

基于瞳孔追踪的网络课堂注意力检测

邓景云,赵丁皓,秦慧伶,胡锦涛,钱程,陈雅茜

(西南民族大学计算机与人工智能学院,四川成都610041)

摘要:近年来,基于瞳孔识别的注意力检测技术因其无感感知和低成本优势成为了智慧教育领域的创新解决方案。然而,由于眼睛特征的个体差异、不断变化的光照条件、多样的头部姿势以及眼动方向的复杂性等因素,现有追踪算法在实时性和精确度方面还有很大提升空间。同时,如何有效地从视线追踪数据中推断出注意力仍是一个待解决的关键问题。因此,提出了基于瞳孔追踪的注意力检测方法,首先对瞳孔数据集进行预处理,采用 Adaboost 算法和 OTSU 动态阈值获取瞳孔中心区域位置特征并结合分类和回归损失组合模型实现实时视线追踪,然后基于分割网络和决策网络模型,利用眼睛开合度和注视离散度来综合判断注意力集中程度。实验结果显示,所提方法在实时视线追踪的精度和处理速度方面均优于现有算法,且在眼睛开合度预测方面达到了 96.72% 的高准确率。

关键词:瞳孔追踪;注意力检测;深度学习

中图分类号:TP391.41;TP18

文献标志码:A

文章编号:2095-4271(2025)02-0206-10

Attention detection in online classroom based on pupil tracking

DENG Jingyun, ZHAO Dinghao, QIN Huiling, HU Jinzong, QIAN Cheng, CHEN Yaxi

(School of Computer and Artificial Intelligence, Southwest Minzu University, Chengdu 610041, China)

Abstract: In recent years, attention detection technology based on pupil recognition has emerged as an innovative solution in the field of smart education, owing to its advantages of unobtrusive perception and low cost. However, due to factors such as individual differences in eye features, constantly changing lighting conditions, diverse head postures, and the complexity of eye movement directions, there is still significant room for improvement in the real-time performance and accuracy of existing tracking algorithms. Meanwhile, effectively inferring attention from gaze tracking data remains a key issue that needs to be addressed. Therefore, a pupil-tracking-based attention detection method is proposed. Firstly, the pupil dataset is preprocessed, and the Adaboost algorithm and OTSU dynamic threshold are used to obtain the positional features of the pupil center region. A combined classification and regression loss model is then employed to achieve real-time gaze tracking. Subsequently, based on segmentation and decision network models, eye opening and closing degrees, as well as gaze dispersion, are utilized to comprehensively assess the level of attention concentration. Experimental results show that the proposed method outperforms existing algorithms in terms of both the accuracy and processing speed of real-time gaze tracking, and achieves a high accuracy rate of 96.72% in predicting eye opening and closing degrees.

Keywords: pupil tracking; attention detection; deep learning

在信息化技术迅速发展的时代背景下,线上线下混合式教学模式正逐渐成为推动教育向更高质量发展

的重要引擎。然而在线学习过程中,由于缺乏实时的监督和反馈,导致过程性学习效果难以评估。因此,

收稿日期:2024-09-16

通信作者:陈雅茜(1981-),女,副教授,博士,研究方向:人机交互、信息可视化、深度学习。E-mail:yaxichen@swun.cn

基金项目:西南民族大学中央高校优秀学生培养工程项目(2023NYXXS041,ZYN2024116);西南民族大学横向项目(横20240096)

注意力检测系统应运而生,通过视线追踪技术来实时监测学生的注意力状态.其中基于瞳孔识别的检测系统因无感知和低成本优势备受关注.但在实际应用中,该类技术面临着诸多挑战:眼睛外观的个体差异使得系统难以准确地识别每个用户的眼睛特征;光照条件的变化会严重影响视线追踪的准确性;头部姿势的多样性和双眼注视方向的灵活性也增加了视线追踪的难度.针对上述问题,本文提出一种基于瞳孔追踪的注意力检测方法,通过提取瞳孔中心区域位置特征并结合分类与回归损失(Losses and Classification, LC)组合模型,进行实时视线追踪.然后基于改进的分割和决策网络(Segmentation Network and Decision Network, SNDN)模型,利用眼睛开合度和注视离散度来综合判断注意力集中程度.

1 相关工作

视线追踪算法是一种计算机视觉技术,能够精确测定个体的视线方向,从而推测其注视点在屏幕或空间中的位置.自20世纪60年代至今,视线追踪算法已得到了广泛研究和应用,它们主要分为基于硬件的算法和非侵入式算法两大类.其中,基于硬件的视线追踪算法可分为眼电图法:Woodworth^[1]等人记录眼球的左右运动方向,通过放大电位差变化获取眼球位置信息;电磁感应法:Robinson于1963年首创电磁感应法^[2],用于监视双眼的运动情况;卡尔曼滤波法:胡鹏^[3]将系统的先验信息和实时传感器测量数据进行动态融合,实现对视线位置的精准估计;瞳孔角膜反射法:陈秋香^[4]用红外光源照射眼球表面,在角膜和瞳孔上形成反射,并利用相机捕捉反射图像分析其位置和移动情况,以此推断眼睛的视线方向,实现了视线追踪等四类.考虑到该类基于特定设备的算法成本较高,在测量技术上存在较大的复杂性,以及在目标物体受到遮挡或快速移动等复杂场景中的追踪效果会有所下降.因此,近年来出现了不依赖于设备的非侵入式视线追踪算法.

