

基于多尺度特征提取和时间分割的梅杰综合征检测

李碧草¹, 衣本泽¹, 王贝², 刘志涛³, 郭旭伟⁴, 王岩¹

1. 中原工学院信息与通信工程学院, 河南 郑州 450007; 2. 中原工学院附属医院, 河南 郑州 451191; 3. 河南省直第三人民医院梅杰诊疗中心, 河南 郑州 450018; 4. 河南科技大学第一附属医院儿科, 河南 洛阳 471000

【摘要】梅杰综合征的诊断主要依赖医生的临床评估, 由于该疾病的症状与其他神经性疾病有相似之处且确诊较为复杂, 因此诊断过程对医生和患者相当关键。本研究采集31名梅杰综合征患者的发病视频, 建立梅杰综合征检测数据集, 开发出一种应用于未修剪视频的梅杰综合征自动检测系统(MS-Net)。首先, 利用RetinaNet和UNet3+构建时间检测分支和时间分割分支来进行多尺度特征提取和时间分割; 其次, 通过时间检测分支和时间分割分支解码分别生成检测窗口的概率向量和每帧发病的概率; 最后, 使用多层感知器处理两分支的概率预测, 为每个窗口生成一个更加准确的概率。通过使用额外的损失函数和数据增强等技术优化模型性能, 并使用临床医生可解释特征进行操作。MS-Net可以辅助诊断梅杰综合征, 提高早期诊断的准确性、便捷性和效率。与其他先进的网络进行比较, 结果表明MS-Net在使用临床实践中所需的可解释特征的同时在平均精度方面也取得相当的性能。

【关键词】梅杰综合征; 时序动作检测; 多尺度特征提取; 时间分割

【中图分类号】R318

【文献标志码】A

【文章编号】1005-202X(2025)07-0962-07

Detection of Meige's syndrome based on multi-scale feature extraction and temporal segmentation

LI Bicao¹, YI Benze¹, WANG Bei², LIU Zhitao³, GUO Xuwei⁴, WANG Yan¹

1. School of Information and Communication Engineering, Zhongyuan University of Technology, Zhengzhou 450007, China; 2. University Infirmary, Zhongyuan University of Technology, Zhengzhou 451191, China; 3. Diagnosis and Treatment Center of Meige's Syndrome, the Third People's Hospital of He'nan Province, Zhengzhou 450018, China; 4. Pediatric Department, the First Affiliated Hospital of He'nan University of Science and Technology, Luoyang 471000, China

Abstract: The diagnosis of Meige's syndrome predominantly relies on the clinical assessment by physicians. Given the complexity and similarity of its symptoms to other neurological disorders, the diagnosis is crucial for both doctors and patients. Herein a detection dataset for Meige's syndrome is compiled from video recordings of 31 patients, and an automated diagnostic system for Meige's syndrome (MS-Net) applicable to untrimmed videos is developed. The system utilizes RetinaNet and UNet3+ to construct temporal detection and segmentation branches for multi-scale feature extraction and temporal segmentation, obtains probability vectors for detection windows and the probability of disease onset per frame via the decoding of temporal detection and segmentation branches, and finally generates a refined probability for each window by processing the probability predictions from both branches using a multi-layer perceptron. The model performance is optimized using additional loss functions and data augmentation techniques, operating on features interpretable by clinical physicians. MS-Net can assist in the diagnosis of Meige's syndrome, improving the accuracy, convenience, and efficiency of the early diagnosis. The comparison of MS-Net with other state-of-the-art networks indicates that MS-Net achieves comparable performance in terms of average precision while utilizing interpretable features required in clinical practice.

Keywords: Meige's syndrome; temporal action detection; multi-scale feature extraction; temporal segmentation

【收稿日期】2024-12-14

【基金项目】国家自然科学基金(61901537); 河南省高校科技创新人才支持计划(23HASTIT030); 中原工学院学科青年硕博培育计划(SD202207); 中原工学院自然科学基金(K2025ZD006); 中原工学院研究生科研创新计划(YKY2024ZK26)

【作者简介】李碧草, 博士, 副教授, 研究方向: 人工智能与医学图像处理, E-mail: lbc@zut.edu.cn

前言

目标检测作为计算机视觉领域内一项基础且充满挑战的任务, 其重要性在安全监控、自动驾驶和医疗检测等实际应用中日益凸显, 这一技术的核心目标是在图像或视频帧中识别并定位特定类别的语义对象实例^[1-3]。得益于深度学习的进步, 尤其是神经网络结构的创新和图形处理单元计算能力的提

