

DOI:10.3969/j.issn.1005-202X.2024.05.007

医学影像物理

基于非局部卷积和卷积注意力模块的眩晕眼震诊断方法

贺斌, 高永彬

上海工程技术大学电子电气工程学院, 上海 201620

【摘要】鉴于良性阵发性位置性眩晕的复杂致病因素和诊断困难等问题,提出一种新的基于非局部卷积和卷积注意力模块(CBAM)的眩晕眼震诊断方法。首先,通过构建目标检测模型定位瞳孔,从而捕捉眼球运动并提取水平和垂直运动轨迹时序数据。其次,采用分类模型进行分类检测,该分类模型采用非局部卷积来捕获眼震数据中的远程依赖关系特征,并引入CBAM来提取特征层中的高级和低级语义信息,从而提高了分类模型的检测性能。在耳鼻喉科医院提供的视频眼震数据集上进行了实验,结果表明,与主流方法相比,本文所提出的诊断方法在精确率、召回率、准确率、平均F1值等评估指标上比主流方法分别提高了1.82%、2.09%、1.62%和1.96%,表明了本文方法的显著性优势。

【关键词】良性阵发性位置性眩晕;医学图像处理;时序数据分类;目标检测;视频眼震数据分类

【中图分类号】R318

【文献标志码】A

【文章编号】1005-202X(2024)05-0571-08

Diagnostic methods for nystagmus in vertigo based on non-local convolution and convolutional block attention module

HE Bin, GAO Yongbin

School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai 201620, China

Abstract: In view of complex pathogenic factors and diagnostic difficulties of benign paroxysmal positional vertigo, a novel method for diagnosing nystagmus in vertigo based on non-local convolutional and convolutional block attention module (CBAM) is proposed. An object detection model is constructed to locate the pupil, thereby tracking eye movement and extract temporal data of horizontal and vertical motion trajectories. Subsequently, a classification model is employed for detection and classification, utilizing a non-local convolutional module to capture remote dependency relationships in nystagmus data, and introducing CBAM to extract high-and low-level semantic information in the feature layer for enhancing the detection performance. Experiments were conducted on a video nystagmus dataset provided by the Eye, Ear, Nose and Throat Hospital. The results show that compared with the best mainstream method, the proposed method improves precision, recall rate, accuracy, and average F1 score by 1.82%, 2.09%, 1.62%, and 1.96%, respectively, demonstrating its superiority.

Keywords: benign paroxysmal positional vertigo; medical image processing; temporal data classification; object detection; video nystagmus data classification

前言

眩晕(Vertigo)是耳鼻喉科的常见主诉^[1],也是急诊和神经内科患者的常见症状^[2]。流行病学调查显示^[3],大约21%的人一生中会经历至少一次眩晕^[4]。眩晕常导致患者生活质量受到影响^[5]。眩晕的致病

因素众多,且诊断困难^[6]。其中,良性阵发性位置性眩晕(Benign Paroxysmal Positional Vertigo, BPPV),俗称耳石症,是临床医生公认的眩晕疾病中发病率最高也是最难诊断的疾病之一^[7]。依据最新的BPPV诊断和治疗指南,年发病率为(10.7~600)/10万,年患病率约为1.6%^[8],目前世界上并无可以直接用于临床的计算机辅助眼震检查手段^[9]。因此,BPPV诊断中最为重要的眩晕眼震诊断环节并无足够的技术支持^[10]。为了通过使用计算机来辅助检测眩晕眼震^[11],前人做了一些尝试,文献[12]使用圆形算子来定位瞳孔位置,并且提取眼球运动特征,文献[13]使用了一种基于重心的算法来跟踪瞳孔中心,采用圆形霍夫变换检测椭圆瞳孔。如果检测到瞳孔,则使

【收稿日期】2023-12-18

【基金项目】上海市科委“科技创新行动计划”社会发展科技攻关项目(21DZ1204900)

【作者简介】贺斌,硕士研究生,研究方向:医学图像处理,E-mail: hebin99@foxmail.com

【通信作者】高永彬,博士,副教授,研究方向:计算机视觉,智慧医疗,E-mail: gaoyongbin@sues.edu.cn

用边缘检测和椭圆拟合算法来定位瞳孔的中心。但是该方法容易受到睫毛遮挡和眨眼等干扰,从而导致瞳孔定位存在误差,影响检测精度。文献[14]通过定位瞳孔位置,得到眼震的水平、垂直和虹膜旋转运动特征,并将其特征数据使用1D-CNN进行训练,从而进行特征分类,判断出受试者所患的眼震疾病。文献[15]通过使用连续动态前庭评估(CAVA)设备捕获的角视网膜电位来预测眼球震颤,但是,该方法无法检测扭转性眼震,并且患者由于需要长时间佩戴该设备,诊断体验感受较差,在临床上很难进行应用。文献[16]提出使用光流法来提取瞳孔运动特征,但这种方法需要的计算资源较大,也很难在临床上应用。文献[17]提出了一种基于卷积-卷积神经网络(ConvNet)的眼动视频压缩方法。在对捕捉到的帧序列中的移动瞳孔进行标定时,将Hough变换与基于模板匹配的轨迹跟踪相结合,提高了对睫毛遮挡和瞳孔变形的鲁棒性,提出使用光流法来提取虹膜旋转特征。但是该方法仅检测了虹膜旋转,并没有检测其他特征,不能直接应用于疾病检测。

