-attention module is constructed to cross-fuse the outputs of the two branches, which promotes the model to extract richer features and thus improves the model performance. The experimental results show that the accuracy of DPVF on the test set reaches 98.5%, which can effectively meet the practical requirements.

**Key words** acute ischemic stroke; vision Transformer; dual-path; feature fusion

静脉溶栓可以一定程度上开通闭塞血管, 恢复血流灌注, 是治疗急性缺血性脑卒中 (Acute Ischemic Stroke, AIS) 的有效方式。对静脉溶栓治疗效果的评估常常需要借助 X 射线数字减影血管造影 (Digital Subtraction Angiography, DSA) 成

像, DSA 成像是诊断脑血管疾病的重要方法, 它的基本原理是将造影前后拍摄的 X 射线图像进行减影, 以消除血管造影影像上的骨骼和软组织结构, 从而获得清晰的血管影像。在获得 DSA 图像后, 医生可以基于图像对 AIS 治疗后的再灌注程度

收稿日期: 2023-03-21; 修回日期: 2023-05-15

基金项目: 科技部科技创新 2030-重大项目 (2020AAA0108703); 广西高校人工智能与信息处理重点实验室基金 (2022GXZDSY001)

作者简介: 张桃红, 博士, 教授, 主要从事机器视觉、深度学习方面的研究。

\*通信作者 E-mail: 05057@hcnu.edu.cn

进行 mTICI 评分。mTICI 评分根据血管再通程度分为 5 级，分别为 0 级、1 级、2a 级、2b 级和 3 级。为了分级更准确，经常采用正面和侧面的 DSA 显影图像对以获取更充分的信息。然而，对 DSA 图像的识别、诊断和分级工作通常是由专业的医生来完成。近年来，随着人工智能、深度学习的快速发展，使用计算机辅助诊断可以显著提高诊断效率<sup>[1]</sup>。其中，基于深度学习的图像分类是计算机辅助诊断的常用方法，将医疗图像作为输入，通过训练好的模型对其进行预测，输出病患病情进行智能辅助诊断。

在脑卒中辅助智能诊断模型研究中，文献 [2] 提出了一种基于视频的卒中损伤评估系统，使用 Mask R-CNN<sup>[3]</sup>、级联金字塔网络和时域卷积网络模型实现了自动评分。文献 [4] 基于血管造影参数成像 (Angiographic Parametric Imaging, API) 图，设计了一个能够自动评估机械血栓切除术 (Mechanical Thrombectomy, MT) 过程中神经血管的再灌注情况的卷积神经网络，对血管是否再通成功的预测准确率达 81%。文献 [5] 基于卒中患者核磁共振图像 (Magnetic Resonance Imaging, MRI) 研究了一种基于深度学习和机器学习的混合方法，用于预测患者的语言障碍严重程度，使用 CNN 的高级特征和主成分分析 (PCA) 的图像特征作为岭回归的输入，实现了比仅使用深度学习或机器学习模型更好的性能。文献 [6] 基于医疗服务使用和健康行为数据，利用深度神经网络和 PCA 预测患者卒中的概率，AUC 值达 83.48%，对于具有较高卒中风险患者的早期发现具有重要意义。文献 [7] 使用卷积神经网络进行了急性缺血性卒中患者组织病变体积的预测，以便于医生根据患者病变体积制定科学的治疗方案，表明了使用深度卷积神经网络对卒中患者组织形态和治疗效果预测的有效性。文献 [8] 使用集成网络结合多个平面的 API 图来评估再灌注水平，使用 CNN 将 API 图分类为充分/不充分的再灌注；对于模型的输出，采用网格搜索算法对每个网络输出进行加权，结果表明使用来自多个视图的模型评估再灌注水平比使用单一视图更有效。文献 [9] 提出了一种基于 CNN 的全自动的定量 TICI 评分算法 autoTICI，首先，利用多路径卷积神经网络将每个 DSA 图像序列划分为 4 个时期，分别为非对比度期、动脉期、实质造影期以及静脉期；其次，使用运动校正的动脉期和实质造影

期的图像序列计算最小强度图，在最小强度图上，分割血管、灌注和背景；最后将 autoTICI 评分量化为治疗后的再灌注像素比率，实现对再灌注水平的定量分析。

以上研究表明：当前深度学习模型应用于 AIS 辅助影像智能诊断的研究工作主要基于 CNN，且处理的大多是单面影像；文献 [9] 提出的模型可以同时处理正面和侧面影像，但两个视频流输入模型不仅导致数据处理量大，而且无法适应不同成像设备导致的视频流规格不一致问题，此外，提出的 autoTICI 的定量分析方法具有 4 个阶段，无法实现端到端训练。且 CNN 模型感受野较小，难以捕获图像全局特征。为了获得图像全局信息，并结合临床需要的正、侧面图像结合诊断，本文设计了一种基于 Transformer 的双路径图像分类模型 Dual-Path Vision Transformer (DPVF) 用于 AIS 辅助诊断，模型的两个路径分别用于提取患者正面和侧面 DSA 图像的信息特征。