非侵入式视线追踪算法通过综合利用摄像头和图像处理技术,无须通过物理接触便能实现视线追踪.根据实现方式可分为基于卷积神经网络的视线追踪、基于人体测量学特征的方法以及多项式标定算法等三类.李钰卿^[5]使用轻量级的线性变换特征图替换

原来的卷积特征图,将ResNet-18作为子模型构建多区域输入视线追踪模型.该方法在不同的应用场景中都有很好的适应性,但模型较为复杂.王文成和常亮发提出以区域投影为基础的人眼定位技术^[6],该技术将人脸图像在垂直和水平方向上进行分离,提取投影过程中的二维特征,通过跟踪边界并分析灰度特性来精准定位人眼瞳孔,从而实现视线追踪.张新^[7]提出了一种基于人工神经网络的动态标定算法,将人工神经网络应用到视线追踪系统的标定过程中,解决运行前需要进行标定以及误差矫正算法较复杂等问题,提高了系统运动追踪的鲁棒性.黄亚勤^[8]采用基于霍夫变换的圆检测方法实现眼部图像中虹膜的准确定位,但标定过程容易受到外部环境、标定位置等因素的影响.张远辉等人^[9]通过图像处理算法获取左右眼的(Pupil Center Corneal Reflection, PCCR)矢量信息,并用二阶多项式映射函数计算二维平面注视点,利用眼球光心标定方法获取三维坐标,结合平面注视点计算视线方向,得到初步注视点估计.但PCCR在光照条件不稳定时会影响矢量的准确度,进而影响注视点估计的准确性.以上视线追踪算法对头部姿势的敏感性较高,可能需要较复杂的校准和调整.而基于瞳孔检测的视线追踪算法能够学习到眼球结构和纹理等更复杂的特征,自动完成瞳孔位置的定位,提高对不同个体眼部特征的适应性和在多变环境下的鲁棒性.并通过精准捕捉瞳孔的细微变化来提升视线追踪的准确性,进而在此基础上可实现注意力检测.

目前教育等应用领域中出现了对注意力检测的研究.Lin等人^[10]使用眼动仪记录学生眼动情况,分析其眼睛注视路径.Zhang等人^[11]和Zhu等人^[12]通过集成到可穿戴设备上的相机、惯性测量单元和脉搏检测器来综合评估学生的注意力水平.以上方法都需要运用到额外的硬件设备.Asteriadis等人^[13]通过眼睛注视方向与头部旋转的相关性来量化注意力检测结果,将头部姿势和眼睛注视方向映射为注意力水平.吾部力喀斯木等人^[14]通过头环检测学习者头部姿态和运动,判断学习者的学习行为和注意力.张杰^[15]提出基于人脸识别算法的大学课堂抬头率检测方法,用于监测学生听课时的注意力水平.党俊俊^[16]基于面部图像与脑电信号的双模态注意力检测模型,判断学习者的注意力状态.这些方法均需要结合除人脸外的其他多类特

征,计算复杂度较高.孙玥等人^[17]通过判断驾驶员视线状态实现对驾驶员注意力的判断,光照条件不好时准确率较低.常建霞等人^[18]结合信息象征功能理论和情感诉求理论,利用眼动追踪方法考察游客的视觉注意和感知吸引力,但未能实现实时注意力检测.

综上所述,现有注意力检测算法普遍存在开销较大、计算复杂度高、精度偏低等问题,在实时性和准确性方面仍有待提升.同时,从视线追踪到注意力的有效转换问题也亟待解决.

2 本文工作

本文对基于瞳孔追踪的注意力检测方法进行研究,通过实时视线追踪算法 LCET 实现实时视线追踪,然后结合分割网络和决策网络的优势以及眼睛开合度和注视离散度实现注意力检测.首先从 CASIA-Iris-Twins 数据库中选取图片,通过边缘检测、图像分割和细线化等预处理操作,得到了瞳孔区域更加精确和连通的数据集.然后利用 Adaboost 算法和 OTSU 动态阈值获取瞳孔中心区域的位置特征,并结合改进的 LC 组合模型实现实时视线追踪算法 LCET.最后基于 SNDN 模型,结合眼睛开合度和注视离散度来判断注意力集中程度,从而实现视线追踪到注意力的转换.

2.1 数据集与数据预处理

本文从 CASIA-Iris-Twins 虹膜数据库^[19]提供的双眼瞳孔图片中随机选取了 3 183 张图片进行研究. CASIA-Iris 数据库总共包含来自 1 800 多个真实主体和 1 000 个虚拟主体的 54 601 张虹膜图像,分为 CASIA-Iris-Interval、CASIA-Iris-Lamp、CASIA-Iris-Twins、CASIA-Iris-Distance、CASIA-Iris-Thousand 和 CASIA-Iris-Syn 6 个子集,分辨率为 640 * 480 像素.

为提高后续瞳孔追踪的精准度,本文对选取的数据集进行了预处理.首先通过边缘检测、迭代阈值分割及细线化处理技术得到了连通性准确的瞳孔图片.然后采用 Sobel 算法进行瞳孔边缘检测^[20],获取到清晰的瞳孔边缘轮廓图像^[21].在此基础上,使用迭代阈值图像分割及细线化处理技术^[22],得到了更为清晰精确以及瞳孔区域轮廓连通性更高的瞳孔图像,为后续研究提供了可靠数据保障.

2.2 实时视线追踪算法 LCET

2.2.1 瞳孔中心区域位置特征提取

本文利用 Haar 特征和级联分类器技术精确识别并提取人脸中的 12 个关键眼部特征点^[23],然后利用 Harris 角点检测算法获取内眼角位置.在获取到内眼角位置后,使用高斯滤波对其进行降噪处理,以降低后续实时视线追踪的计算负担,确保系统在复杂环境下依然能够对视线进行实时追踪.目前主流的滤波算法有均值滤波、中值滤波、高斯滤波等.考虑到高斯滤波不仅可以保留图像中的细节信息,还能够抵制图像中的噪声.因此本文选择高斯滤波算法进行降噪处理^[24]:首先计算滤波权重,对图像进行边缘扩充,使得图像原有的所有像素点可以取到邻域大小为 $(2n + 1)^2$ 的矩形窗口,对其进行权重计算,减少图像噪声.公式(1)为二维高斯函数的标准形式, $W(x, y)$ 表示二维高斯函数的权重,用于计算图像每个像素点的滤波权重. σ 为标准差, σ 越大,权重分布越均匀,滤波效果越好, x 和 y 分别表示距离中心点的水平和垂直距离:

$$W(x, y) = \frac{1}{2\pi\sigma^2} * e^{-\left(\frac{x^2+y^2}{2\sigma^2}\right)}. \quad (1)$$

由于高斯滤波二维加权算法较为耗时,可能会影响计算的实时性,因此根据可分离性原则推导高斯滤波的权重公式,通过逐一检查每个像素点,并利用矩形窗口内的像素点的加权和进行滤波处理,将高斯滤波值的二维加权和转换为一维加权和进行计算,以减少计算的复杂度.如公式(2)(3)所示,对于图像中任意一点 $P(x, y)$, $Conx(x, y)$ 和 $Cony(x, y)$ 分别表示沿行、列方向的一维加权和, $W(x)$ 、 $W(y)$ 为一维权重:

$$Conx(x, y) = \sum_{i=0}^{2n} P(x-n+i, y) * W(x). \quad (2)$$

$$Cony(x, y) = \sum_{i=0}^{2n} Con(y-n+i, y) * W(y). \quad (3)$$

如图 1 所示,高斯滤波在抑制原始图像中的噪声干扰方面表现良好,在消除原始图像干扰噪声的同时,保留了原图中人眼、人脸的细节和结构等重要信息.

