

DOI: 10.13382/j.jemi.B2407941

# 融合面部外观与生理表征的作业人员疲劳判别

颜文琴<sup>1,2</sup> 郑秀娟<sup>1,2</sup> 殷中平<sup>3</sup> 张学刚<sup>4</sup> 贾明<sup>3</sup> 刘伯相<sup>1,2</sup> 涂海燕<sup>1,2</sup>

(1. 四川大学电气工程学院 成都 610065; 2. 四川大学信息与自动化技术四川省高校重点实验室 成都 610065;  
3. 深圳中广核工程设计有限公司 深圳 518172; 4. 中广核工程有限公司核电安全技术与装备全国重点实验室 深圳 518172)

**摘要:**在工业生产中,长时间和高强度的作业易导致人员疲劳,从而增加安全事故的风险。已有研究表明,接触式生理特征能有效表征疲劳状态,但在工业环境中采用接触式设备获取生理信号进行疲劳判别会干扰正常作业。因此,基于监控视频的疲劳判别成为更实际的选择,然而现有方法主要关注嘴部和眼部特征,未能全面反映疲劳状态。为此,提出了基于视频的融合面部外观与生理表征的无干扰式疲劳判别方法,通过双支路网络模型实现对作业人员疲劳判别。首先,在视频中定位面部感兴趣区域并进行子区域划分,通过提取皮肤反射光变化获取视频隐含的生理表征信息,进而构建生理时空图。接着,搭建双支路三维卷积网络分别提取面部外观和生理表征特征。最后,将两者特征融合并输入全连接层,以映射最终的疲劳判别结果。通过模拟工业生产任务获取的疲劳数据集验证了所提方法的性能。实验结果表明,基于视频的融合面部外观与生理表征的疲劳判别准确率达到88%,相较于现有技术具有更高的准确性和更强的现场适用性。

**关键词:** 疲劳检测; 视频; 面部外观; 生理时空图; 无干扰

**中图分类号:** TP39; TN91 **文献标识码:** A **国家标准学科分类代码:** 510

## Fatigue identification of workers by integrating facial appearance and physiological characteristics

Yan Wenqin<sup>1,2</sup> Zheng Xiujuan<sup>1,2</sup> Yin Zhongping<sup>3</sup>  
Zhang Xuegang<sup>4</sup> Jia Ming<sup>3</sup> Liu Boxiang<sup>1,2</sup> Tu Haiyan<sup>1,2</sup>

(1. College of Electrical Engineering, Sichuan University, Chengdu 610065, China; 2. Key Laboratory of Information and Automation Technology in Sichuan Province, Chengdu 610065, China; 3. China Nuclear Power Design Co., Ltd (Shenzhen), Shenzhen 518172, China; 4. State Key Laboratory of Nuclear Power Safety Technology and Equipment, China Nuclear Power Engineering Co., Ltd., Shenzhen 518172, China)

**Abstract:** In industrial production, prolonged and high-intensity operations can lead to worker fatigue, increasing the risk of safety incidents. Existing research has shown that contact-based physiological features can effectively represent fatigue status, but using contact-based instruments to monitor fatigue in industrial environments interferes with operations. Therefore, fatigue detection based on surveillance video has become a more practical choice. Current methods mainly focus on mouth and eye features, failing to comprehensively reflect fatigue status. To address this issue, we propose a non-intrusive fatigue detection method that integrates facial appearance and physiological representation, utilizing a video-based dual-branch network model for monitoring worker fatigue. First, we locate the facial areas of interest in the video and segment these areas. By extracting changes in skin reflectance due to variations in capillary blood volume, we construct a physiological spatiotemporal map. Next, we build a dual-branch 3D convolutional network to extract facial appearance and physiological feature representations separately. Finally, we fuse these features and input them into a fully connected layer to map the final fatigue detection results. The proposed method is validated using a fatigue dataset obtained from simulated industrial production tasks. Experimental results demonstrate that the fatigue detection accuracy, based on the integration of facial appearance and physiological features from video, reaches 88%, offering higher accuracy and stronger applicability in industrial settings compared to existing technologies.