## 1 DPVF 模型结构

本研究采用 mTICI 评分的正面和侧面 DSA 图像，如图 1 所示，0 级表示无灌注，血管闭塞远端无顺向血流；1 级表示渗透性灌注，远端分支灌注有限；2a 级表示前向血流部分灌注小于一半下游缺血区；2b 级表示前向血流部分灌注大于一半下游缺血区；3 级表示前向血流完全灌注下游缺血区。当 mTICI 评分为 2b 级或 3 级时，表示患者血管再通成功。

本文提出的双路径视觉 Transformer 用于急性缺血性脑卒中辅助诊断模型 DPVF 的整体结构如图 2 所示。DPVF 每一路径均基于轻量化模型 EdgeViT<sup>[10]</sup> 进行设计，为了提升轻量化模型分类精度，重新设计 EdgeViT 的多头自注意力模块为稀疏空间-通道自注意力模块 (Sparse Dual Attention)；为了融合两路径提取的正面和侧面图像信息，构建了交叉注意力模块 (Cross Attention) 以实现两路径特征的交叉融合，从而使模型提取到更丰富的信息，提高模型的可信度。本文设计的 DPVF 模型主要包含 3 个模块：LGL (Local-Global-Local) 模块、Cross Attention 模块以及 Head 模块。LGL 模块用于提取输入图像特征，Cross Attention 模块用于 DPVF 模型两分支的特征融合，Head 模块用于将融合后的特征输出为卒中 DSA 图像的 mTICI 评分类别。

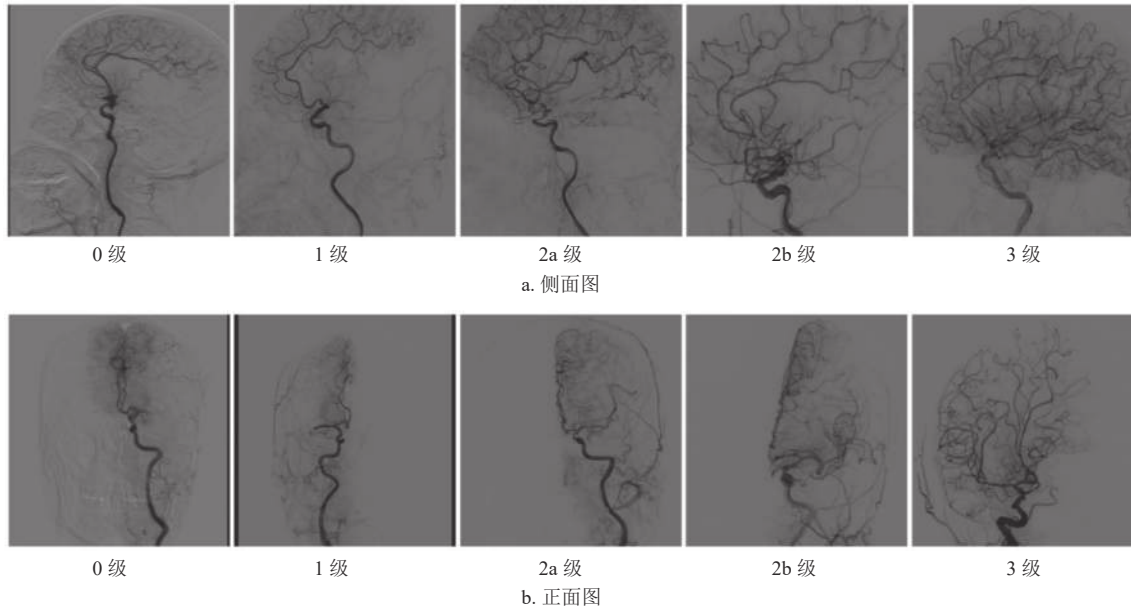


图1 正面和侧面 DSA 图像实例图

## 1.1 LGL 模块

Transformer<sup>[11]</sup> 最初是在自然语言处理 (NLP) 领域中提出的, 通过引入自注意力机制来捕捉输入序列中的全局关系, 取得了比传统循环神经网络更好的效果和性能。随着深度学习的发展, Transformer 也被引入计算机视觉领域, 典型代表模型是 ViT<sup>[12]</sup>, 它将输入图像分成若干个图块 (patches), 将这些 patches 转化为一个序列 (tokens), 再利用多头自注意力机制学习图像的全局语义信息。然而, 由于图像具有较高的空间冗余性, 在每个空间位置都执行自注意力的计算是低效的。EdgeViT 的 LGL 模块只使用输入 tokens 的一部分子集计算自注意力, 但可以实现完全的空间交互, 在保证性能的前提下减少了特征的计算量和参数量。