目前眩晕眼震智能辅助检测面临的问题有:(1)没有相关的眩晕眼震公开数据集。(2)眼震视频检测算法速度慢,瞳孔定位主要依赖传统边缘检测方法。(3)目前已有的方法基于传统卷积神经网络(CNN)模型进行设计,无法避免传统CNN具有归纳偏置、缺乏长距离注意力的问题。

针对以上问题,本研究提出了一种基于非局部

卷积和卷积注意力模块(Convolutional Block Attention Module, CBAM)的眩晕眼震诊断方法,主要贡献点如下:(1)针对没有相关眩晕眼震公开数据集,创建了相关眩晕眼震视频数据集以及相应的眼球运动时序数据集。(2)针对传统瞳孔定位算法检测速度慢、误差大等问题,提出了目标检测瞳孔定位算法,捕捉眼震视频中眼球在水平和垂直的运动轨迹,转换为时序数据供眼震检测模型训练使用。(3)在眩晕眼震计算机辅助诊断领域,提出一种基于非局部卷积和CBAM的眩晕眼震诊断方法,本研究通过将非局部卷积^[18]以及CBAM^[19]融入到网络结构中,充分利用了眼震时序数据中的序列关系,提高了眩晕眼震的分类效果。

1 本研究方法

本研究构建了一个眩晕眼震诊断方法,整体流程图如图1所示,首先将眩晕眼震视频帧中的无效帧进行剔除,减少眨眼和睫毛遮挡对瞳孔定位的影响,其次在特征提取部分,为了降低模型的计算负担,本研究引入了YOLOv5目标检测模型^[20],通过定位瞳孔位置来提取眼震运动特征,得到一组包含眼球运动轨迹的时序特征曲线数据,随后将处理过的眼震运动时序数据送入到基于非局部卷积和CBAM的眩晕眼震分类网络中进行训练,最后,通过分类层网络得到最终的分类预测结果^[21]。

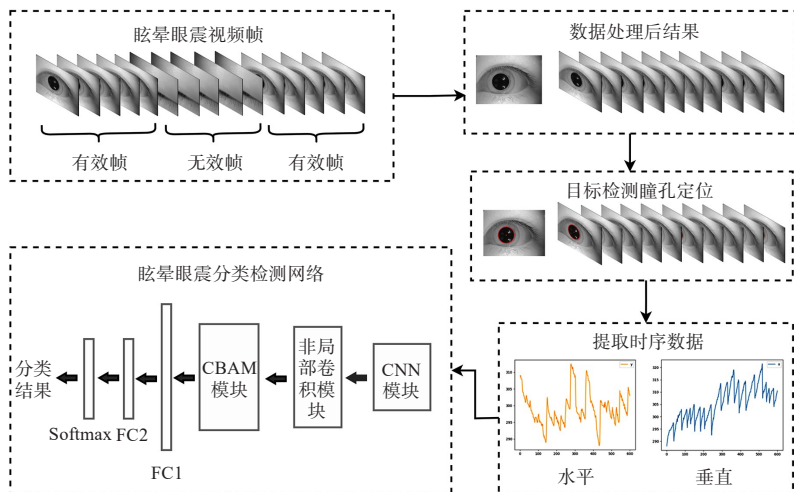


图1 眩晕眼震检测方法整体流程图
Figure 1 Overall flow chart of vertigo nystagmus detection method

1.1 基于YOLOv5的瞳孔定位模块

目前已有的瞳孔定位方法主要有基于传统的轮廓检测以及Hough变换等方法^[22-23],这些方法存在着

检测有误差、受视频数据质量影响大等问题,所以为了解决这些问题,本研究提出使用深度学习算法来定位瞳孔位置,使用训练好的目标检测模型来定位

瞳孔位置。从而提高定位瞳孔的准确率和泛化性。本研究制作了一个瞳孔定位目标检测数据集,在该数据集上训练出一个基于YOLOv5架构的目标检测模型,然后使用该模型来定位瞳孔,从而提高对瞳孔定位的准确性。