**Keywords:**fatigue detection; video; facial appearance; physiological spatiotemporal map; distraction-free

0 引 言

工业生产中作业人员的身心状态与生产安全关系密切。持续长时间、高强度的重复性劳动容易引发作业人员认知疲劳状态<sup>[1]</sup>,直接影响到作业人员的反应速度和决策能力。作业人员疲劳作业可能导致操作失误,甚至引发严重的安全事故。因此,对作业人员进行疲劳判别对于保障工业生产和保护生命安全十分重要。

目前,针对疲劳判别问题已有大量研究方法被提出,主要集中于基于面部外观和基于接触式生理信息的方法<sup>[2-3]</sup>。基于面部外观方法主要通过面部运动等参数,利用计算机视觉算法判别疲劳状态<sup>[4-5]</sup>。Savas 等<sup>[6]</sup>利用 OpenCV 和 Dlib 库提取面部运动,包括眨眼频率、打哈欠次数、嘴巴闭合度和眨眼次数等,通过支持向量机(support vector mechine, SVM)分类器判别疲劳状态。Hou 等<sup>[7]</sup>定位面部特征点计算嘴部和眼部的特征信息,通过机器学习方法判别疲劳状态。这些方法主要集中于眼部和嘴部特征,忽略了生理信号特征和其他面部特征,无法全面反映个体疲劳。

基于接触式生理信号的方法主要通过如脑电图、心电图和皮肤电反应等不同生理参数的变化来检测疲劳状态<sup>[8-9]</sup>。人体的心率、呼吸率、心率变异性等生理参数会随人体的疲劳状态变化。在针对驾驶疲劳的生理特征研究中指出,随着疲劳增加,驾驶员均出现心率、呼吸率降低,心率变异性整体张力增大的特点<sup>[10]</sup>。Ouyang 等<sup>[11]</sup>结合由疲劳监测试验获得建筑工人的心率变异性数据(由心电信号计算得到)和皮肤电信号,利用 SVM 模型对疲劳状态进行判别。而基于接触式生理信息的疲劳检测方法由于需要专用的测量设备,有时还会干扰作业人员的正常工作,并不适用实际生产场景<sup>[12]</sup>。研究发现可以从面部视频的裸露皮肤区提取得到生理信息,包括心

率、呼吸率、心率变异性等生理参数<sup>[13]</sup>。Niu 等<sup>[14-15]</sup>通过视频信息建立面部时空图捕捉人体生理信息。因此,前期研究成果表明,通过视频同时获取面部外观和生理信息的多模态表征具有可行性。

随着深度学习在各个领域的应用,研究者们提出基于三维卷积神经网络模型从视频序列中提取时空特征,如 Resnet-3D<sup>[16]</sup>、C3D<sup>[17]</sup>、S3D<sup>[18]</sup>等方法,虽然这些方法在动作识别等领域获得了较好精度和性能,但并非针对工业生产场景进行疲劳判别设计,因此无法直接应用于疲劳判别。此外,它们的参数量较大,对计算资源要求高,对于边缘设备的实时应用具有局限性。

针对上述问题,本文提出了一种融合面部外观与生理表征的作业人员认知疲劳判别方法。该方法基于视频设计了一种双支路网络模型,将面部外观与生理时空图相结合,用于疲劳特征的学习。为充分考虑疲劳状态的时间特性,采用三维卷积网络捕捉时间维度上的特征信息。实验结果表明,面部外观与生理表征的融合能够实现不同模态之间的优势互补,疲劳检测的准确率达到 88%。同时在与其它方法的比较中,本方法兼顾了精度和时间复杂度。因此,本文方法在工业生产中具有良好的应用前景,可有效实现认知疲劳状态的检测与判别。

1 原理和方法

鉴于工业生产中常用的传感设备主要为监控设备,本文研究基于视频提取面部外观和生理表征两种模态特征。通过双支路网络提取疲劳特征,并融合这些特征进行疲劳判别。总体框架如图 1 所示。该模型包含 3 个功能模块:提取面部感兴趣区域获得面部外观信息;面部感兴趣区域子区域划分,通过像素平均构建生理表征时空图;使用双支路三维卷积网络进行多模态特征提取,最终融合特征后输入全连接层映射进行疲劳判别。