瞳孔中心区域位置特征提取需要依靠瞳孔检测方法,目前主流的方法有以下 3 种:

1) 基于粒子滤波和卡尔曼滤波的瞳孔检测方法^[25-27]:是一种高效地解决瞳孔运动中非线性和非高斯问题的技术.这种组合滤波方法不仅有效克服了传

瞳瞳孔位置追踪方法不稳定的问题,还显著提高了瞳孔状态参数的提取精度,使算法的鲁棒性显著增强.但在处理高维度问题时,计算资源消耗较多,降低了算法的实时性.

2) 由粗及精的瞳孔定位方法^[28-32]: 该类方法采用由粗及精的瞳孔定位思想.基于瞳孔区域的直方图,采用改进的最大类间方差法自适应地分割瞳孔区域,实现粗略定位.利用瞳孔灰度梯度特性,进行瞳孔边缘点的精确检测.随后,基于检测到的瞳孔边缘点,结合亚像素定位方法以及椭圆拟合技术,对瞳孔中心位置进行估计.但面临较大遮挡时,该算法可能出现瞳孔参数检测精度下降的问题.

3) 基于 OTSU 动态阈值的瞳孔中心定位方法^[33-35]: OTSU 动态阈值算法基于图像的灰度分布特性,通过阈值法将图像划分为背景和目标两个部分^[36].首先,计算图像的灰度直方图及每个灰度级的出现概率.随后,算法将遍历所有潜在阈值,依据每个阈值将图像分为两部分:灰度值低于阈值的区域被视为背景,而高于阈值的区域则被认定为目标.对于每一个考察的阈值,计算背景和背景集合的平均灰度、类内方差以及类间方差.选择使类间方差最大化的阈值为最优阈值,从而实现图像的有效分割.在算法的最后阶段,利用纵向腐蚀滤波技术提取虹膜区域,以确定瞳孔中心的位置.这一算法在虹膜定位中表现出了较高的精确度和稳健性,能够准确地定位瞳孔的中心.故本文采用 OTSU 动态阈值算法获取瞳孔区域位置信息:根据图像的灰度分布自动确定一个适合的阈值,利用该阈值对图像进行二值化分割,从而得到瞳孔区域.然后通过遍历的方式来确定最佳阈值 T^* , 保证分割后的前景和背景之间的差异最大化,在计算这两个类别像素灰度方差的过程中,当类间方差达到最大时,意味着前景和背景之间的差异性最明显,分割效果最好,此时的阈值为最佳阈值 T^* .具体实现步骤如下:首先获取图像直方图范围定位 $[0, 1]$, 其次计算类间方差: $\sigma^2(t) = \omega_0(t) * \omega_1(t) * [\mu_0(t) - \mu_1(t)]^2$, 其中 $\omega_0(t)$ 和 $\omega_1(t)$ 是 t 阈值下两个类别的像素权重, $\mu_0(t)$ 和 $\mu_1(t)$ 是两个类别的平均灰度值,最后获取类间方差 $\sigma^2(t)$ 的最大阈值 T^* , 其中 $T^* = \operatorname{argmax} \sigma^2(t)$.考虑到图像分割过程中可能会引入噪声,因此需进行去噪处理.由于虹膜通常呈狭长形,因此本文

对二值化图像进行横向腐蚀滤波,以消除不连续且长度较短的黑色噪点.然后对去噪后的图像分别进行行、列遍历,以锁定虹膜的最小包络矩形框,最终将矩形框的几何中心视为瞳孔中心位置.



(a) 原始图像

(a) Original image



(b) 高斯滤波

(b) Gaussian filtering

图 1 高斯滤波结果

Fig.1 Result of Gaussian filtering

2.2.2 基于改进 LC 组合模型的实时视线追踪

以瞳孔注视摄像头的视线方向为 Z 轴建立摄像头坐标系.在这一坐标系下,将摄像头视频流中捕捉到的瞳孔中心区域位置特征作为 LC 组合模型的输入,以推测瞳孔在垂直和水平方向上的夹角:俯仰角 pitch 和水平移动角 yaw .然后建立一个真实世界坐标系和图像坐标系之间的映射关系,用以实现真实世界中注视点与图像中注视点之间的对应关系,从而实现视线到屏幕上的映射,如图 2 所示. G 为双眼瞳孔中心 P 在屏幕上的注视点, Q 为 G 的水平延长线与屏幕侧边的交点.首先我们需要测量摄像头与 P 之间的距离 D_p 、 Q 与 P 之间的距离 D_u 、水平视场角 α 及垂直视场角 β 等一系列数据^[37].其中,水平视场角 α 和垂直视场角 β 分别决定了注视者的左右以及上下视野的边界,这两个参数共同构成注视者的注视空间区域.基于这些数据进行夹角计算完成视线映射,从而估计出 pitch 、 yaw .