### 1.1.1 EdgeViT 的 LGL 模块简介

如图 2 最右侧虚线框所示, LGL 模块主要包含 3 部分, 分别为 Local Aggregation、Global Sparse Attention、Local Propagation。Local Aggregation 模块的主要功能是对每个 token, 利用深度可分离卷积在大小为  $k \times k$  的局部窗口中聚合信息, 如图 3a 所示。Global Sparse Attention 模块的主要功能是对均匀分布在空间中的一组稀疏代表性 token 进行采样, 每个  $r \times r$  ( $r$  表示下采样率) 窗口对应一个代表性 token, 然后对这些选定的 token 应用自注意力进行全局建模, 如图 3b 所示。Local Propagation 模块的主要功能是通过转置卷积将代表性 tokens 中编码的全局上下文信息传播到其相邻的 tokens, 如图 3c 所示。

### 1.1.2 改进的 LGL 模块

LGL 模块使用稀疏空间自注意力机制来建模输入序列 (tokens) 中每个位置之间的空间关系, 通过计算输入序列中每个位置之间的相关性得分来决定哪些位置对当前位置的表示具有更大的贡献, 以使模型能够更好地关注于目标相关的信息。为提升 LGL 模块的建模能力, 本文对 LGL 模块的 Global Sparse Attention 子模块进行了如下改进: 构建空间-通道稀疏自注意力模块来高效地实现全局建模, 使用通道自注意力建立输入序列中不同通道之间的关系, 通过计算不同通道之间的相关性得分来决定哪些通道对当前位置的表示具有更大的贡献, 从而能够更好地学习到输入序列中的特征。将空间自注意力和通道自注意力进行结合, 从两个正交的角度进行自注意力的计算, 可以帮助模型更好地捕捉输入序列中的空间和通道特征, 从而提高模型的性能。

改进的 Global Sparse Attention 模块结构图如图 4a 所示, 这里使用空间-通道双注意力机制来对输入图像进行特征建模, 并采取残差结构以避免模型退化。空间自注意力子模块如图 4b 所示, 计算公式为:

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Concat}(h_1, h_2, \dots, h_H) \quad (1)$$

$$h_i = \text{Attention}(\mathbf{Q}_i, \mathbf{K}_i, \mathbf{V}_i) \quad (2)$$

$$\text{Attention}(\mathbf{Q}_i, \mathbf{K}_i, \mathbf{V}_i) = \text{softmax} \left[ \frac{\mathbf{Q}_i \mathbf{K}_i^T}{\sqrt{C/H}} \right] \mathbf{V}_i \quad (3)$$

式中,  $h_i$  表示自注意力计算的第  $i$  个 head,  $H$  表示多头自注意力头 (head) 的个数, 第  $i$  个 head 的

$Q_i, K_i, V_i$  通过如下线性映射得到:

$$Q_i = X_i W_i^Q \quad (4)$$

$$K_i = X_i W_i^K \quad (5)$$

$$V_i = X_i W_i^V \quad (6)$$

通道自注意力子模块如图 4c 所示, 计算过程可以用如下公式表示:

$$\text{Attention}(Q, K, V) = \text{Concat}(h_1, h_2, \dots, h_G) \quad (7)$$

$$h_i = \text{Attention}(Q_i, K_i, V_i) \quad (8)$$

$$\text{Attention}(Q_i, K_i, V_i) = \text{softmax} \left[ \frac{Q_i^T K_i}{\sqrt{C/G}} \right] V_i^T \quad (9)$$

式中,  $G$  表示通道的组数,  $Q_i, K_i, V_i$  的线性投射方式与空间自注意力一致, 而在通道自注意力计算

时, 将  $Q_i, K_i, V_i$  的维度进行反转, 便可实现在通道维度上的自注意力的计算。

因此, 改进的 LGL 模块可以表示为:

$$X = \text{LocalAgg}(\text{Norm}(X_{\text{in}})) + X_{\text{in}} \quad (10)$$

$$Y = \text{FFN}(\text{Norm}(X)) + X \quad (11)$$

$$Y' = \text{GlobalSparseAttn}(\text{Norm}(Y)) + Y \quad (12)$$

$$Z = \text{LocalProp}(\text{Norm}(Y')) + Y' \quad (13)$$

$$X_{\text{out}} = \text{FFN}(\text{Norm}(Z)) + Z \quad (14)$$

式中,  $X_{\text{in}}$  表示 LGL 模块的输入;  $X_{\text{out}}$  表示 LGL 模块的输出; Norm 表示层归一化<sup>[13]</sup> (Layer Normalization); LocalAgg 表示 LGL 模块的 Local Aggregation 子模块; GlobalSparseAttn 表示空间-通道稀疏自注意模块; LocalProp 表示 LGL 模块的 Local Propagation 子模块。