由于笔者需要定位临床视频中的瞳孔,所以,第

一步是对视频进行拆帧处理,对每一帧进行瞳孔定位。如图2所示,首先将拆分的每一帧输入到训练好的YOLOv5模型中,进行特征提取,从而得到定位好的瞳孔位置。图3为YOLOv5瞳孔定位对于不同瞳孔位置以及有上下眼皮遮挡的定位效果。

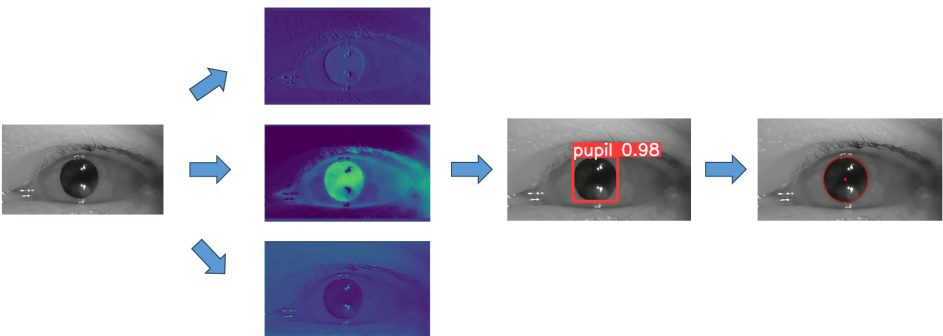


图2 瞳孔定位算法流程
Figure 2 Pupil positioning algorithm flow

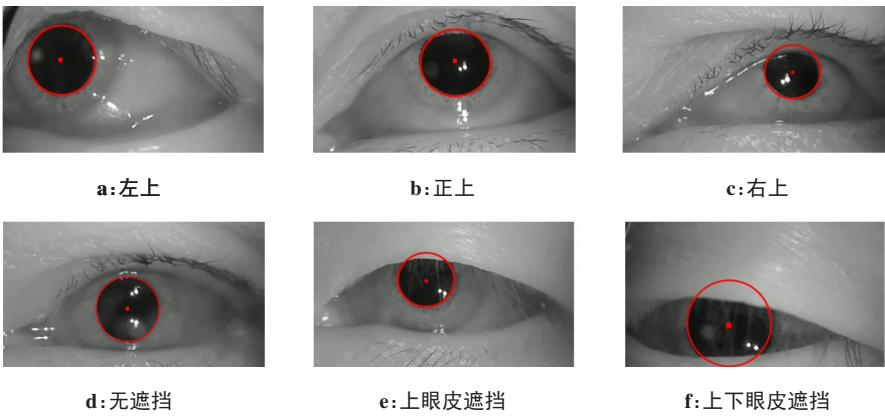


图3 瞳孔定位效果
Figure 3 Pupil positioning results

1.2 数据处理模块

为了提高模型分类的准确度,本研究将视频数据进行裁剪,将长短不一的视频统一裁剪为10 s的视频,统一数据尺寸。并且将获取到的时序数据进行数据预处理,虽然在此之前已经对数据进行处理,去除了大量眨眼和睫毛遮挡的无效帧,但在瞳孔定位期间还是会有个别数据定位出现偏差,所以,本研究提出使用插值算法对定位异常数据进行插值处理,更好的避免异常数据对训练模型的干扰,具体插值算法公式如下:

$$y = y_0 + (x - x_0) \frac{y_1 - y_0}{x_1 - x_0}$$

(1)

其中, x_0, x_1 为时序序列数据中的位置坐标, y_0, y_1 为时

序序列数据中的具体数值,通过以上公式得到 $[x_0, x_1]$ 区间内某一位置 x 对应的检测异常点处的 y 值。

通过使用该插值算法,解决了数据中异常数据点的问题,最终得到每一个视频对应一个 2×600 的矩阵时序数据。为了避免水平运动眼震时序数据和垂直运动眼震时序数据的运动尺度范围不同影响模型的训练效果,本研究提出了使用数据标准化处理,将眼震运动时序数据调整为均值为0和标准差为1的分布,消除了不同眼震运动之间的尺度影响,确保不同眼震震动幅度对模型的权重和影响是平等的,提高模型的训练效果,标准化公式如下:

$$x_{\text{normalization}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$$

(2)

其中, x_{\max} 为眼震时序数据中的最大值, x_{\min} 为眼震时序数据中的最小值。通过瞳孔定位, 得到了眼球水平和垂直运动轨迹, 如图4所示, 本研究采取提取特