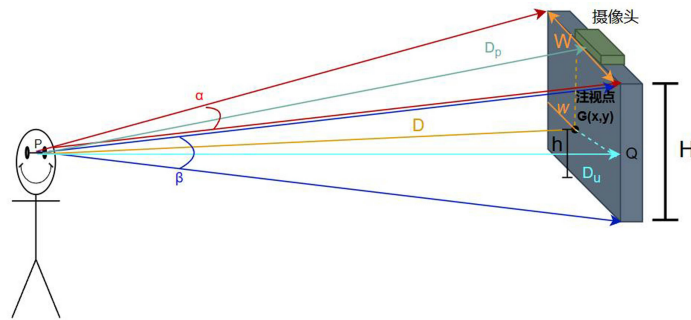


图 2 摄像头、眼睛和屏幕之间的位置示意图

Fig.2 Illustration of the camera, eyes and screen's positions

摄像头与 P 之间的距离 D_p 以及 Q 与 P 之间的距离 D_u 的计算方法如公式(4)所示,其中 D 为摄像头到注视者的水平距离, H 为摄像头距电脑屏幕底边的高度, h 为注视点 G 距屏幕底边的垂直距离, W 为屏幕的宽度, w 为注视点 G 距屏幕侧边的水平距离. 垂直视场角 β 的计算方法如公式(5)所示. 水平视场角 α 的计算如公式(6)所示. 通过上述方法得到水平视场角 α 和垂直视场角 β 后, 接下来计算夹角 $pitch$ 、 yaw . 首先计算注视点 G 相对于屏幕宽度和高度的位置比例 S_x 和 S_y , 如公式(7)所示, 其中 x 、 y 分别为注视点 G 在屏幕上的横纵坐标, 取值范围为 $[0, 1]$.

$$D_p = \sqrt{D^2 + (H-h)^2}, D_u = \sqrt{D^2 + (W-w)^2}. \quad (4)$$

$$\beta = 2 \arctan \frac{\frac{H}{2}}{D_u}. \quad (5)$$

$$\alpha = 2 \arctan \frac{\frac{w}{2}}{D_p}. \quad (6)$$

$$S_x = \frac{x}{W}, S_y = \frac{y}{H}. \quad (7)$$

为了便于后续夹角的计算, 将原点从屏幕左下角

移至屏幕中心处, 则 S_x 和 S_y 的取值范围从原始的 $[0, 1]$ 调整至 $[-0.5, 0.5]$. 设调整后 G 相对于屏幕宽度和高度的位置比例为 S'_x 和 S'_y , 利用公式(8)和(9)计算出 $pitch$ 和 yaw :

$$pitch = 2 S'_y * \frac{\beta}{2}. \quad (8)$$

$$yaw = 2 S'_x * \frac{\alpha}{2}. \quad (9)$$

2.2.3 基于角度预测的实时视线追踪

基于 LC 组合网络架构的视线角度估计模型, 如图 3 所示, 将摄像头视频流中捕捉到的瞳孔中心区域位置作为输入, 送入主干网络 ResNet-50 获取 $pitch$ 和 yaw 角度, 并通过计算概率分布的期望获取注视点. 本文采用两个独立的全连接层, 分别预测每个角度以提高角度的预测精度, 让两个全连接层共享相同的卷积层 backbone, 以确保网络在特征提取方面的一致性. 此外, 对每个注视角度使用两个单独的损失函数, 包括注视二值分类和回归. 这两种损失通过网络进行反向传播, 实现对网络权重的精确微调, 提高网络的泛化性能. 最后, 每个全连接层都输出.

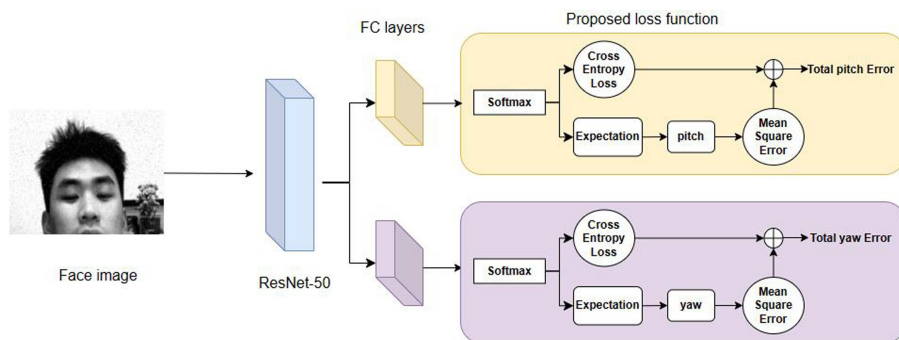


图 3 LC 组合模型网络架构

Fig.3 Network architecture of LC combination model

首先使用 softmax 层将网络输出 logits 转换为概率分布.如公式(10)所示:

$$P(x_i) = \frac{e^{x_i}}{\sum_{j=1}^n e^{x_j}} \quad (10)$$

然后,应用交叉熵损失函数计算输出概率与目标分类标签之间的分类损失,交叉熵损失函数定义计算如公式(11)所示, x_i 表示独热编码向量, i 表示类别的索引:

$$H = - \sum_i x_i \log P(x_i) \quad (11)$$

随后利用计算概率分布的期望 E ,实现对注视点预测的细粒度分析,如公式(12)所示.其中 v_i 代表与第 i 个目标分类标签相对应的注视点的具体数值:

$$E = \sum_i v_i P(x_i) \quad (12)$$

为了更全面地评估预测的准确性,最后计算该预测的均方误差,并将其添加到分类损失中.将均方误差引入到输出中,有助于改进注视预测的准确性,提高模型的性能和适应性.预测值和真实值之间的均方误差 MSE 计算如公式(13)所示,其中 N 为样本数量, y_i 为第 i 个样本的真实值, p_i 为第 i 个样本的预测值:

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - p_i)^2 \quad (13)$$

将均方差与损失进行线性组合得到综合损失 CLS,如公式(14)所示,其中 β 为回归系数,用于调节 MSE 在总损失中的权重:

$$CLS = H + \beta * MES \quad (14)$$

将均方误差和损失函数进行线性组合,形成综合的优化目标,调整模型中非线性预测层的参数,以细

化预测过程.

2.3 基于分割网络和决策网络模型的注意力检测机制

2.3.1 基于 SNDN 模型的眼睛开合度计算

在注意力检测任务中,眼睛开合程度的获取是一项至关重要的任务,对检测的准确性和实时性有极大影响.通常基于图 4 所示的眼部关键点,通过公式(15)来计算眼睛开合度 EAR.