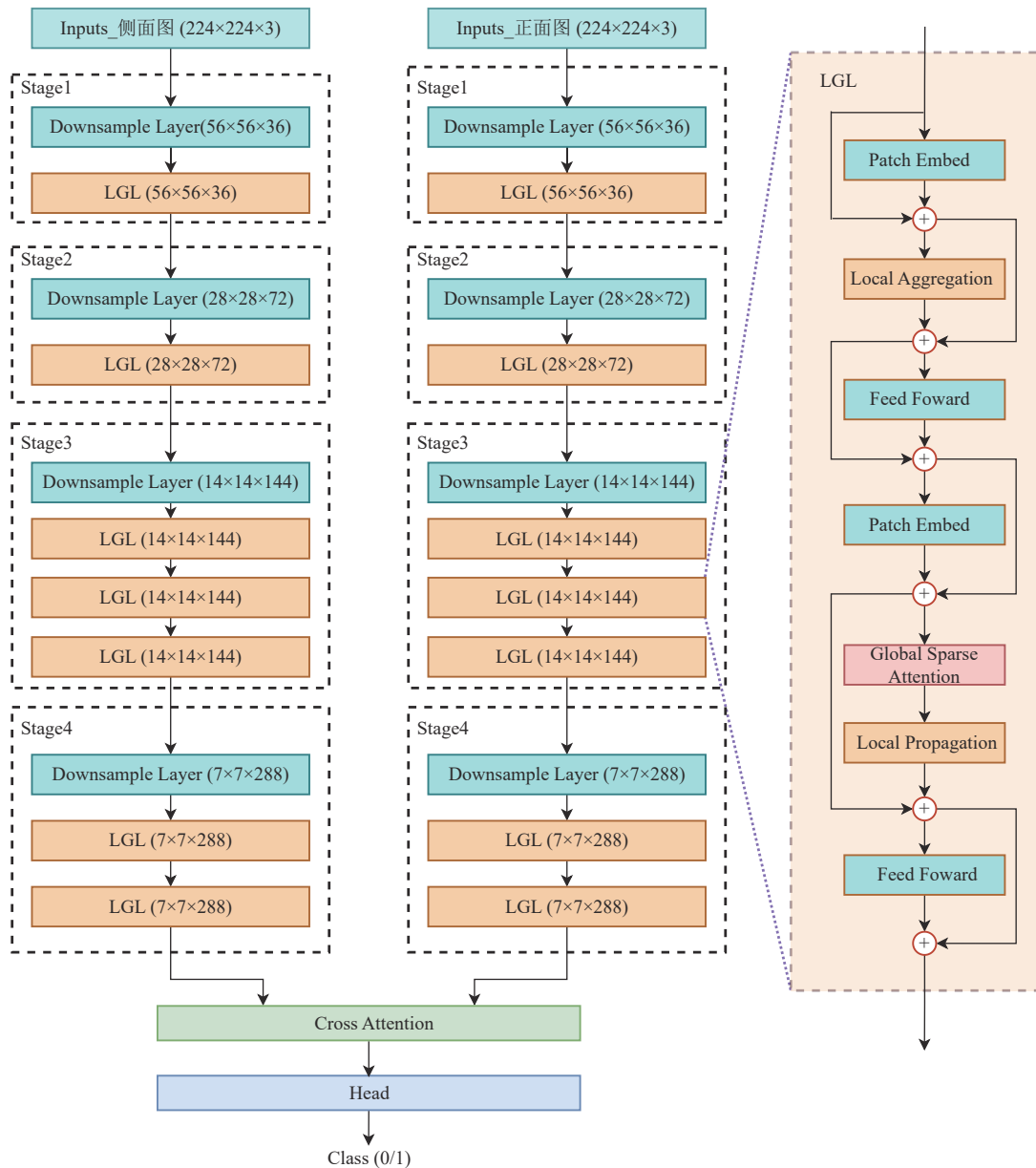


图 2 DPVF 网络结构图

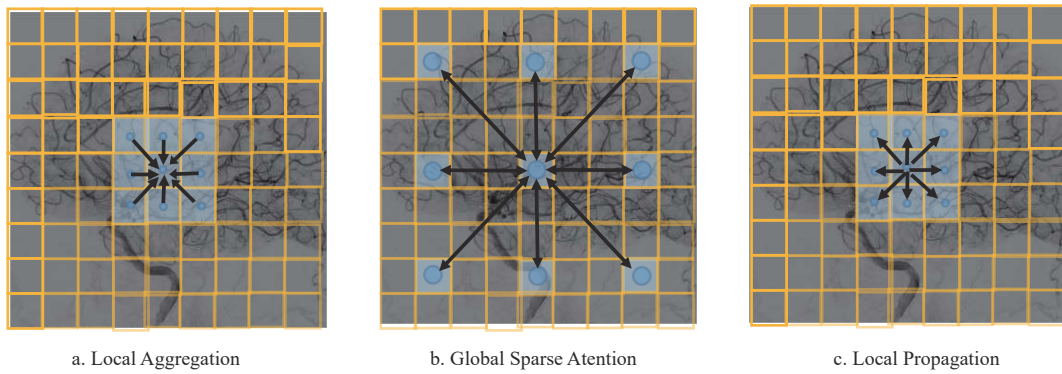


图 3 LGL 模块的 3 个子模块操作示意图

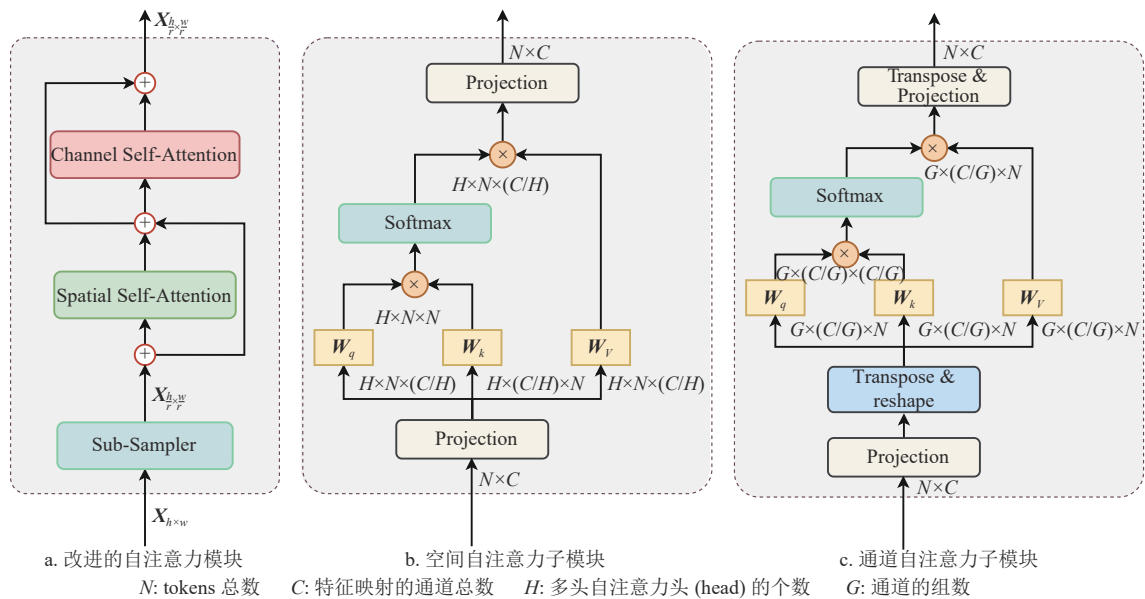


图 4 空间-通道稀疏自注意力模块

### 1.2 Cross Attention 模块

为了充分融合 AIS 患者的正面和侧面 DSA 图像信息, 构建交叉注意力模块实现双分支网络的特征融合。如图 5 所示, 将经过改进的 LGL 模块训练的正面图像 tokens 和侧面图像 tokens 进行交叉融合, 为降低自注意计算的运算量, 将正面图像 tokens 进行全局平均池化, 得到一个包含正面图像全局信息的 token, 将此 token 与侧面图像的 tokens 进行多头自注意的计算, 输出包含正侧面图像信息的 token。