征的方式, 提取出了眩晕眼震中最关键的特征, 并以时间维度进行合并, 生成时间序列数据。

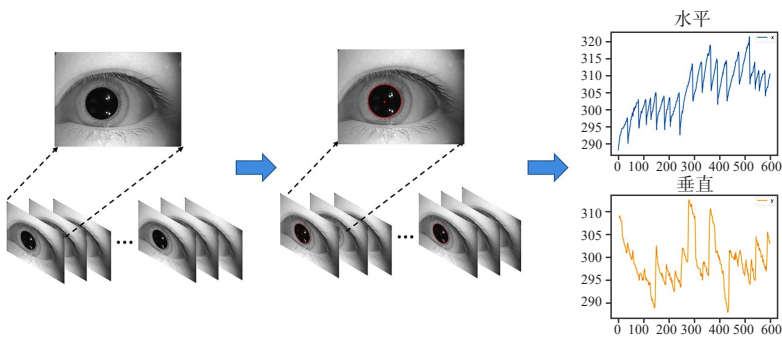


图4 眼震视频数据转换为时序数据流程图
Figure 4 Flowchart of converting nystagmus video data into temporal data

1.3 眩晕眼震分类检测网络模块

该网络组成部分包括卷积层、非局部卷积+CBAM层、全局平均池化(GMP)层以及全连接分类层。其整体结构如图5所示, 图中的卷积模块包含卷

积层和 ReLu 激活层。分类部分由 GMP 层、3 层全连接层以及 Softmax 分类器组成。详细网络模型结构如表1所示。

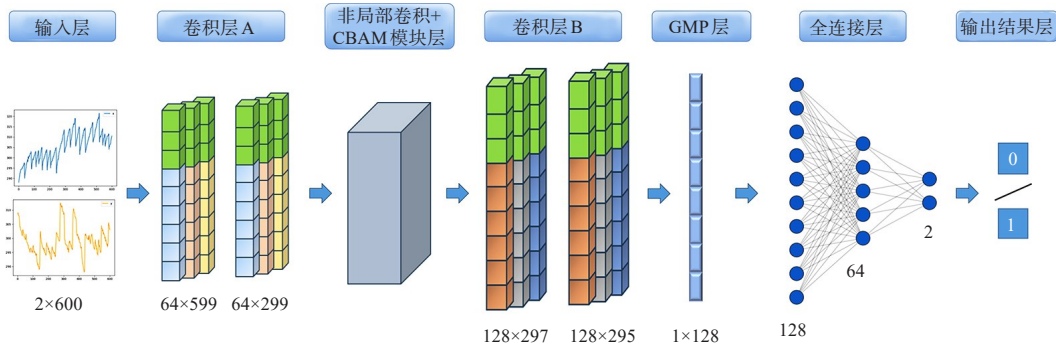


图5 模型网络结构图
Figure 5 Model network structure

| 表1 详细网络模型结构参数 | | |
|---|------|---------|
| Table 1 Detailed network model structure parameters | | |
| 神经网络模块结构 | 参数 | 尺寸大小 |
| - | 卷积核 | 3×3 |
| 卷积层A | 输入大小 | 2×600 |
| | 输出大小 | 64×299 |
| 非局部卷积+CBAM模块层 | 输入大小 | 64×299 |
| | 输出大小 | 64×299 |
| 卷积层B | 输入大小 | 64×299 |
| | 输出大小 | 128×295 |

1.3.1 非局部卷积模块 本文在骨干网络中构建了非局部注意力层^[24], 非局部注意力层的架构如图6所示。本文模型的非局部注意力层的计算公式为:

$$y_i = \frac{1}{\sum_{\forall j} e^{\theta(x_i)^T \phi(x_j)}} \sum_{\forall j} e^{\theta(x_i)^T \phi(x_j)} g(x_j) \tag{3}$$

其中, i 表示输出层在位置 i 处的输出, 而 j 表示特征图中除了位置 i 之外的其他位置。 x 是输入的特征, y 是输出的特征, 它们具有相同的维度大小。 $\sum_{\forall j} e^{\theta(x_i)^T \phi(x_j)}$ 代表了成对函数, 用于计算位置 i 和位置 j 之间的相似性。这个成对函数将 x_i 和 x_j 映射到高斯空间, 以计

算两个位置特征图之间的相似度^[25]。

图6中 C 和 L 分别表示特征图的通道数量和长度。 \otimes 用于表示矩阵相乘操作,而最上面的 \oplus 表示元素求和,也即是公式(3)所表述的内容,即将原始特征 x_i 与通过注意力机制计算得到的特征 $W_z y_i$ 相加。