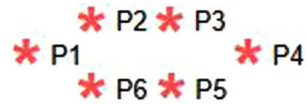


图 4 6 个眼部关键点

Fig.4 6 Eye key points

$$EAR = \frac{|y_{P2} - y_{P6}| + |y_{P3} - y_{P5}|}{2 |x_{P1} - x_{P4}|} \quad (15)$$

其中, y_{P2} 、 y_{P3} 、 y_{P5} 、 y_{P6} 分别为第 2、3、5、6 个人眼特征点的纵坐标值, x_{P1} 、 x_{P4} 为第 1、4 个人眼特征点的横坐标值.在眼睛处于睁开状态时,EAR 倾向于收敛于一个固定值,而当眼睛闭上时,EAR 数值将变为 0.传统的 EAR 计算方法多依赖于静态图像的分析,在捕捉眨眼或快速移动等眼睛动态变化方面可能存在误差叠加.

考虑到分割网络具有像素级别的精度优势,而决策网络可以对各类输入数据进行处理,对感兴趣像素进行识别学习,降低其分辨率以执行全局最大化,实现更快速、更准确的眼睛开合度计算.因此,本文对这两个网络模型进行端到端的组合,形成 SNDN 模型,如图 5 所示,旨在克服传统 EAR 计算方法中的不足,提升眼睛开合度计算的准确性和稳定性.

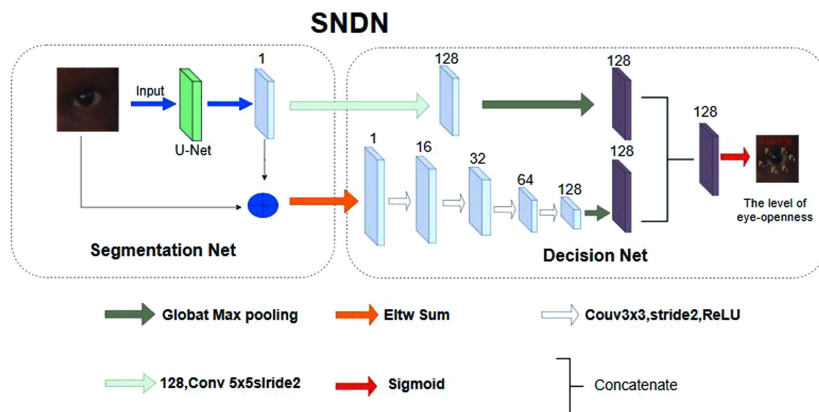


图 5 SNDN 模型结构图

Fig.5 Structure diagram of SNDN model

在分割网络中,将瞳孔中心区域像素进行二分类,将感兴趣像素输出到决策网络.如图 6 所示,分割网络的设计采用了 U-Net 结构.U-Net 是计算机视觉领域应用较广泛的经典网络,由医学图像分割领域的研究者 Ronneberger 等人于 2015 年提出^[38],主要用于生物医学图像中组织和器官的分割.U-Net 网络结构包含对称的上采样和下采样模块,两者通过一个跳跃连接层相连,整体分为收缩路径和扩张路径两大部分,呈现出明显的对称性.在收缩路径中获取上下文信息,在扩张路径中实现精确定位.本文所采用的分割网络,通过左侧收缩路径和右侧扩张路径的协同工作,实现对图像的高效处理.左侧路径通过连续的 3×3 卷积层和最大池化层,逐步降低图像分辨率,提高特

征的抽象性,以便更好地捕捉图像的细节信息.而右侧路径则通过上采样操作,将这些高度抽象的特征映射回原始图像尺寸,同时保持空间信息的连续性,实现精确的像素级分类和区域分割.此结构设计将特征与空间细节相结合,确保对目标区域的准确界定,从而提升图像分割的精度和效果.因此,本文采用的分割网络将简化特征图和上采样输出进行连接,能够具备处理不同尺寸与分辨率眼部特征图片的能力,提高分割的准确性,实现对眼部区域各类信息的细致分割,同时兼顾图像局部与全局信息的处理.并且由于不存在全连接层,相较于同等深度的网络,网络所需的参数数量显著减少,泛化能力和速度得以提升.

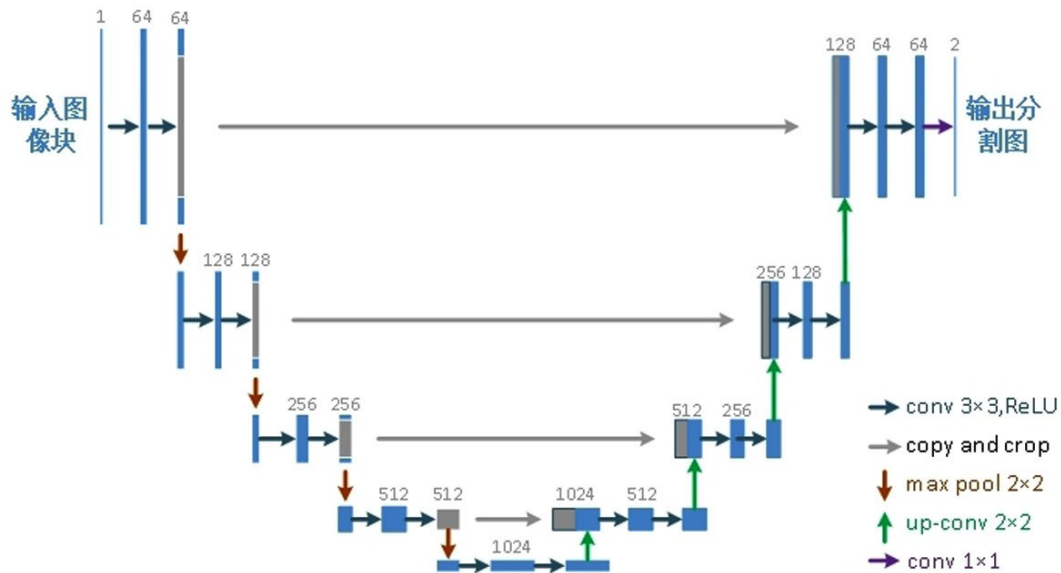


图 6 分割网络结构图

Fig.6 Structure diagram of segmentation network

在决策网络中,本文使用多层卷积对输入的感兴趣像素进行处理.卷积神经网络通过参数共享和局部感知的卷积操作能够有效提取图像各个层次上不同抽象级别的特征表示.具体而言,通过多个卷积层和池化层的堆叠,可以提取出高级语义特征,从而完成复杂的图像处理任务.与传统的全连接神经网络相比,卷积神经网络能够更精准地提取特征,在减少参数数量的同时保持算法性能.