侧面图像 tokens 与正面图像 tokens 之间的交叉融合与此过程一致。经过 Cross Attention 模块的最终输出是正面和侧面 DSA 图像 tokens 交叉融合后的两个 token。

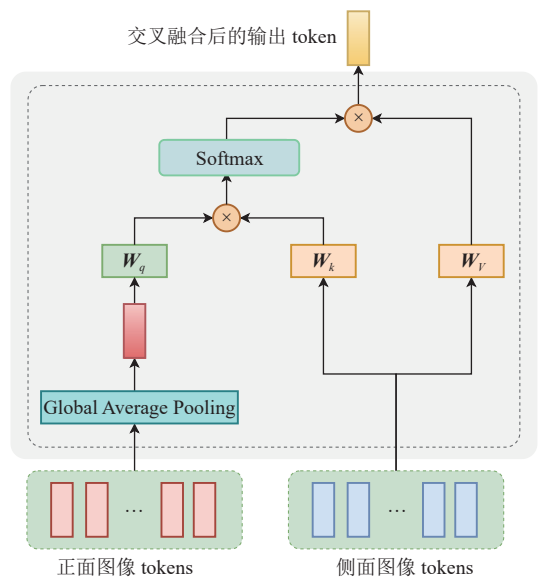


图 5 Cross Attention 模块

### 1.3 Head 模块

对于 Head 模块，它的主要作用是根据 Cross Attention 模块最终输出的 tokens 实现卒中 DSA 图像的 mTICI 评级分类。如图 6 所示，Head 模块将经过 Cross Attention 模块融合后的两分支的输出进行拼接，对拼接结果应用两个线性层，以输出模型最终的分类结果。