升,目标检测的性能得到革命性的增强。深度学习模型已被广泛应用于从输入数据中提取特征、进行分类和定位,从而推动目标检测技术的发展,这些模型通常作为目标检测系统中的骨干网络,负责特征提取,而检测网络则利用这些特征来预测对象的类别和边界框。目标检测的研究不仅限于通用对象的识别,还涵盖特定领域目标的检测,如场景文本识别、人脸检测和行人识别等。这些研究方向不仅拓展目标检测的应用范围,也提高检测系统的准确性和鲁棒性。此外,目标检测技术在特定应用场景下的重要性也日益增加,如在安全领域中对可疑行为的识别、在交通管理中车辆流量的监控、在医疗影像分析中对病变的定位以及在军事侦察中对特定目标的跟踪等,这些应用体现了目标检测技术在现实世界问题解决中的潜力和价值。

受上述研究的启发,本研究运用目标检测技术对梅杰综合征进行分析。梅杰综合征是一种罕见的节段性肌张力障碍性疾病,又称特发性眼睑痉挛-口下颌肌张力障碍综合征,属成人多动症的一种^[4],其特征包括眼睑痉挛、睁眼困难,可能伴有口及下颌的不自主运动,症状可能在疲劳、日光刺激、注视、紧张时加重,而在睡眠时消失,此外,患者可能伴有精神疾患、牙科疾患、头颈部的肌张力障碍等,通常在40~70岁人群中发病,女性患者数量略多于男性;病因目前尚不完全明确,但可能与心理因素、药物(如抗精神病药物、抗震颤麻痹药物、抗组胺药物等)、颅脑创伤、脑炎等导致的脑内神经介质失衡有关。由于梅杰综合征的罕见性和诊断难度,患者常常难以确诊,一个易于部署的自动检测系统可以为患者和医生提供帮助,降低诊断的难度。少有研究从视频数据中解决检测问题,如Bernabei等^[5]使用放置在患者躯干上的三轴加速计的信号,通过分类记录信号的2 s非重叠窗口来实现检测;Shute等^[6]使用脑深层刺激手术(一种侵入性的神经调节疗法)收集并分类脑深层刺激信号的分段时间窗口,采用支持向量机进行分类;Elakkiya等^[7]使用LeNet-5来分类无线信道信息信号的分段时间窗口;Wu等^[8]研究是少数利用视频数据进行检测的研究之一,使用类似于SimCLR^[9]的无监督学习框架,从分段的1 s视频剪辑中提取鲁棒的视频特征,然后将这些特征输入到长短时记忆网络进行分类。尽管这些研究做出过杰出的贡献,但有两个主要缺点:(1)大多数研究依赖于复杂设备设置的数据,很难在普通临床环境或患者家中实施;(2)所有先前的研究都是在定位良好的时间窗口上进行分类,忽略长期的时间相关性,限制了检测的准确性和使用场景。为缓解上述方法的局限

性,提高工具的应用性和准确性,本研究提出使用智能手机录制的视频,此外,由于面部发病在梅杰患者中非常常见,因此本研究提出MS-Net架构,在未修剪的视频中进行梅杰综合征检测。

视频中的面部检测是活动检测的一个特殊实例。活动检测是视频分析领域的一项重要且具有挑战性的任务,在视频推荐、智能监控、机器人等领域都有实际应用。得益于计算能力的提高和深度学习的巨大成功以及包含数千个带注释视频的大规模公共数据集的可用性,已经提出许多成功的时间动作检测方法^[10-11]。Lin等^[12]开发出一种边界敏感的方法(边界匹配网络),用于获取候选行动建议;He等^[13]提出时间上下文聚合网络,使用局部和全局时间上下文逐步获得准确的动作边界和可靠的概率预测。对于梅杰综合征检测,还存在非常大的挑战:(1)由于患者的隐私和机密性,公开可用的视频很少,也没有注释,而大多数深度学习方法需要大型数据集进行训练;(2)梅杰综合征的面部发病在持续时间上变化很大,在患者之间和患者内部的表达也不尽相同,尽管它们可能属于同一类别,例如不同患者的“眨眼”可能在持续时间、重复次数和力度上都有所不同,甚至在同一个人身上每周都有所不同;(3)缺乏可解释性是将当前深度学习技术应用于梅杰综合征检测的另一个障碍,因为患者和临床医生可能无法在深度学习的背景下解释结果。

本研究的MS-Net可以有效解决从智能手机获得的未修剪视频中的梅杰综合征检测问题,主要贡献如下:(1)提出一种全新的时间动作检测应用,即运用多尺度特征提取和时间分割检测梅杰综合征,这是在梅杰综合征上的首次应用;(2)构建出首个针对梅杰综合征的数据集,包括患者面部的发病位置与强度;(3)使用临床有意义的面部肌肉单元运动强度作为输入特征,不仅可以获得临床可解释的检测结果,还能保持良好的性能。