如图 5 右半部分所示,决策网络中包含四次重复的卷积层和 ReLU 激活函数层,每组包括一个内核大小为 3×3 的卷积层和一个 ReLU 层.考虑到特征分辨率随着层数增加而降低的特性,为确保每层的计算需

求相对一致,采取逐层递增的设计思路,即随着层数增加,每层的通道数量由 16 逐层增加至 128.分割输出映射阶段涉及卷积和全局最大池化操作,用以综合分割网络的输出.在进行图像处理或分析时,关注当前层的输出,并且综合利用各层网络产生的特征信息,通过对分割结果进行卷积和全局最大池化处理,能有效提取更高层次的语义特征,有助于对图像中目标区域进行更准确的划分.此外,根据网络输出的质量以及场景的复杂程度,调整分割网络输出与逐像素分类结果之间的权重,随后通过损失函数为分割任务和分类任务分配不同的权重,进一步强化网络性能,使其能够更好地适应复杂的视觉场景,提高分割精

度.因此,本文所采用的决策网络对分割网络输出进行处理时,不仅将输出的眼部区域图像进行映射,而且融合由分割网络的输入与输出联合生成的一种特征映射,以减少计算复杂度,消除不必要的特征映射使用,进而提高网络的处理效率.

综上所述,对眼睛开合度的准确估计的具体步骤为:首先在分割网络中用多尺度分析技术,对瞳孔中心区域的像素进行二分类处理,筛选出关键像素将其传递至决策网络.然后在决策网络中,运用多层卷积对输入的感兴趣像素进行处理,以降低其分辨率,并执行全局最大化操作.最后,结合分割网络的输出结果进行逐像素分类,并通过权重加权的方式,实现对眼睛开合度的准确估计.图 7 所示眼睛开合度的计算结果为 0.39,表示眼睛为张开状态.

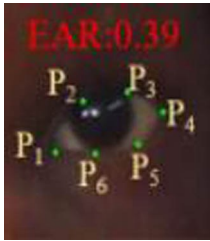


图 7 眼睛开合度计算结果

Fig.7 Result of eye opening degree calculation

本文采用的 SNDN 模型运用多尺度分析技术提取更为精确的眼睛特征,解决了传统 EAR 计算方法对眨眼或快速移动情况的处理局限性.通过在决策网络中利用眼部区域图像的映射输出,识别图像中的除眼睛外的区域和背景,并将这些特征映射与眼部区域的映射输出结合,减少多余特征映射的使用,得到更准确的眼睛区域,从而提升眼睛开合度计算的准确性和稳定性.

2.3.2 基于注视离散度的注意力检测

通过 SNDN 网络模型实现对眼睛开合度的精确估计.在眼睛开合度不为 0 的情况下,进行基于注视离散度的注意力检测.在现有注意力检测算法中,信息熵考虑了数据的分布信息,但是计算较为复杂,变分自编码器(Auto-Encoding Variationa, VAE)^[39]通过计算空间中分布的信息熵或 KL 散度,利用数据的特征信息进行注意力检测,以提高检测效果.但需要大量的数据和训练时间,并且还会受到模型结构和超参数选择的影响.综合考虑注意力检测的实时、适用等

因素,本节采用更为直观、适用于各种维度且计算简单的欧氏距离作为基本特征,通过离散度计算实现注视点到注意力的转换.

首先,按照知识点对教学视频进行切片,对每个视频切片中的理论注视区域 D_A 进行标注.在观看视频切片的过程中,每间隔 2 秒采样一次,设每个视频切片采样次数为 N ,将每次采样时的注视离散度 S_i 定义为用户实际注视位置 D 与理论注视区域 D_A 之间的欧式距离,如公式(16)所示.

$$S_i = \sqrt{(D_x - D_{Ax})^2 + (D_y - D_{Ay})^2}. \quad (16)$$

然后计算每个视频切片的注视离散度 S_A ,如公式(17)所示.将 S_A 与阈值 T_1 进行比较,如果 S_A 大于预设阈值 T_1 ,则判断该段视频切片的注意力为不集中,反之则判断为集中.

$$S_A = \frac{1}{N} \sum_{i=1}^N S_i. \quad (17)$$

最后,为了对整段视频的整体注意力情况做评估,利用公式(18)计算整段视频的全程离散度 S_U ,其中 $t_j(j=1 \cdots M)$ 为每段视频切片的时长, t 为整段视频的总时长, M 为整段视频的切片数, S_{A_j} 为第 j 段视频切片的注视离散度:

$$S_U = \sum_{j=1}^M \frac{t_j}{t} S_{A_j}. \quad (18)$$

将全程离散度 S_U 与预设阈值 T_2 进行比较,如果 S_U 大于阈值 T_2 ,则判断该段视频的总体注意力为不集中,反之则判断为集中.

3 实验结果分析

整个实验过程用 Python 实现.本次实验的实验环境如下:系统 Windows 10(64 位),处理器 Intel(R) UHD Graphics 630NVIDIA GeForce GTX 1050,摄像头为电脑自带摄像头.

3.1 提出的 LCET 算法和角度精度较高的算法对比实验结果分析

本文从 CASIA-Iris-Twins 虹膜数据库提供的双眼瞳孔图片中随机选取 3 183 张图片作为实验数据集,经过 2.1 节所述的预处理后,将其中 2 000 张作为 LC 组合模型的训练样本,剩余 1 000 张用于本实验.

选取凝视角误差作为评估指标,凝视角误差是指预测或测量的凝视角度与实际凝视角度之间的差异.