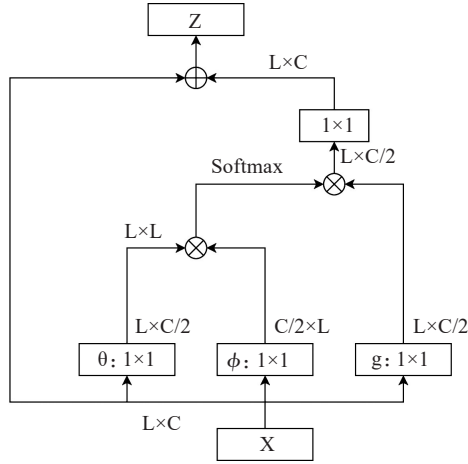


图6 非局部卷积模块结构图

Figure 6 Structure diagram of non-local convolutional module

最终, $\sum_{vj} e^{\theta(x_i)^T \phi(x_j)}$ 的计算结果将代表位置 i 处的特征与其他所有位置特征之间的相似度权重的加权和。关于 $g(x_j)$ 的表达式如下:

$$g(x_j) = W_g x_j \quad (4)$$

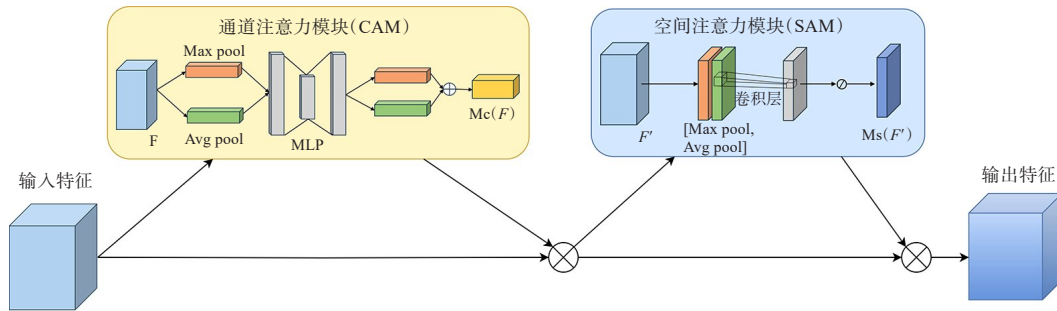


图7 CBAM结构图

Figure 7 CBAM structure

CAM通过最大池化和平均池化对输入特征图 F 的空间信息进行聚合,生成最大池化特征和平均池化特征两种不同类型的特征^[27]。然后,将它们转发到一个权重共享网络,该网络由一个多层感知器(MLP)组成,然后接进一步的最大池化层、平均池化层以及一个非线性激活单元,从而产生一个通道注意力图 $M_c(F)$ 。

其中, $g(x_j)$ 用于对输入特征 x_j 进行映射变换,而 W_g 则表示可学习的权重矩阵。最终的非局部注意力模块可以用以下方式表示:

$$Z_i = W_z y_i + x_i \quad (5)$$

其中, W_z 表示学习得到的权重矩阵,符号 $+$ 表示按元素相加, $W_z y_i + x_i$ 表示将非局部注意力模块前的特征与非局部注意力模块后的特征进行直接相加。这种操作的目的在于保持原有语义特征的基础上,同时捕获了任意两个位置之间的远程依赖关系,而不仅仅局限于相邻点的依赖。这相当于使用了一个与特征具有相同大小的卷积核来执行卷积操作,从而能够捕获全局的语义信息。

1.3.2 CBAM CBAM属于混合域注意力模块^[26],其模块结构如图7所示,由通道注意力模块(CAM)和空间注意力模块(SAM)组成。在输入特征依次经过CBAM的CAM、SAM后,输入特征图中的空间、通道特征信息得到自适应特征细化。

CBAM计算的整体过程可以总结如下:

$$\begin{aligned} F' &= M_c(F) \otimes F \\ F'' &= M_s(F') \otimes F' \end{aligned} \quad (6)$$

其中, F 表示中间映射特征图, F' 表示经过通道子模块后的输出, F'' 表示最终的细化输出; \otimes 表示逐元素乘法, $M_c(F)$ 表示1D通道注意力, $M_s(F')$ 表示1D空间注意力。

$$M_c(F) = \sigma \left(\text{MLP}(\text{AvgPool}(F)) + \text{MLP}(\text{MaxPool}(F)) \right) \quad (7)$$

其中, σ 表示sigmoid函数。

SAM沿通道维度对输入特征映射 F' 进行最大池化和平均池化操作,得到两个特征映射^[28]。将两个特征图通过卷积层进行连接和卷积,得到最终的空间注意力图 $M_s(F')$:

$$M_s(F') = \sigma\left(f^{7 \times 7}\left([\text{AvgPool}(F'); \text{MaxPool}(F')]\right)\right) \quad (8)$$