1 方法

1.1 网络构造

本研究提出的MS-Net研究总体分为3部分:时间检测分支、时间分割分支和最终的精度融合器,它们共享相同的编码器与精度融合器,开创性地结合来自两个分支的概率预测以生成最终检测的精细概率(图1)。时间分割分支是一种编码器-解码器架构,将面部肌肉运动单元强度序列作为输入^[14],并预测梅杰综合征患者每帧发病的概率;因此需要进行一些后处理,将分割概率转化为检测概率。时间检测分支通过编码器-解码器架构直接产生具有相关检测概

率的时间窗口。MS-Net通过融合分割和检测概率以获得两全其美的效果,具体来说,通过使用具有少量层的紧凑体系结构以及各种数据增强技术来减轻对大型训练集的需求。然后,MS-Net使用k-means聚类来获得更合适的锚大小,并使用额外的回归损失,即

有效交集超过联合损失(Efficient-IoU, EIoU)^[15]来筛选不精确的预测窗口,从而获得可变速持续时间的精确窗口。最后,MS-Net使用面部肌肉单元运动强度作为特征,产生可解释的检测。

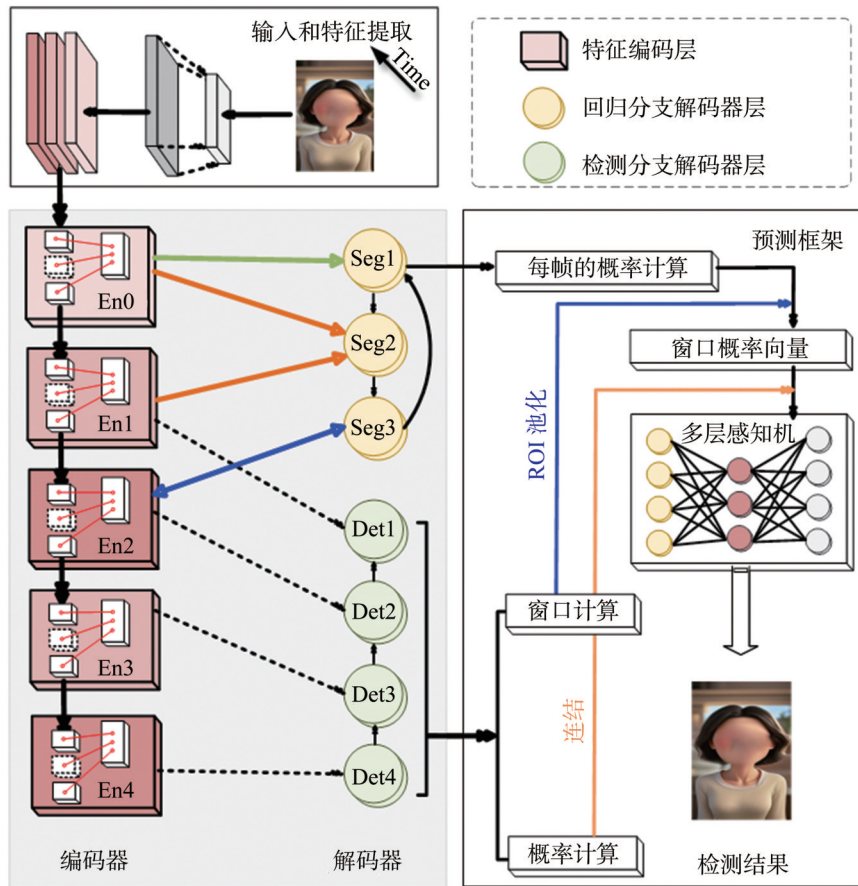


图1 网络整体构架

Figure 1 Overall network architecture

1.2 时间动作检测分支

时间动作检测的目标是在视频中检测到一个动作的实例。目标检测器可以根据其操作方式分为两大类:(1)两阶段检测器采用提议-分类方案,首先通过提议生成网络生成多个候选边界框,然后对其进行分类^[16-17]; (2)单阶段检测器绕过提议生成阶段,直接检测对象^[18-19]。虽然两级检测器通常更准确,但一级检测器通常更快更有效。

时间动作检测方法遵循与目标检测方法类似的方法,只是2D边界框被时间窗口取代。具体来说两阶段方法学习一个与类别无关的时间动作建议生成器,然后将生成的建议反馈到动作分类器,产生最终的检测结果;单阶段方法联合学习动作建议和分类网络,效率更高^[20-21]。由于本研究的目标是开发一个紧凑模型,可以从有限数据中进行训练,本研究的