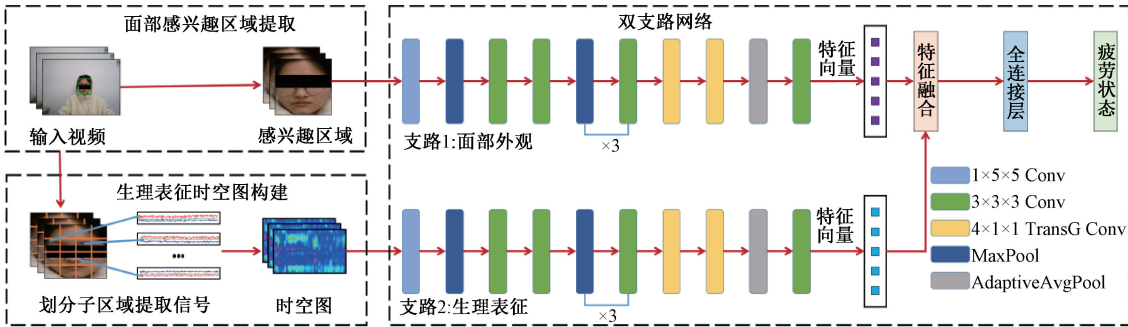


图 1 方法总框架图  
Fig. 1 Framework of method

### 1.1 面部感兴趣区域提取

人体的疲劳状态主要体现在眼睛、嘴巴等面部特征中,体现为眼睛睁开的幅度变小、闭眼、打哈欠、面部肌肉松弛等现象。因此本文使用 Mediapipe 多媒体框架<sup>[19]</sup>进行人脸感兴趣区域的定位提取,去除背景噪声。该算法通过神经网络模型从视频帧中获得人脸边界框,包括人脸起始点的横纵坐标以及边界框的宽度和高度。面部感兴趣区域提取过程如图 2 所示。

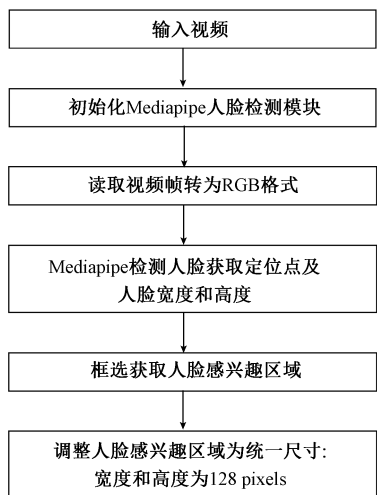


图 2 感兴趣区域提取流程

Fig. 2 Region of interest extraction flowchart

人体疲劳状态往往是逐渐变化的,不会出现突变现象,因此为了有效捕捉视频中人脸外观的动态变化,同时避免重复样本的引入,本研究每隔  $t$  秒提取单帧面部感兴趣区域用于表征外观信息。该提取方法不仅有效减少了相邻帧之间的高度相关性,也降低了模型在训练过程中的冗余信息,增强了特征学习的效率。

### 1.2 生理表征时空图构建

皮肤反射光强度受到血管中血液体积和血氧饱和度的影响。当心脏搏动时,血液流入血管,导致皮肤下容积量的变化,从而改变反射光的强度,在视频中体现为皮肤的微弱颜色变化,原理如图 3 所示。王宇等<sup>[13]</sup>的研究表明,可以从视频中定位人脸裸露皮肤区域进行生理测量,获取如心率、脉搏波等生理信号。人脸不同区域包含的生理信号有差异,例如脸颊及额头区域包含更强烈的生理信号。为了捕捉生理信号的空间信息和时间信息,通过对人脸外观感兴趣区域划分子区域获得多个生理信号组成时空图,提取流程如图 1 中生理表征时空构建模块。

考虑到人体头部一般呈椭圆形,在矩形框定的人脸区域的四角一般包含背景或头发等非皮肤区域,因此首先将选定的脸部区域分为  $M \times N$  块子区域,并舍弃四角子区域,最终获得  $n$  块感兴趣子区域。第  $i$  个感兴趣子区