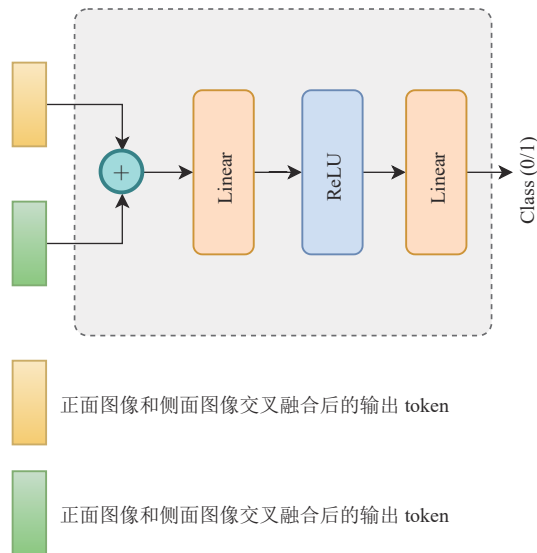


图 6 Head 模块结构图

## 2 实验与分析

### 2.1 数据集

本实验使用的数据集涉及 194 名 AIS 患者，使用 DSA 设备采集减影血管图像序列，每个序列包含 15~30 帧不等，对这些患者在治疗过程中拍摄得到的 DSA 图像进行筛选分级，得到 1 019 对质量良好的正侧面图像。根据脑血管再通是否成功作为分类依据，将 mTICI 评级为 2b 或 3 标为类别 1，血管未再通成功（mTICI 评级为 0、1 或 2a）标注为类别 0。再通成功数据共计 537 对，未再通成功共计 482 对，具体如表 1 所示。训练集和测试集按照 8:2 的比例随机划分。

表 1 数据集详情表

mTICI 级别	数量/对	所属分类
0	368	0
1	40	0
2a	74	0
2b	170	1
3	367	1

### 2.2 评价指标

对于医学图像分类模型，常用的评价指标包括

准确率 (Accuracy)、精确率 (Precision)、召回率 (Recall) 和 F1 得分 (F1 score)，分别表示为：

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (15)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (16)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (17)$$

$$\text{F1 score} = \frac{2 \times TP}{2 \times TP + FP + FN} \quad (18)$$

式中，TP、TN、FP、FN 分别代表真阳性、真阴性、假阳性、假阴性，也即分别代表“预测为正样本，且预测正确”“预测为负样本，且预测正确”“预测为正样本，但预测错误”“预测为负样本，但预测错误”。

### 2.3 实验结果分析

在 RTX 2080Ti 上基于深度学习框架 Pytorch 进行模型的训练，迭代次数为 100，batch size 设置为 8，初始学习率为 0.000 1，使用 Adam 优化器<sup>[14]</sup>。DPVF 模型的训练过程如图 7 和图 8 所示，可以看出，DPVF 的准确率随 epoch 增加而增加，损失值随 epoch 增加而减小，约迭代 20 次后，准确率和损失都收敛到一个较小的区间范围，说明 DPVF 经过了充分的训练后能够快速收敛。

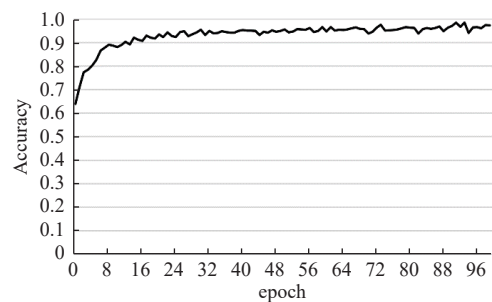


图 7 DPVF 模型训练准确率随 epoch 变化曲线

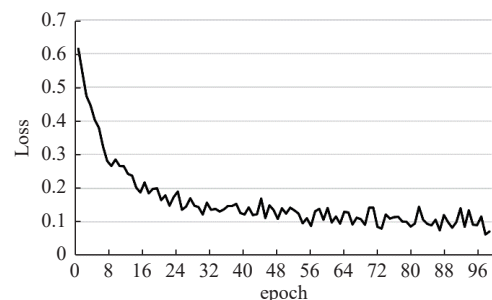


图 8 DPVF 模型训练损失随 epoch 变化曲线

对经过训练的 DPVF 模型与现有主流模型在测试集上测试, 其中, 每个经典模型的测试包括 3 方面, 分别为输入单张正面图像的单分支模型、输入单张侧面图像的单分支模型, 以及同时输入