MS-Net 检测分支使用一个来自 RetinaNet 的单阶段时间动作检测器^[22],见图2。

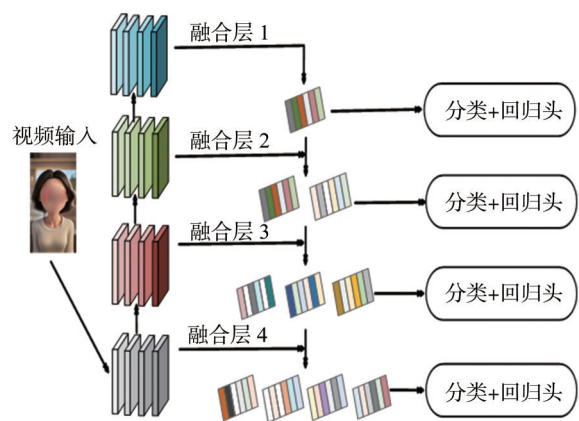


图2 时间动作检测分支架构

Figure 2 Temporal action detection branch architecture

1.3 时间动作分割分支

时间动作分割的目标是预测视频中每一帧的动作标签,需要捕获帧之间的短期和长期关系,捕获这种时间相关性的主流模型包括隐马尔可夫模型^[23]、条件随机场^[24]、循环神经网络^[25]和时间卷积网络^[26],然而,这些方法在训练和推理过程中的计算复杂性限制其对短距离相关的适用性,可能导致过度分割。多阶段时间卷积网络通过使用多阶段的细化来缓解这一问题^[27-28];然而,这种网络的参数非常多,在有限的数下无法获得良好的性能。时间语义分割网络通过使用多尺度编码器-解码器架构缓解这一问题,该架构也可以捕获帧之间的长期相关性,因此有可能在帧级产生更准确的发病概率估计。

本研究使用时态语义分割网络作为MS-Net的分割分支。采用UNet3+架构^[29],该架构最初是为医学图像分割任务提出的(图3)。UNet3+通过其全尺寸跳跃连接机制能有效捕获多尺度的语义信息,这对于分割不同尺寸的器官至关重要。鉴于梅杰综合征发作的持续时间具有显著的多变性,该架构保留多尺度细节特征的能力,在时序UNet3+模型中同样得到继承,从而显著提升梅杰综合征发作事件的分割性能。

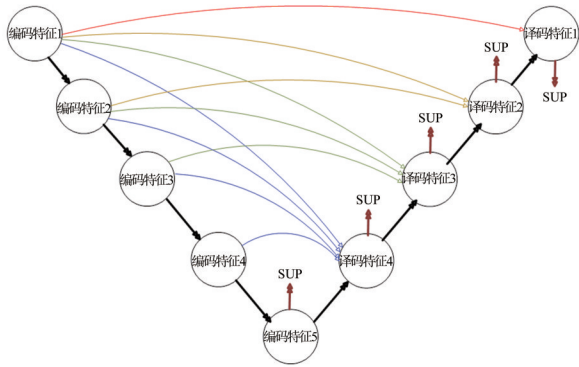


图3 UNet3+架构

Figure 3 UNet3+ architecture

1.4 网络总体检测策略

在临床实践中,临床医生通过评估患者运动的力度、突发性和位置来判断面部,这包括肌肉激活强度的组合,涉及肌肉的位置和组合以及这些激活之间的时间关系。面部动作单元是面部动作编码系统用于描述面部肌肉运动的基本单元。每一个动作单元都对应着一组特定的面部肌肉或肌肉群的活动,这些活动可以独立地产生面部表情的变化。由于动作单元代表面部肌肉或肌肉群的基本运动,因此动作单元可以捕获临床医生用于识别发病的特征和空

间特征,这说明使用动作单元强度作为特征的模型将更容易被临床医学所解释,而不像深度神经网络提取的视觉特征难以解释。

设 $A = \{a_l\}_{l=1}^L$ 表示帧长为 L 的未修剪视频,其中, a_l 为第 l 帧。将 A 的检测注释表示为 $Z = \{z_n\}_{n=1}^N$, 其中, $Z_n = (t_n^s, t_n^e)$ 是一个检测实例,其开始帧为 t_n^s , 结束帧为 t_n^e 。对于分割任务,将 Z 转换为标签序列 $S = \{S_l\}_{l=1}^L$, 如果 $l \in [t_n^s, t_n^e]$, $S_l=1$ (发病), 否则 $S_l=0$ (正常)。

本研究采用双流式网络TSN^[30]直接从原始视频中提取深度特征。该网络由两个卷积框架组成,分别处理RGB和光流数据,然后进行全局平均池化操作。连接两个数据流的输出,得到帧 a_l 的深度特征 $f_l \in \mathbb{R}^C$, 其中 $C=3\ 072$ 是特征的维数。为提高解释性也可以考虑可解释的特征,即使用OpenFace^[31]从视频中提取的面部肌肉单元运动强度。OpenFace提取17个面部肌肉单元运动的强度,得到帧 a_l 的特征 $f_l \in \mathbb{R}^C$, 其中 $C=17$ 为动作单元强度的维数。因此,给定一个未经修剪的视频 A , 可以提取出一个特征序列 $F = \{f_l\}_{l=1}^L \in \mathbb{R}^{C \times L}$, 其中, $C \in \{C^v, C^a\}$ 。