域中提取的 RGB 三通道的远程生理信号记作:

$$f_i(t) = \frac{\sum_{x,y \in R_i} P(x,y,t)}{|R_i|} \quad (1)$$

式中:  $f_i(\cdot) \in \mathbf{R}^c$  表示第  $i$  个感兴趣区域提取得到的远程生理信号;  $c$  表示视频的 RGB 三颜色通道;  $P(x,y,t)$  表示属于第  $i$  个感兴趣区域  $t$  时刻的像素点值;  $|R_i|$  表示第  $i$  个感兴趣区域的总像素点数。通过对像素值取平均值可以去除相机产生的噪声影响。为了统一尺度,将每个信号进行归一化:

$$f_i'(t) = \frac{f_i(t) - \bar{f}_i(t)}{\text{std}(f_i(t))} \quad (2)$$

式中:  $\bar{f}_i(t)$  表示信号平均值;  $\text{std}(f_i(t))$  表示信号的标准差。进一步使用三阶巴特沃斯带通滤波器对信号进行滤波,频率范围为  $0.667 \sim 4$  Hz,并通过线性最小二乘法拟合的结果进行去趋势化处理。将  $n$  个子区域获取到的信号排列起来,可以为每个人脸视频序列形成一个时空表示图:

$$\mathbf{F} = \begin{bmatrix} f_1'(t_0) & f_1'(t_0+1) & \cdots & f_1'(t_0+T) \\ f_2'(t_0) & f_2'(t_0+1) & \cdots & f_2'(t_0+T) \\ \vdots & \vdots & \ddots & \vdots \\ f_n'(t_0) & f_n'(t_0+1) & \cdots & f_n'(t_0+T) \end{bmatrix} \quad (3)$$

式中:  $\mathbf{F} \in \mathbf{R}^{n,T,c}$  表示面部视频的时空图;  $t_0$  表示视频的开始帧;  $T$  表示提取视频总帧数。通过时空图,能够清晰地观察到不同脸部区域在时间序列上的信号强度变化,从而识别出人体相关生理特征信息<sup>[14-15]</sup>。同时一张时空图中包含  $T$  帧脸部视频的时间信息和空间信息,既有助于网络学习生理特征,也有利于减少网络的时间复杂度。

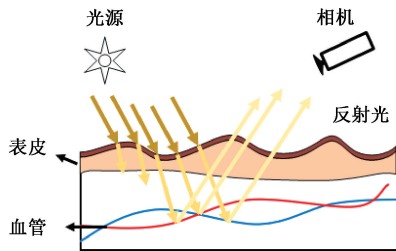


图 3 基于视频的生理测量原理示意图

Fig. 3 Schematic diagram of video-based physiological measurement

### 1.3 双支路网络结构

本文提出基于面部外观和生理表征的双分支网络结构,外观支路网络的输入为  $\mathbf{A}_{\text{input}} \in \mathbf{R}^{N_1 \times h \times w \times c}$ ,表示  $N_1$  帧人脸视频帧堆叠,其中  $h, w$  表示图像高度和宽度,  $c$  表示 RGB 三颜色通道。作为外观支路的补充信息,生理表征

网络支路的输入为  $F_{input} \in \mathbf{R}^{N_2, n \times (60/t), t \cdot fps, c}$ , 表示  $N_2$  帧时空图堆叠, 其中  $n$  表示面部感兴趣区域子区域数量,  $t$  表示视频帧采样间隔,  $fps$  表示视频每秒帧数,  $c$  表示 RGB 三颜色通道。通过时空图的堆叠, 模型能够获取更丰富的时序表达和特征表达, 不仅补充了外观支路网络的生理时空信息, 也让网络不仅仅局限于单帧数据, 从而增强疲劳判别的准确性。

本文采用三维卷积网络进行时空特征提取, 以扩展二维卷积神经网络在时间维度上的卷积能力, 示意图如图 4 所示。卷积输入输出特征图尺寸之间的关系为:

$$H' = \frac{H + 2p - f}{s} + 1 \quad (4)$$

$$W' = \frac{W + 2p - f}{s} + 1 \quad (5)$$

$$C' = \frac{C + 2p - f}{s} + 1 \quad (6)$$

式中:  $H$ 、 $W$ 、 $C$  表示输入特征图尺寸;  $H'$ 、 $W'$ 、 $C'$  表示输出尺寸图尺寸;  $p$  表示填充维度;  $f$  表示卷积核大小;  $s$  表示步长。三维卷积模型能够有效识别图像在时间序列中的变化, 从而增强对上下文信息的感知能力。通过共享卷积核的方式, 可以显著减少模型的参数数量, 降低计算复杂度, 并提高训练效率。特征处理过程中, 采用最大池化层和自适应平均池化层进行特征降维, 同时保留全局信息。转置卷积用于特征上采样, 优化对应的模态生成特征。