设凝视方向为 A , 预测的凝视向量为 B , 则预测注视方向与实际注视方向之间的角度误差 X 如公式(19)所示, 其中 $\|A\|$ 和 $\|B\|$ 分别表示两个向量的模。

$$X = \frac{A * B}{\|A\| * \|B\|} \quad (19)$$

实验结果如表 1 所示: LCET 算法在前 180° 和前 20° 范围内的实时视线追踪的平均角度误差分别为 10.41° 和 9.02°, 均小于角度精度较高的 Dilated-Net 等 3 种算法。

表 1 角度误差比较

Table 1 Comparison of angle errors

| 方法 | 前 180°/° | 前 20°/° |
|-----------------------------|--------------|-------------|
| Gaze360 模型 ^[40] | 11.4 | 11.1 |
| Dilated-Net ^[41] | 13.73 | 无法检测 |
| RT-Genie ^[42] | 12.26 | 无法检测 |
| 本文方法 | 10.41 | 9.02 |

在性能方面, 本算法与处理速度较快的粒子滤波视线追踪等 3 种算法进行对比。实验结果如表 2 所示, 本算法的平均处理速度为 11.8 毫秒/帧, 平均迭代次数为 1.25, 均明显超过目前较为优越的三种对比算法。

表 2 处理速度比较

Table 2 Comparison of processing speed

| 方法 | 平均耗时/帧 (ms) | 平均迭代次数(次) |
|--------------------------------------|-------------|-------------|
| 粒子滤波视线追踪 ^[25] | 18.7 | 1.80 |
| 基于帧差法的 Mean-Shift 算法 ^[43] | 16.5 | 1.60 |
| 粒子滤波过滤追踪 ^[44] | 16.95 | 1.65 |
| 本文方法 | 11.8 | 1.25 |

3.2 眼睛开合度的预测准确性验证实验

CASIA-Iris-Twins 虹膜数据库主要收集了双胞胎及多胞胎的虹膜图像, 但没有足够的眼睛开闭状态下的样本。因此本文采用 NTHU-DDD 数据集^[45] 进行眼睛开合度的实验。该数据集包括多种眼睛状态数据, 可用于眼睛开合度的分析。我们从中选择 400 张图片作为训练集, 260 张作为测试集, 随后通过对训练数据进行数据增强(旋转、缩放、裁剪、翻转、调整亮度对比度等变换), 生成增强的训练样本, 这样既尽可能保持眼睛开合度关键特征的同时增加数据的多样性, 又有效防止过拟合风险, 提高了模型的泛化能力。随后, 将 SNDN 模型和准确率较高的 FDRNet、GP-BCNN、DDD 三个网络模型进行对比。实验结果如表 3 所示, 本文提出的 SNDN 模型对眼睛开合度的预测准确率

达 96.72%, 优于对比模型。

表 3 眼睛开合度的预测结果

Table 3 Prediction result of eye opening and closing degree

| 方法 | Precision(%) | Recall(%) | F1(%) |
|-------------------------|--------------|--------------|--------------|
| FDRNet ^[46] | 93.10 | 85.71 | 89.25 |
| GP-BCNN ^[47] | 90.62 | 92.06 | 91.34 |
| DDD ^[48] | 83.60 | 80.95 | 82.26 |
| SNDN | 96.72 | 93.63 | 95.16 |

4 结论

本文对基于瞳孔追踪的注意力检测机制进行研究。首先对瞳孔数据集进行预处理, 然后运用 Adaboost 算法与 OTSU 动态阈值方法提取瞳孔中心区域位置特征, 并结合分类与回归损失 LC 组合模型, 以实现高效的视线追踪功能。进而基于 SNDN 模型, 利用眼睛开合度和注视离散度来综合判断注意力集中程度。未来我们将探索更为精确的注视点标定和映射算法, 探讨基于视频内容分析的理论注视区域自动标注方法, 并结合用户表情、肢体动作、语音等多模态信息, 实现更为客观全面的注意力检测。

参考文献

- [1] WOODWORTH R, SCHLOSBERG H, KLING J W, et al. Woodworth & Schlosberg's Experimental psychology[J]. Holt, Rinehart and Winston, 1972: 115-119.
- [2] ROBINSON D A. A method of measuring eye movement using a scleral search coil in a magnetic field[J]. IEEE Transactions on Bio-Medical Engineering, 1963, 10: 137-145.
- [3] 胡鹏. Kalman 滤波在视频目标跟踪中的应用研究[D]. 重庆: 重庆大学, 2010.
- [4] 陈秋香. 基于瞳孔-角膜反射的视线跟踪算法研究[D]. 重庆: 重庆大学, 2014.
- [5] 李钰卿. 基于深度学习的视线追踪技术研究[D]. 广州: 广东工业大学, 2021.
- [6] 王文成, 常发亮. 一种基于区域投影的人眼精确定位方法[J]. 光电子激光, 2011, 22(4): 618-622.
- [7] 张新. 基于单个网络摄像头的视线追踪系统研究[D]. 北京: 北京理工大学, 2018.
- [8] 黄亚勤. 基于视线跟踪技术的眼控鼠标研究与实现[D]. 成都: 西华大学, 2011.
- [9] 张远辉, 段承杰, 朱俊江, 等. 眼球光心标定与距离修正的 3 维注视点估计[J]. 中国图象图形学报, 2019, 24(8): 1369-1380.
- [10] LIN Y T, WU C C, HOU T Y, et al. Tracking students' cognitive processes during program debugging—An eye-movement approach[J]. IEEE Transactions on Education, 2016, 59(3): 175-186.
- [11] ZHANG X, WU C W, FOURNIER-VIGER P, et al. Analyzing students'