正面和侧面两张图像的采用 Concat 融合方式的双分支模型, 分别计算每个模型的混淆矩阵、准确率、精准率、召回率和 F1 得分, 对比实验结果如表 2 所示。

表 2 DPVF 与其他分类模型在测试集上的实验结果对比

模型	TP	FP	FN	TN	Accuracy/%	Precision/%	Recall/%	F1 score/%
本文DPVF	94	2	1	106	98.5	97.9	98.9	98.4
EdgeViT (正+侧)	95	1	5	102	97.0	99.0	95.0	97.0
EdgeViT (正+侧) 拼接	93	3	4	103	96.6	96.9	95.9	96.4
EdgeViT (正)	91	5	4	103	95.6	94.8	95.8	95.3
EdgeViT (侧)	90	6	2	105	96.1	93.8	97.8	95.8
ViT-B/16 (正+侧)	83	13	10	97	88.7	86.5	89.2	87.8
ViT-B/16 (正)	96	0	26	81	87.2	100	78.7	88.1
ViT-B/16 (侧)	83	13	16	91	85.7	86.5	83.8	85.1
ShuffleNet V2 (正+侧)	88	8	3	104	94.6	91.7	96.7	94.1
ShuffleNet V2 (正)	92	4	10	97	93.1	95.8	90.2	92.9
ShuffleNet V2 (侧)	93	3	9	98	94.1	96.9	91.2	94.0
ResNet-50 (正+侧)	95	1	4	103	97.5	99.0	96.0	97.5
ResNet-50 (正)	90	6	2	105	96.1	93.8	97.8	95.8
ResNet-50 (侧)	95	1	5	102	97.0	99.0	95.0	97.0
MobileNet V2 (正+侧)	76	20	3	104	88.7	79.2	96.2	86.9
MobileNet V2 (正)	73	23	18	89	79.8	76.0	80.2	78.0
MobileNet V2 (侧)	75	21	15	92	82.3	78.1	83.3	80.6
ConvNeXt (正+侧)	95	1	4	103	97.5	99.0	96.0	97.5
ConvNeXt (正)	92	4	2	105	97.0	95.8	97.9	96.8
ConvNeXt (侧)	94	2	4	103	97.0	97.9	95.9	96.9
AlexNet (正+侧)	94	2	20	87	89.2	97.9	82.5	89.5
AlexNet (正)	83	13	21	86	83.3	86.5	79.8	83.0
AlexNet (侧)	78	18	12	95	85.2	81.3	86.7	83.9

使用折线图对每个模型在测试集上的准确率进行可视化, 如图 9 所示, 可以看出, 本文提出的 DPVF 模型在测试集上的准确率达到最高的 98.5%, 优于基于 EdgeViT 的正面加侧面图像输入的双分支模型, 说明了本文改进的 LGL 模块的有效性; 此外, 可以看出, EdgeViT 的双分支采用交叉注意力的融合方式的准确率达 97.0%, 而采用拼接融合方式的准确率为 96.6%, 说明本文构建的交叉注意力模块可以达到比拼接方式更好的特征融合效果, 证明了本文构建的交叉注意力模块的有效性; 此外, 输入正面和侧面两幅图像的双分支模型的准确率均高于输入单幅图像对应

的单分支模型, 说明同时输入正面和侧面两幅图像有助于模型提取到更丰富的特征, 可以有效提升图像分类的准确率和精度。结合表 2 可以看出, DPVF 在准确率、精准率、召回率和 F1 得分上均取得了较好的表现, 其中准确率达 98.5%, F1 得分、精准率和召回率分别达到了 98.4%、97.9%、98.9%。与单分支 ViT 模型及其他模型相比, DPVF 在各个指标上均具有一定的优势, 说明本文构建的正面和侧面图像交叉融合有效地提取了正面和侧面图像多角度的特征, 优于两幅图像的简单拼接融合, 从而提升了模型的 mTICI 自动分级准确率。