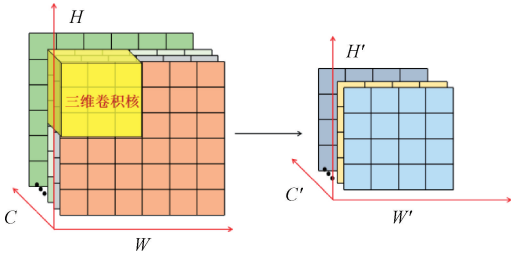


图 4 三维卷积示意图

Fig. 4 Three-dimensional convolution diagram

在该网络结构中, 外观支路和生理表征支路分别提取特征后进行融合, 最终通过全连接层输出融合特征。使用 sigmoid 激活函数实现疲劳判别。此外, 采用二元交叉熵损失函数作为模型的损失计算标准:

$$Loss = -\frac{1}{m} \sum_{i=1}^m [y_i \log(p_i) + (1 - y_i) \log(1 - p_i)] \quad (7)$$

式中:  $m$  表示样本数量;  $y_i$  表示第  $i$  个样本的真实标签;  $p_i$  表示预测第  $i$  个样本为正类的概率。

## 2 数据采集与评价指标

### 2.1 疲劳激发试验采集数据集

卡罗林斯卡嗜睡量表 (Karolinska sleepiness scale, KSS) 是一种用于评估个体主观睡意程度的量表<sup>[20]</sup>, 广泛应用于睡眠研究、疲劳监测以及驾驶安全等领域。其评分范围通常为 1~10, 每个分数对应一个疲劳状态描述, 如表 1 所示, 研究将 KSS 量表作为主观疲劳标签。

表 1 KSS 量值对应的疲劳状态描述

Table 1 Descriptions of fatigue states corresponding to KSS scores

KSS 量值	疲劳状态描述
1	极度警醒
2	非常警醒
3	警醒
4	比较警醒
5	不太警醒, 也无困意
6	有一点困意倾向
7	有困意, 但是不需要努力保持清醒
8	有困意, 需要一定的努力保持清醒
9	非常困倦, 需要极大努力保持清醒
10	极度困倦, 无法保持清醒

奇异球 (oddball) 实验范式常用于心理认知科学领域事件相关电位 (event-related potential, ERP) 实验范式之一, 常被用于研究注意力和认知过程<sup>[21]</sup>。经典的 Oddball 范式为在一项实验中, 随机呈现作用于同一感觉通道的两种刺激——标准刺激和偏离刺激。在该范式中, 通过被试对靶刺激的反应来观察大脑对异常事件的处理过程。当被试处于疲劳状态时, 身体反应将变得迟钝, 持续性注意力变差。这些状态变化将客观反映为对靶刺激的反应时间延长, 甚至直接忽略靶刺激, 并且准确率也有所下降。因此反应时间及准确率指标可以作为疲劳程度的客观测量。

疲劳激发试验基于视觉通道刺激的 oddball 范式, 可以有效诱发和测量被试者在长时间高度认知任务中产生的疲劳状态。试验数据采集示意图如图 5 所示。本文所设计的 oddball 范式界面包含 5 个方块以及对应的 5 个校正按钮。在正常界面中, 5 个方块均为绿色, 即标准刺激, 此时被试无需作出反应; 而在异常界面中, 5 个方块中有 1 个方块变为红色, 即为靶刺激, 此时要求被试者快速点击红色方块下的校正按钮。频繁转换注意力的任务会不断消耗被试者的认知, 逐步引发认知疲劳。本文研究设置每个界面持续时间为 2 s, 正常界面与异常界面的出现次数之比为 4:1。一次完整的 oddball 范式实验持续 10 min, 每次试验连续进行 3 次, 总时长为 30 min。这