由于未修剪的视频具有可变的时间长度,并且通常太大而无法装入内存进行计算。因此本研究将使用重叠滑动窗口将一个长度为 L 的视频分成 X 个固定长度为 T 的视频片段。使用 $T=416$ 帧的长观测窗口,得到 X 个特征序列 $\{F_x\}_{x=1}^X$, 其中 $F_x = \{f_l\}_{l=1}^T \in \mathbb{R}^{C \times T}$ 。

本研究将RetinaNet改编为一个时间动作检测器用于检测。使用ResNet-18作为共享编码器,并在其之上构建一个4级多尺度时间特征网络FPN^[32]作为检测分支的解码器。在FPN中,将每个ResNet残差阶段的输出输入到FPN的相应阶段,计算多尺度输出。然后,对FPN的输出应用一个两分支单层卷积器,分别进行分类和时间边界窗口回归。综上所述,检测分支以 B_D 特征序列 F_x 作为输入和输出具有相应概率的检测窗口,即:

$$B_D(F_x) = D_x = \{b_i, c_i\}_{i=1}^{N_b}, b_i = (\hat{t}_i^s, \hat{t}_i^e) \quad (1)$$

其中, b_i 表示第 i 个检测窗口, \hat{t}_i^s 和 \hat{t}_i^e 分别为起始帧和结束帧, c_i 表示 b_i 为发病实例的概率。

该系统在时序ResNet-18的基础上构建一个4层TFPN,随后是一个执行分类和回归的双分支卷积器。本研究将UNet3+应用于时间分割,UNet3+通过在编码器中结合相同和较小规模阶段的特征图以及在解码器中结合较大规模的输出来利用全尺寸信息。在

ResNet-18(共享编码器)的前3个阶段之上构建一个3阶段的UNet3+,即预残差阶段和前两个ResNet残差阶段。分割分支 B_s 也以特征序列 F_x 为输入,输出各阶段每帧的概率预测,即:

$$B_s(F_x) = \{P_x^k\}_{k=1}^K, P_x^k = \{P_j^k\}_{j=1}^T \quad (2)$$

其中, $K=3$ 表示UNet3+的阶段数, P_j^k 表示UNet3+解码器的第 K 阶段第 j 帧为发病的概率。

在精度融合器中,MS-Net将两个分支的概率预测启发式地结合起来以得到精确的概率。首先,根据 $\{b_i\}_{i=1}^{N_p}$,对UNet3+解码器的最低尺度阶段(即 p_x^1)的输出执行1D RoIPooling,以提取每个检测窗口 b_i 的概率向量 V_i 。然后,将这个概率向量与来自检测分支的概率预测 c_i 连接起来。最后,用多层感知器(MLP)对拼接结果进行处理,得到精细化的概率 \tilde{c}_i ,形成精细化的检测 \tilde{D}_x 。具体可以表示为:

$$\tilde{c}_i = \text{MLP}(V_i \oplus c_i), V_i = \phi(b_i, P_x^1), i = 1, \dots, N_p \quad (3)$$

$$B_H(D_x, P_x^1) = \tilde{D}_x = \{b_i, \tilde{c}_i\}_{i=1}^{N_p}, b_i = (\hat{t}_i^s, \hat{t}_i^e) \quad (4)$$

其中, \oplus 表示串联, ϕ 表示一维RoIPooling操作。

对于锚点,按照YOLO9000^[33]的方法使用k-means聚类来生成合适的锚大小。如果锚点的有效交集大于0.3,则锚点与真值注释(正)匹配;如果锚点的有效交集小于-0.4,则锚点与背景(负)匹配。有效交集超过EIoU是IoU的精细化版本,不仅考虑IoU本身,还考虑锚点与真值注释之间的中心点距离和长度差异。具体来说,可以将锚点表示为:

$$A = (t_A^s, t_A^e), m_A = (t_A^s + t_A^e)/2, d_A = t_A^e - t_A^s \quad (5)$$

其中, t_A^s 和 t_A^e 分别表示该锚点的起始帧和结束帧, m_A 表示该锚点的中心点, d_A 表示长度。同时也可以表示一个中心点为 m_c ,长度为 d_c 的真实注释 $G = (t_c^s, t_c^e)$,以及最小的封闭窗口 $C = (t_c^s, t_c^e)$,其中 $t_c^s = \min\{t_A^s, t_c^s\}$ 和 $t_c^e = \max\{t_A^e, t_c^e\}$,中心点为 m_c ,长度为 d_c 。

最后,锚点与注释之间的有效交集超过EIoU可以表示为:

$$\text{EIoU} = \text{IoU} - \frac{(m_A - m_c)^2}{d_c^2} - \frac{(d_A - d_c)^2}{d_c^2} \quad (6)$$