其中, σ 表示sigmoid函数, $f^{7 \times 7}$ 表示滤波器大小为7×7的卷积运算。

2 试验

2.1 试验数据集介绍

本研究建立数据集所用的VNG视频数据来自某眼耳鼻喉科医院。在位置试验中,使用红外视频眼动记录仪记录和保存患者在收到外部刺激时真实准确的眼球运动视频,视频格式是mp4,视频帧大小为640×480,帧率60。

数据集总共收集了从2020年6月到2021年6月期间录制眼震患者的3 034个初始视频段,所有的数据均由耳科专家进行标注。并经过数据预处理最终得到规范数据段5 160个,按照6:2:2的比例分别划分为训练集、验证集和测试集。其中数据分布如表2所示。

表2 数据集分布
Table 2 Dataset distribution

| 数据集 | 无眩晕眼震症状 视频个数 | 有眩晕眼震症状 视频个数 | 总视频个数 |
|------|-----------------|-----------------|-------|
| 总数据集 | 2 189 | 2 971 | 5 160 |
| 训练集 | 1 314 | 1 782 | 3 096 |
| 验证集 | 437 | 594 | 1 031 |
| 测试集 | 437 | 594 | 1 031 |

2.2 试验环境与参数

本研究均在Ubuntu上使用CUDA并行计算架构,并在Cudnn加速计算库的基础上搭建PyTorch框架,然后进行加速计算。机器显卡为NVIDIA GeForce GTX 2080 (8 G),内存为16.0 GB。CPU为Intel(R) Xeon(R) CPU E5-2603 v4 @1.70 GHz。迭代次数为100,优化器选择Adagrad^[29],优化参数选择如下:迭代学习率为1e-4,epoch为100,批大小为16,dropout为0.2。在训练过程中,经过多次迭代,在验证集上评估模型性能,并选择最佳的模型参数配置,最后在独立的测试集上对模型进行评估。

2.3 评估指标

为了评估模型性能,本研究采用二分类任务中常用的评价指标,采用准确率(Accuracy)、精确率(Precision)、召回率(Recall)、F1分数作为评价指标,对模型进行整体评估,并且训练模型时选用了交叉熵损失函数,该损失函数为常用分类任务损失函数。

计算公式如下所示:

$$\text{Accuracy} = \frac{(\text{TP} + \text{TN})}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} \quad (9)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (10)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (11)$$

$$\text{F1 Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (12)$$

其中,TP表示真正例,即模型正确预测为正例的样本数量;FP表示假正例,即模型预测为正例,但实际上是反例的样本数量;FN表示假反例,即模型预测为反例,但实际上是正例的样本数量;TN表示真反例,即模型正确预测出为反例的样本数量。

准确率是指模型正确分类的样本数量与总样本数量之间的比例,而精确率则评估了模型在预测为正类别的样本中有多少是真正的正类别。召回率衡量了模型在所有真正的正类别样本中有多少被成功预测为正类别。通常情况下,精确率和召回率这两个度量是相互制衡的,因此使用它们的调和平均数,即F1分数,作为性能指标。

2.4 对比试验

为了验证本研究提出的眩晕眼震检测模型的性能优势,笔者使用该方法与主流的眩晕眼震检测方法进行对比,其试验结果如表3所示。

表3 本文方法与其他方法对比(%)
Table 3 Comparison between the proposed method and other methods (%)

| 方法 | 精确率 | 召回率 | 准确率 | 平均F1值 |
|-----------------------|-------|-------|-------|-------|
| Lim ^[13] | 78.42 | 77.42 | 78.62 | 77.92 |
| Zhang ^[17] | 79.23 | 78.57 | 79.14 | 78.90 |
| Lu ^[14] | 82.73 | 84.15 | 83.63 | 83.43 |
| 本文方法 | 84.55 | 86.24 | 85.25 | 85.39 |

通过表3可以看出,与其他方法相比,本研究提出的方法具有明显的优势,Lim等^[13]提出使用传统算法来跟踪瞳孔中心,使用圆形霍夫变换用于检测椭圆瞳孔,使用传统的边缘检测和椭圆拟合算法来定位瞳孔的中心,在定位瞳孔阶段容易出现误差。Zhang等^[17]提出的算法只使用了光流检测虹膜旋转变化,并没有关注眩晕眼震的水平和垂直方向运动特征,检测效果一般。Lu等^[14]提出的算法基于传统CNN来识别眼震,由于CNN缺少捕捉时序之间相关性的能力,存在检测精度低,误差大等问题,本研究基于眩晕眼震的特点,提出使用基于非局部卷积和

CBAM的眩晕眼震诊断方法来捕捉时序特征,如表3所示,本研究所提出的诊断方法与其他主流方法的比较中,相比于主流方法中检测效果最好的Lu等^[14]方法在精确率、召回率、准确率、平均F1值等评估指标中分别又提升了1.82%、2.09%、1.62%、1.96%,表明所提方法的优越性。