- attention in class using wearable devices [C]//2017 IEEE 18th International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM).China;IEEE,2017;1-9.
- [12] ZHU Z W, OBER S, JAFARI R. Modeling and detecting student attention and interest level using wearable computers [C]//2017 IEEE 14th International Conference on Wearable and Implantable Body Sensor Networks (BSN). Eindhoven, Netherlands; IEEE, 2017; 13-18.
- [13] ASTERIADIS S, KARPOUZIS K, KOLLIAS S. The importance of eye gaze and head pose to estimating levels of attention [C]//2011 Third International Conference on Games and Virtual Worlds for Serious Applications. Athens, Greece; IEEE, 2011; 186-191.
- [14] 吾部力喀斯木·吾布力艾散, 穆巴热科·斯迪克. 基于头部姿态与运动的注意力检测与反馈系统的设计与实现 [J]. 中国信息界, 2024(3):246-248.
- [15] 张杰. 面向大学课堂的抬头率检测系统的研究与实现 [D]. 武汉: 华中科技大学, 2018.
- [16] 党俊俊. 双模态注意力检测与反馈分析研究 [D]. 西安: 西安电子科技大学, 2023.
- [17] 孙玥, 杨国为, 陈雪鑫. 基于计算机视觉的疲劳及注意力检测算法 [J]. 计算机与数字工程, 2021, 49(6):1195-1198+1239.
- [18] 常建霞, 李君轶, 张妍妍, 等. 语言景观对游客视觉注意的影响——基于眼动追踪实验的分析 [J]. 旅游学刊, 2023, 38(10):119-133.
- [19] GAUTAM G, RAJ A, MUKHOPADHYAY S. Identifying twins based on ocular region features using deep representations [J]. Applied Intelligence; The International Journal of Artificial Intelligence, Neural Networks, and Complex Problem-Solving Technologies, 2021, 51(1):1-18.
- [20] 李迪, 吴奇, 杨浩森. 基于改进 Sobel 算子的边缘检测系统的设计与实现 [J]. 信息技术与网络安全, 2022, 41(3):13-17.
- [21] 王文辉. 基于深度学习的视网膜血管图像分割技术研究 [D]. 太原: 太原理工大学, 2022.
- [22] 王元元. 数字图像细线化处理方案探讨 [J]. 数字通信世界, 2019(6):105.
- [23] 周天相. 基于 Adaboost 的人脸检测研究 [D]. 北京: 北京邮电大学, 2013.
- [24] 蒋延杰, 谢蓉蓉, 李云红, 等. 基于图像边缘轮廓的角点检测算法研究 [J]. 电子设计工程, 2022, 30(21):185-188.
- [25] 王长元, 金瑞铭, 田宝杰, 等. 一种基于粒子滤波的瞳孔跟踪方法 [J]. 计算机与数字工程, 2015, 43(11):2044-2046+2077.
- [26] 董文会, 吴晓娟, 曲培树. 基于规则与卡尔曼滤波的人眼跟踪 [J]. 计算机工程与科学, 2006, 28(11):27-29.
- [27] 王长元, 张文强, 薛鹏翔. 粒子滤波和卡尔曼滤波组合的瞳孔跟踪方法 [J]. 计算机与数字工程, 2018, 46(4):739-742.
- [28] 周永修. 人眼瞳孔精确跟踪算法研究 [D]. 重庆: 重庆理工大学, 2016.
- [29] 李擎, 胡京尧, 迟健男, 等. 视线追踪中一种新的由粗及精的瞳孔定位方法 [J]. 工程科学学报, 2019, 41(11):1484-1492.
- [30] 杨秀, 张轩雄. 一种快速有效的虹膜定位方法 [J]. 软件导刊, 2019, 18(1):61-64.
- [31] 杨敏, 徐万明. 一种改进的瞳孔定位方法 [J]. 智能计算机与应用, 2016, 6(3):122-124.
- [32] 陈静瑜, 林丽媛, 刘冠军, 等. 一种基于几何形状特征的实时瞳孔定位追踪技术 [J]. 天津科技大学学报, 2021, 36(3):65-71.
- [33] 徐红霞, 成艳真. 基于动态阈值的运动目标检测算法 [J]. 科技信息, 2011(15):601+611.
- [34] 方超, 王斌斌, 陈立生. 基于双眼独立动态阈值的人脸预处理及识别 [J]. 电脑知识与技术, 2012(3):1618-1621.
- [35] 杨泽霖, 杨坚, 涂前彦, 等. 一种基于动态阈值的人脸识别方法、系统及存储介质: CN111814570B [P]. 2024-04-30.
- [36] 金丽丽. 探究基于动态阈值图像分割法的人脸识别技术研究 [J]. 计算机产品与流通, 2019(12):89.
- [37] 赵鸣雷. 基于机器视觉的仪表示数检测系统研究 [D]. 天津: 天津工业大学, 2022.
- [38] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional networks for biomedical image segmentation [C]// Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015. Cham; Springer International Publishing, 2015:234-241.
- [39] KINGMA D P, WELLMING M. Auto-encoding variational Bayes [J]. ArXiv e-Prints, 2013; arXiv:1312.6114.
- [40] KELLNHOFFER P, RECASENS A, STENT S, et al. Gaze360: Physically unconstrained gaze estimation in the wild [C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, Korea; IEEE, 2019;6911-6920.
- [41] CHEN Z K, SHI B E. Appearance-based gaze estimation using dilated-convolutions [C]// Computer Vision - ACCV 2018. Cham; Springer International Publishing, 2019;309-324.
- [42] FISCHER T, CHANG H J, DEMIRIS Y. RT-GENE: Real-time eye gaze estimation in natural environments [C]// Computer Vision - ECCV 2018. Cham; Springer International Publishing, 2018;339-357.
- [43] 岳昊恩, 袁亮, 吕凯. 结合帧差法与 Mean Shift 的抗遮挡跟踪算法 [J]. 现代电子技术, 2019, 42(12):180-182+186.
- [44] JACOB P E, MURRAY L M, RUBENTHALER S. Path storage in the particle filter [J]. Statistics and Computing, 2015, 25(2):487-496.
- [45] WENG C H, LAI Y H, LAI S H. Driver Drowsiness Detection via a Hierarchical Temporal Deep Belief Network [C]// Asian Conference on Computer Vision. Springer, Cham, 2016;117-133.
- [46] XUE C H, TIAN Z C, ZHAN F N, et al. Fourier document restoration for robust document dewarping and recognition [C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, LA, USA; IEEE, 2022;4563-4572.
- [47] WANG Y, HUANG R, GUO L. Eye gaze pattern analysis for fatigue detection based on GP-BCNN with ESM [J]. Pattern Recognition Letters, 2019, 123:61-74.
- [48] REDDY B, KIM Y H, YUN S, et al. Real-time driver drowsiness detection for embedded system using model compression of deep neural networks [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Honolulu, HI, USA; IEEE, 2017;438-445.

(责任编辑:张阳,殷锋,付强,和力新,肖丽;英文编辑:周序林,郑玉才)