其中, m_A 、 m_c 、 d_A 和 d_c 的含义与式(5)相同。

1.5 损失函数与参数设置

对于检测分支,本研究首先按照单步多框目标检测的方法进行在线难负例挖掘,使用的正负样本比例为1:3,以解决不平衡问题,从而提高检测性能。然后,应用二进制交叉熵损失对所有正向和负向进行分类,并应用SmoothL1(sl)损失对所有正向进行回归,此外将有效交集超过EIoU损失也用于回归,这使得检测尽可能与基准真实值注释相似。检测分支的

总损失可以表示为分类损失与回归损失的加权和,即:

$$L_x^D = \frac{1}{N_p} \sum_i L_{x,bce}(\tilde{c}_i, \tilde{y}_i) + \gamma(L_{x,sl} + \lambda L_{x,EIoU})(b_i, y_i) \quad (7)$$

其中,如果生成 b_i 的锚匹配到一个基准真实值(正向),则 $\tilde{y}_i = 1$,否则 $\tilde{y}_i = 0$ 。其中 $y_i = (t_i^s, t_i^e)$ 表示基准真实值。 N_p 是正向值。对于超参数设 $\gamma=2, \lambda=1.5$ 。

对于分割分支,采用focal loss和Dice loss,并对UNet3+中所有阶段的预测概率序列进行深度监督,即 $\{P_x^k\}_{k=1}^K$ 。分割分支的损失可以表示为:

$$L_x^S = \frac{1}{T} (L_{x,fl} + L_{x,dice})(\{P_x^k\}_{k=1}^K, \{S_j\}_{j=1}^T) \quad (8)$$

最后,MS-Net的总损失可以表示为检测分支损失 L_x^D 、分割分支损失 L_x^S 和 L_2 正则化项的加权和:

$$L_x = L_x^D + \alpha L_x^S + \beta L_2 \quad (9)$$

其中, $\alpha=3, \beta=0.001$ 。

2 实验结果与分析

2.1 实验设置和评价指标

本研究所提出的网络在公共的PyTorch平台上实现。训练和测试使用CUDA的11.3版本,并在一个具有80GB内存的NVIDIA A100 GPU上进行训练。训练轮次为149,来自所有多尺度的检测结果通过一个概率阈值为0.2的标准进行合并,然后应用EIoU的非极大值抑制,阈值设为0.2,得到最终的检测结果。

为定量评估检测性能,计算平均精度(Average Precision, AP)作为主要参考指标,用于衡量目标检测系统在不同召回率水平上的AP,它是在特定召回率阈值下计算精度的平均值。以IoU阈值0.5和概率阈值0.2计算,即只有预测概率(该检测为发病)大于0.2的检测才会被考虑,如果检测与真实注释之间的IoU大于0.5,则认为发病已被检测。实验遵循两种交叉验证策略:从12个视频中随机挑选1个视频(方法一)和从12个视频中随机挑选3个视频(方法二)用作测试。

2.2 梅杰综合征检测数据集

本研究对梅杰综合征患者的面部运动进行分析。在每次访问中为31名不同的梅杰综合征患者录制31个2min的视频。在每次访问中,他们均进行无意识的面部行为。所有视频均使用30帧每秒的智能手机进行拍摄。梅杰综合征检测数据集主要分为两个部分(VIA梅杰综合征检测数据集和OpenFace梅杰综合征检测数据集)。VIA梅杰综合征检测数据集中包含面部发病的时间边界窗口和相应的运动种类(眼睑抽搐、耸鼻子、皱眉等),OpenFace梅杰综合征检测数据集包含面部各个肌肉运动的强度。

使用开源软件 VIA^[34](VGG Image Annotator)对时间边界窗口进行标注。开源软件 VIA 是一个简单且独立的手动标注工具,用于图像、音频和视频,可以在网络浏览器中运行,不需要任何安装或设置。它是一个轻量级、独立且离线运行的软件包,完整软件包含在一个单一的 HTML 页面中,可以在大多数现代网络浏览器中作为离线应用程序运行。VIA 支持多种区域形状的标注,包括矩形、圆形、椭圆、多边形、点和折线,允许导入和导出区域数据为 CSV 和 JSON 文件格式。VIA 同时支持视频标注,用户可以通过选择视频帧中的区域并添加注释来标记视频中的对象或事件,视频标注功能包括播放控制、时间轴导航、区域标注和属性编辑等。每个视频首先由两个训练有素的注释者标记,随后一名高级研究员检查这两组注释并解决所有不一致的问题。