2.5 消融试验

为了评估模型方法的有效性,本研究选择将Lu等^[14]提出的网络结构作为基础网络,分别将非局

部卷积模块、CBAM加入网络中进行眼震时序数据分类的消融试验。根据表4中的数据可以看出,一旦引入非局部卷积模块到基础网络中,试验结果中在精确率,召回率,准确率,平均F1值等评估指标都有不同程度的提升,这表明非局部卷积模块有效地捕捉了任意两个位置之间的远程依赖关系,从而更好地捕获了全局上下文信息,而不仅仅限于相邻点的依赖,提升了模型对时序数据中复杂模式的识别能力。

表4 消融试验结果(%)
Table 4 Ablation experiment results (%)

| 模型 | 精确率 | 召回率 | 准确率 | 平均F1值 |
|------------|-------|-------|-------|-------|
| 基础网络 | 82.73 | 84.15 | 83.63 | 83.43 |
| 基础网络+非局部卷积 | 83.14 | 84.93 | 84.24 | 84.03 |
| 基础网络+CBAM | 83.41 | 84.72 | 84.45 | 84.06 |
| 本文方法 | 84.55 | 86.24 | 85.25 | 85.39 |

此外,引入CBAM后,试验结果的各项指标也都有不同程度的提高,这表明CBAM有效结合了特征层中高级和低级的语义信息,并且通过空间和通道注意力将局部特征与全局特征自适应融合,从而提高了模型的检测效果。实验证明,引入以上两个模块都能在不同程度上提高眼震诊断的分类准确度。

3 结 论

通过上述试验数据对比,本研究提出的一种基于非局部卷积和CBAM的眩晕眼震诊断方法取得了较高的检测效果,与现有眩晕眼震检测方法不同,本研究采用非局部卷积模块中强大的远程依赖关系建模能力对眼震时序数据进行特征提取,并且根据各自的特点,进行有针对性的提取。更加关注时序之间的特征,相对于CNN的特性,解决了CNN具有归纳偏置,缺乏长距离注意力的问题,通过使用CBAM有效地结合了特征层中高级和低级语义信息,并且通过空间和通道注意力将局部特征与全局特征自适应融合更好地提升了模型的分类效果。但目前的研究还存在一些问题,更进一步的研究可以聚焦于精确度的提升和更精细的眼震分类等方面,后续会加强对眩晕眼震数据特点的研究,进一步改进相关方法。

【参考文献】

[1] Edlow JA, Kerber K. Benign paroxysmal positional vertigo: a practical approach for emergency physicians[J]. Acad Emerg Med, 2023, 30(5): 579-588.

[2] Alyono JC. Vertigo and dizziness: understanding and managing fall risk [J]. Otolaryngol Clin North Am, 2018, 51(4): 725-740.

[3] Pandya S, Suresh NV, Shah V, et al. A critical appraisal of clinical practice guidelines for the diagnosis and management of benign paroxysmal positional vertigo [J]. J Laryngol Otol, 2023, 137(2): 121-126.

[4] 戴春富. 前庭医学发展现状[J]. 中国耳鼻咽喉科杂志, 2014, 14(3): 137-141.

[5] Dai CF. Development of vestibular medicine [J]. Chinese Journal of Ophthalmology and Otorhinolaryngology, 2014, 14(3): 137-141.

[6] Richburg HA, Povinelli RJ, Friedland DR. Direct-to-patient survey for diagnosis of benign paroxysmal positional vertigo [C]//2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA). Piscataway, NJ, USA: IEEE, 2018: 332-337.

[7] Fancello V, Hatzopoulos S, Santopietro G, et al. Vertigo in the elderly: a systematic literature review [J]. J Clin Med, 2023, 12(6): 2182.

[8] Eggers SDZ, Bisdorff A, von Brevern M, et al. Classification of vestibular signs and examination techniques: nystagmus and nystagmus-like movements [J]. J Vestib Res, 2019, 29(2-3): 57-87.

[9] 陈太生, 王巍, 徐开旭, 等. 良性阵发性位置性眩晕及其诊断治疗的思考 [J]. 山东大学耳鼻喉眼学报, 2019, 33(5): 1-5.

[10] Chen TS, Wang W, Xu KX, et al. Thoughts on benign paroxysmal positional vertigo and its diagnosis and treatment [J]. Journal of Otolaryngology and Ophthalmology of Shandong University, 2019, 33(5): 1-5.

[11] Lee Y, Lee S, Han J, et al. A nystagmus extraction system using artificial intelligence for video-nystagmography [J]. Sci Rep, 2023, 13(1): 11975.

[12] 周国庆, 孔玉, 高志强, 等. 后半规管和水平半规管BPPV变位实验时眼震特点初步分析 [J]. 中国现代医学杂志, 2017, 27(25): 92-94.