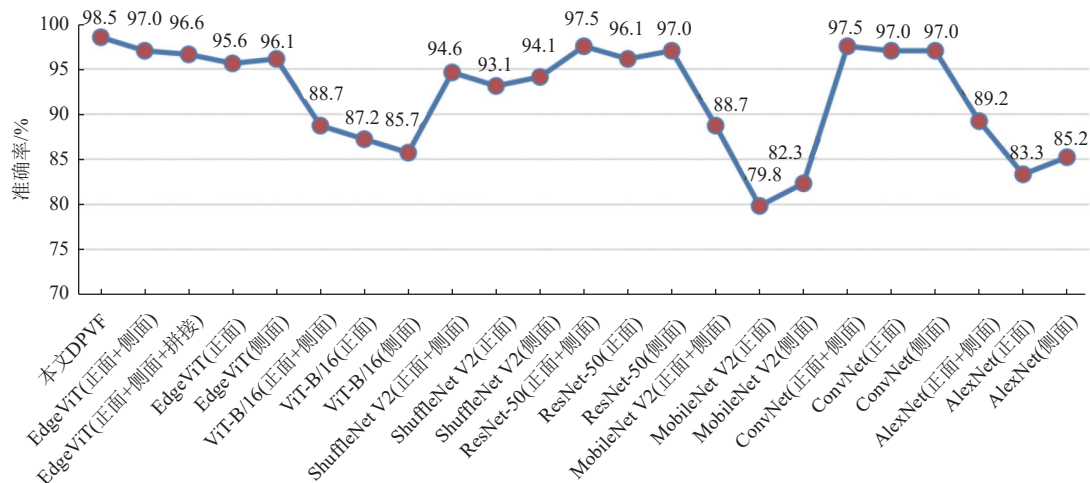


图9 DPVF模型与其他模型在测试集上的准确率比较

### 3 结束语

本文构建了一个基于 Vision Transformer 的轻量级双路径图像分类模型 DPVF 用于急性缺血性脑卒中患者 DSA 图像的自动分级。基于 EdgeViT 的轻量化设计思想进行了模型的构建，并提出空间-通道自注意力模块对原有自注意力模块进行改进，以使模型保持轻量化的同时捕获更全面的特征信息，提高模型表达能力；此外，构建交叉注意力模块对 DPVF 模型的两分支进行交叉融合，促使模型提取更丰富的特征，从而提高模型表现。实验结果表明，本文构建的 DPVF 比其他图像分类模型要好，证明了本文方法的可行性和有效性。

#### 参考文献

- [1] 邱甲军, 吴跃, 惠宇, 等. 肝细胞癌 MR 图像的纹理分类研究[J]. 电子科技大学学报, 2019, 48(4): 619-626.
- [2] 沈子祺, 谢文军, 刘晓平. 基于视频的自动 Fugl-Meyer 评估方法研究[J]. 电子测量与仪器学报, 2022, 36(2): 1-11.
- [3] HE K, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN[C]//Proceedings of the IEEE International Conference on Computer Vision. [S.l.]: IEEE, 2017: 2961-2969.
- [4] SHIRAZ B M M, SNYDER K V, WAQAS M, et al. Use of quantitative angiographic methods with a data-driven model to evaluate reperfusion status (mTICI) during thrombectomy[J]. *Neuroradiology*, 2021, 63(9): 1429-1439.
- [5] CHAUHAN S, VIG L, DE F D G M, et al. A comparison of shallow and deep learning methods for predicting cognitive

- performance of stroke patients from MRI lesion images[J]. *Frontiers in Neuroinformatics*, 2019, 13: 53.
- [6] CHEON S, KIM J, LIM J. The use of deep learning to predict stroke patient mortality[J]. *International Journal of Environmental Research and Public Health*, 2019, 16(11): 1876.
- [7] NIELSEN A, HANSEN M B, TIETZE A, et al. Prediction of tissue outcome and assessment of treatment effect in acute ischemic stroke using deep learning[J]. *Stroke*, 2018, 49(6): 1394-1401.
- [8] BHURWANI M M S, SNYDER K V, WAQAS M, et al. Use of biplane quantitative angiographic imaging with ensemble neural networks to assess reperfusion status during mechanical thrombectomy[C]//Medical Imaging 2021: Computer-Aided Diagnosis. [S.l.]: Springer, 2021, 11597: 328-336.
- [9] SU R, CORNELISSEN S A P, VAN D S M, et al. AutoTICI: Automatic brain tissue reperfusion scoring on 2D DSA images of acute ischemic stroke patients[J]. *IEEE Transactions on Medical Imaging*, 2021, 40(9): 2380-2391.
- [10] PAN J, BULAT A, TAN F, et al. Edgevit: Competing light-weight CNNs on mobile devices with vision transformers[C]//Computer Vision-ECCV 2022: 17th European Conference. [S.l.]: Springer, 2022: 294-311.
- [11] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[EB/OL]. [2022-10-22]. <https://arxiv.org/pdf/1706.03762.pdf>.
- [12] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words: Transformers for image recognition at scale[EB/OL]. [2022-06-15]. <https://arxiv.org/abs/2010.11929v1>.
- [13] BA J L, KIROS J R, HINTON G E. Layer normalization[EB/OL]. [2022-06-25]. <https://arxiv.org/pdf/1607.06450.pdf>.
- [14] KINGMA D P, BA J. Adam: A method for stochastic optimization[EB/OL]. [2022-07-10]. <http://www.arxiv.org/pdf/1412.6980.pdf>.