种渐进式的任务负荷设计贴近实际作业环境中的认知需求,能够有效诱发和测量被试者产生的认知疲劳。

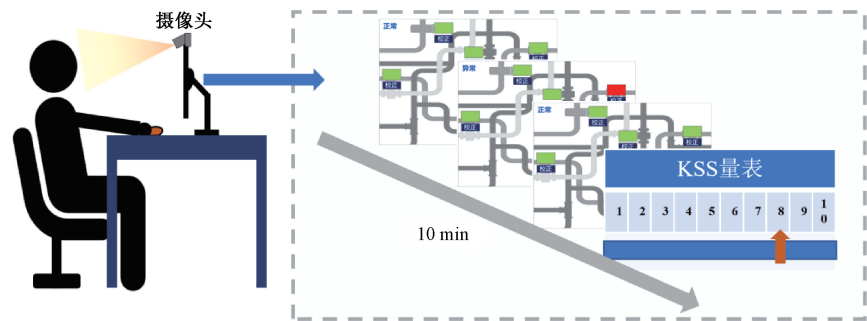


图 5 试验数据采集示意图  
Fig. 5 Diagram of experimental data collection

为了更贴近实际作业环境的易疲劳时间段,研究安排了上午(9:00~12:00)、中午(12:00~14:00)和傍晚(16:00~18:00)3个时间段进行试验。试验过程中将记录被试者每次异常反应的时间、误反应次数和漏反应次数,作为客观的疲劳标签,反映注意力下降、反应减慢等认知疲劳状态。在被试者进行模拟操作的同时,使用网络摄像头(海康威视 USB 摄像头)获取他们的面部视频数据。在每次 oddball 范式试验后被试者填写一次 KSS 量表,作为主观疲劳标签。

本文招募了 36 名志愿者被试者参与疲劳检测试验,其中包括 7 名女性和 29 名男性,年龄分布在 22~35 岁。在试验开始前,所有被试者均签署了书面知情同意书。研究将一位被试者在一次 oddball 试验范式中采集的数据作为一个样本,共采集 309 人次的视频疲劳数据,初始疲劳分布如图 6 所示。

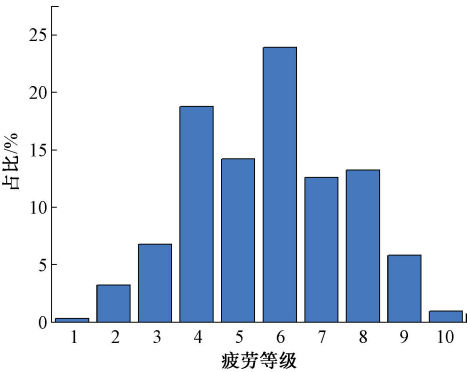


图 6 初始疲劳等级分布  
Fig. 6 Initial fatigue grade distribution

由于 KSS 量表为主观评价量表存在个体差异,例如部分被试可能会高估或低估自身的疲劳状态。因此,在确定疲劳状态标签之前,本研究首先结合客观测量结果对 KSS 量表自评值进行校正对齐。首先,基于被试者在模拟操作中的异常反应时间均值和漏反应次数两种数

据,采用孤立森林算法来刻画不同 KSS 量表自评值所对应样本异常情况。孤立森林算法为每个样本提供异常得分,根据异常得分可以识别出最需要调整 KSS 量表自评值的样本。本文将异常得分排名前 5% 的样本标记为 KSS 自评值异常样本。

采用 K 均值聚类方法对所有样本进行聚类,得到 10 个样本簇(对应 KSS 量值 1~10)。然后计算每个量值在各个簇中出现的频率,将每个簇内最常见的 KSS 量值作为该簇内所有样本的“最可能”KSS 量值标签。之后,将标记的异常样本的 KSS 自评值与估计的“最可能”KSS 量值进行比对,若两者不一致,则将自评值修正为“最可能”KSS 量值。修正后疲劳等级分布如图 7 所示,疲劳等级控制在中间较为常见的范围,没有疲劳极端值 1 和 10,更能反映出作业过程中普遍存在的疲劳程度变化,因为完全没有疲劳或完全耗竭的情况在实际作业环境中较为罕见。