OpenFace 作为一个先进的面部行为分析工具,在梅杰综合征检测领域的应用展现其强大的潜力。本研究利用 OpenFace 深度学习模型来分析梅杰综合征患者的发病行为,以生成梅杰综合征患者面部每一帧的数据,这些数据记录面部肌肉运动位置和抽搐强度相关的详细信息。OpenFace 的应用使得从视频数据中提取关键的面部动作单元成为可能,这些动作单元是理解患者行为的关键,也是诊断和检测梅杰综合征的重要指标,MS-Net 通过对数据的分析,能生成并应用具备临床可解释性的特征。

2.3 消融实验和结果分析

本研究对基于 YOLOV5 和 YOLOV7 的检测系统和 MS-Net 进行全面的评估。YOLOV5 和 YOLOV7 利用从未经处理的视频数据中提取的深度特征,分别实现 21.11% 和 33.43% 的 AP;MS-Net 取得 34.69% 的 AP,优于这两种方法。此外,MS-Net 是基于面部肌肉单元运动强度构建的,在性能上超越那些依赖深层特征作为输入的方法,验证了 MS-Net 的有效性,也表明了使用面部肌肉单元运动强度作为特征替代深层特征可以在不牺牲性能的前提下增强模型的可解释性。进一步的分析揭示了方法一相较于方法二在检测性能上的显著提升,这一发现强调了数据量对于提高检测系统性能的重要性。方法一的 AP 相较于方法二提升了 6.02%。

为深入探究 MS-Net 的结构对性能的影响,本研究进行消融分析,比较在 MS-Net 中添加分割分支与仅使用检测分支的效果(表 1)。实验数据表明,在方法一和方法二中,添加分割分支分别提升了 11.31% 和 11.96% 的 AP,证实了分割分支对于提高 MS-Net 性能的重要性。

表 1 消融实验

Table 1 Ablation study

特征选取	方法	方法一	方法二
面部肌肉单元运动强度	MS-Net 仅检测分支	23.38%	16.71%
面部肌肉单元运动强度	MS-Net 包含分割分支	34.69%	28.67%

3 结语

本研究首先实现梅杰综合征的自动检测,即用于在未修剪的视频中自动检测梅杰综合征的网络 MS-Net,并结合分割概率来优化检测概率。MS-Net 通过促进便捷性和提供准确的检测,帮助解除梅杰综合征诊断的障碍。MS-Net 由于所使用的视频是由智能手机录制的,具有较高的可实现性。实验结果证明,MS-Net 能达到与使用深度特征的基线方法相当的性能。最后,MS-Net 由于使用可解释的面部肌肉单元运动强度作为输入特征,并捕获面部肌肉单元运动强度,而具有极高的可解释性。

【参考文献】

- [1] Chen W, Luo JJ, Zhang F, et al. A review of object detection: datasets, performance evaluation, architecture, applications and current trends[J]. *Multimed Tools Appl*, 2024, 83(24): 65603-65661.
- [2] Contreras M, Jain A, Bhatt NP, et al. A survey on 3D object detection in real time for autonomous driving[J]. *Front Robot AI*, 2024, 11: 1212070.
- [3] Zhu ML, Gong YD, Tian CW, et al. A systematic survey of transformer-based 3D object detection for autonomous driving: methods, challenges and trends[J]. *Drones*, 2024, 8(8): 412.
- [4] 郑为娜,徐振华,庞海英.带你认识梅杰综合征[J].*人人健康*, 2024(13): 13-14.
Zheng WN, Xu ZH, Pang HY. A introducing Meige syndrome[J]. *Health for Everyone*, 2024(13): 13-14.
- [5] Bernabei M, Andreoni G, Mendez Garcia MO, et al. Automatic detection of tic activity in the Tourette syndrome[C]//2010 Annual International Conference of the IEEE Engineering in Medicine and Biology. Piscataway, NJ, USA: IEEE, 2010: 422-425.
- [6] Shute JB, Okun MS, Opri E, et al. Thalamocortical network activity enables chronic tic detection in humans with Tourette syndrome[J]. *Neuroimage Clin*, 2016, 12: 165-172.
- [7] Elakkiya MK, Deje. Novel deep learning models with novel integrated activation functions for autism screening: AutoNet and MinAutoNet[J]. *Expert Syst Appl*, 2024, 238(Part D): 122102.
- [8] Wu JY, Zhou TS, Guo YF, et al. Video-based evaluation system for tic action in Tourette syndrome: modeling, detection, and evaluation[J]. *Health Inf Sci Syst*, 2023, 11(1): 39.
- [9] Yang XY, Zhang ZG, Cui RY. TimeCLR: a self-supervised contrastive learning framework for univariate time series representation[J]. *Knowl Based Syst*, 2022, 245: 108606.
- [10] Zhao F, Wang W, Wu Y, et al. A coarse-to-fine temporal action detection method combining light and heavy networks[J]. *Multimed Tools Appl*, 2023, 82(1): 879-898.
- [11] Shen JY, Li M, Zhang JK. Temporal action detection methods based on deep learning[J]. *Intern J Pattern Recognit Artif Intell*, 2022, 36(3): 2252005.
- [12] Lin TW, Liu X, Li X, et al. BMN: boundary-matching network for temporal action proposal generation[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway, NJ, USA: IEEE, 2019: 3888-3897.