[13] Zhou GQ, Kong Y, Gao ZQ, et al. Characteristics of nystagmus in position test for posterior canal and horizontal canal benign paroxysmal positional vertigo [J]. China Journal of Modern Medicine, 2017, 27(25): 92-94.

[14] 张波, 孙敬武. 良性阵发性位置性眩晕患者裸眼及视频眼震图下眼震特征及定位诊断分析 [J]. 听力学及言语疾病杂志, 2012, 20(3): 235-237.

[15] Zhang B, Sun JW. The observation and diagnosis of 108 patients with benign positional paroxysmal vertigo with naked eyes and VNG [J]. Journal of Audiology and Speech Pathology, 2012, 20(3): 235-237.

[16] Yamada Y, Kobayashi M. Detecting mental fatigue from eye-tracking

- data gathered while watching video: evaluation in younger and older adults[J]. *Artif Intell Med*, 2018, 91: 39-48.
- [13] Lim EC, Park JH, Jeon HJ, et al. Developing a diagnostic decision support system for benign paroxysmal positional vertigo using a deep-learning model[J]. *J Clin Med*, 2019, 8(5): 633.
- [14] Lu W, Li ZZ, Li YN, et al. A deep learning model for three-dimensional nystagmus detection and its preliminary application [J]. *Front Neurosci*, 2022, 16: 930028.
- [15] Newman JL, Phillips JS, Cox SJ. 1D convolutional neural networks for detecting nystagmus[J]. *IEEE J Biomed Health Inform*, 2021, 25(5): 1814-1823.
- [16] Kong S, Huang ZM, Deng WK, et al. Nystagmus patterns classification framework based on deep learning and optical flow[J]. *Comput Biol Med*, 2023, 153: 106473.
- [17] Zhang WL, Wu HY, Liu Y, et al. Deep learning based torsional nystagmus detection for dizziness and vertigo diagnosis[J]. *Biomed Signal Process Control*, 2021, 68: 102616.
- [18] Wang JK, Qiao X, Liu CC, et al. Automated ECG classification using a non-local convolutional block attention module[J]. *Comput Methods Programs Biomed*, 2021, 203: 106006.
- [19] Woo S, Park J, Lee JY, et al. CBAM: convolutional block attention module [C]//Computer Vision-ECCV 2018. Cham: Springer International Publishing, 2018: 3-19.
- [20] Zhu XK, Lyu SC, Wang X, et al. TPH-YOLOv5: improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios[C]//2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW). Piscataway, NJ, USA: IEEE, 2021: 2778-2788.
- [21] 李红利, 丁满, 张荣华, 等. 基于特征融合神经网络的运动想象脑电分类算法[J]. *中国医学物理学杂志*, 2022, 39(1): 69-75.
- Li HL, Ding M, Zhang RH, et al. Motor imagery EEG classification algorithm based on feature fusion neural network[J]. *Chinese Journal of Medical Physics*, 2022, 39(1): 69-75.
- [22] Santini T, Fuhl W, Kasneci E. PuReST: robust pupil tracking for real-time pervasive eye tracking [C]//Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications. New York, NY, USA: Association for Computing Machinery, 2018: 61.
- [23] Eivazi S, Santini T, Keshavarzi A, et al. Improving real-time CNN-based pupil detection through domain-specific data augmentation[C]//Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications. New York, NY, USA: Association for Computing Machinery, 2019: 40.
- [24] Zhang Y, Li K, Li K, et al. Residual non-local attention networks for image restoration[J]. *arXiv preprint arXiv:1903.10082*, 2019.
- [25] Ma WP, Li N, Zhu H, et al. Feature split-merge-enhancement network for remote sensing object detection[J]. *IEEE Trans Geosci Remote Sens*, 2022, 60: 1-17.
- [26] Farag MM, Fouad M, Abdel-Hamid AT. Automatic severity classification of diabetic retinopathy based on DenseNet and convolutional block attention module[J]. *IEEE Access*, 2022, 10: 38299-38308.
- [27] Zhong GJ, Ding WP, Chen L, et al. Multi-scale attention generative adversarial network for medical image enhancement[J]. *IEEE Trans Emerg Top Comput Intell*, 2023, 7(4): 1113-1125.
- [28] Zhan BC, Song EM, Liu H. FSA-net: rethinking the attention mechanisms in medical image segmentation from releasing global suppressed information[J]. *Comput Biol Med*, 2023, 161: 106932.
- [29] Duchi J, Hazan E, Singer Y. Adaptive subgradient methods for online learning and stochastic optimization[J]. *J Mach Learn Res*, 2011, 12: 2121-2159.

(编辑:薛泽玲)