- [13] He YL, Zhong Y, Wang LS, et al. GLFormer: global and local context aggregation network for temporal action detection[J]. Appl Sci, 2022, 12(17): 8557.
- [14] Sanchez-Mendoza D, Masip D, Lapedriza À. Emotions classification using facial action units recognition[C]//Museros L, Pujol O, Agell N. Frontiers in artificial intelligence and applications. Amsterdam, Netherlands: IOS Press, 2014: 55-64.
- [15] Peng HY, Yu SQ. A systematic IoU-related method: beyond simplified regression for better localization[J]. IEEE Trans Image Process, 2021, 30: 5032-5044.
- [16] Akila Agnes S, Anitha J, Arun Solomon A. Two-stage lung nodule detection framework using enhanced UNet and convolutional LSTM networks in CT images[J]. Comput Biol Med, 2022, 149: 106059.
- [17] Jeong NH, Choi JH, Lee G, et al. Feature selection for SAR target discrimination and efficient two-stage detection method[J]. Remote Sens, 2022, 14(16): 4044.
- [18] 李亚飞, 刘娜, 周惠, 等. 运动场景下小目标检测算法设计[J]. 仪表技术, 2024(3): 1-6.
Li YF, Liu N, Zhou H, et al. Design of small target detection algorithm in motion scenes[J]. Instrumentation Technology, 2024 (3): 1-6.
- [19] Kim H, Lee D, Kim C, et al. Fashion image analysis using single-stage detector[C]//2020 International Conference on Image, Video Processing and Artificial Intelligence. Bellingham, WA, USA: SPIE, 2020: 1158418.
- [20] Wu YH, Chang XY, Chen DY, et al. Two-stage salient object detection based on prior distribution learning and saliency consistency optimization[J]. Vis Comput, 2023, 39(11): 5729-5745.
- [21] Ji CF, Liu GZ, Zhao D. ETS-3D: an efficient two-stage framework for stereo 3D object detection[J]. J Vis Commun Image Represent, 2022, 88: 103634.
- [22] Xiong YY, Meng W, Yan JW, et al. A rotation-invariance face detector based on RetinaNet[J]. J Phys Conf Ser, 2023, 2562(1): 012066.
- [23] Goumiri S, Benboudjema D, Pieczynski W. A new hybrid model of convolutional neural networks and hidden Markov chains for image classification[J]. Neural Comput Appl, 2023, 35(24): 17987-18002.
- [24] Ghasemi-Naraghi Z, Nickabadi A, Safabakhsh R. Towards reliable multi-person pose estimation using conditional random fields[J]. Pattern Recognit Lett, 2023, 175: 59-65.
- [25] Tang CL, Zhang D, Tian QC. Convolutional neural network-bidirectional gated recurrent unit facial expression recognition method fused with attention mechanism[J]. Appl Sci, 2023, 13(22): 12418.
- [26] Chen FS, Zhang YF, Chen L, et al. Dynamic traveling time forecasting based on spatial-temporal graph convolutional networks [J]. Front Comput Sci, 2023, 17(6): 176615.
- [27] Wang CZ, Kumar TS, De Raedt W, et al. Drinking gesture detection using wrist-worn IMU sensors with multi-stage temporal convolutional network in free-living environments[C]//2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). Piscataway, NJ, USA: IEEE, 2022: 1778-1782.
- [28] Goldbraikh A, Shubi O, Rubin O, et al. MS-TCRNet: multi-stage temporal convolutional recurrent networks for action segmentation using sensor-augmented kinematics[J]. Pattern Recognit, 2024, 156: 110778.
- [29] Zhang KH, Chen J, Liu B, et al. Deep object co-segmentation via spatial-semantic network modulation[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 12813-12820.
- [30] Liu Z, He XL, Lu YH. Combining UNet 3+ and transformer for left ventricle segmentation via signed distance and focal loss[J]. Appl Sci, 2022, 12(18): 9208.
- [31] Santoso K, Kusuma GP. Face recognition using modified OpenFace [J]. Procedia Comput Sci, 2018, 135: 510-517.
- [32] Sun HT, Yang S, Chen LJ, et al. Brain tumor image segmentation based on improved FPN[J]. BMC Med Imaging, 2023, 23(1): 172.
- [33] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2017: 6517-6525.
- [34] Dutta A, Zisserman A. The VIA annotation software for images, audio and video[C]//Proceedings of the 27th ACM International Conference on Multimedia. New York, NY, USA: ACM, 2019: 2276-2279.

(编辑:谭斯允)