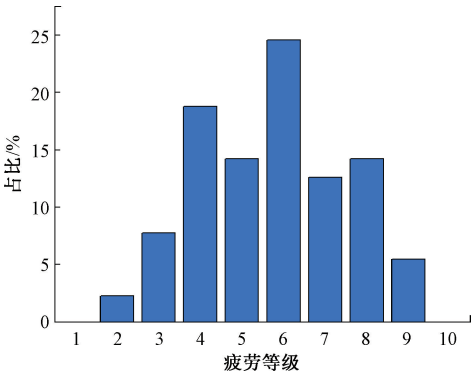


图 7 修正后疲劳等级分布  
Fig. 7 Corrected fatigue grade distribution

## 2.2 评价指标

研究采用包括准确率(accuracy)、查准率(precision)、查全率(recall)和 F1 分数等多种指标对所提疲劳判别模型的性能进行评价,公式如式(8)~(11)所示。

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \times 100\%$$

(8)

$$Precision = \frac{TP}{TP + FP} \times 100\%$$

(9)

$$Recall = \frac{TP}{TP + FN} \times 100\%$$

(10)

$$F1 = \frac{2Precision \times Recall}{(Precision + Recall)} \times 100\%$$

(11)

式中:  $TP$ (true positives) 为正确检测为“疲劳状态”的样本数目;  $TN$ (true negatives) 表示正确检测为“非疲劳状态”的样本数目;  $FP$ (false positives) 为错误检测为“疲劳状态”的样本;  $FN$ (false negatives) 则为错误检测为“非疲劳状态”的样本数目。

3 实验与结果

3.1 参数设置

本文疲劳标签的划分阈值选择为 5,修正后的疲劳等级>5 视为“疲劳状态”,≤5 视为“非疲劳状态”。考虑到数据样本量较小不利于模型训练,容易导致模型快速过拟合,从而产生错误预测偏向。本研究采用添加高斯噪声方法对原视频样本进行数据增强,数据集由原本的 309 例样本扩充 1 倍,至 618 例样本。本研究将数据增强后的样本集随机按 4 : 1 比例划分为训练集和测试集。训练双支路模型的迭代次数设置为 100 次,批大小为 16,选择自适应矩估计(adaptive moment estimation, Adam)作为优化器,学习率设置为 0.001。

3.2 视频帧采样间隔选取

本文每隔  $t$  秒抽取一帧脸部图像作为外观信息以减少相邻视频帧之间存在的高度相关性带来的冗余信息,时空图则由采样间隔  $t$  秒内的视频生成,对应  $t$  秒提取的生理信息。为了探究不同视频帧采样间隔对疲劳判别的影响,研究分别测试了  $t = 5$ 、10、15 s 的疲劳判别网络性能。其中  $t = 5$  s 时所提方法的疲劳判别性能最好。这是由于采样间隔越短人脸外观信息的时间粒度越小,能够抓取面部外观的细微变化,而  $t = 20$  s 时疲劳判别性能最差,准确率仅为 80%。进一步测试不同采样时间模型的训练时长,在配备 NVIDIA GeForce RTX 3090 GPU 的电脑上训练结果如表 2 所示。其中  $t = 5$  s 时的训练时长接近  $t = 10$  s 时训练时长的 2 倍,但准确率只比  $t = 10$  s 时提升了 1%,查准率提升了 2%。因此综合时间复杂度和模型判别准确率,  $t = 10$  s 是最合适的采样时间。

3.3 疲劳判别模型性能比较

比较了本文方法与基于视频帧的端到端三维神经网络模型(Resnet-3D、C3D、S3D)的参数数量。对比模型的输入与本文提出方法相同,训练 100 轮次,结果如表 3 所

示。结果表明,本研究提出的模型具有最小的参数量,模型计算效率高、内存消耗少,对于边缘计算设备的实时应用友好。

表 2 不同视频帧采样间隔结果

Table 2 Results of different video frame sampling intervals

间隔 时间/s	准确率/%	查准率/%	查全率/%	F1 分数/%	训练时间/ min
5	89	94	88	91	36
10	88	92	88	90	20
15	80	96	74	84	13

表 3 不同深度网络模型参数量

Table 3 Parameters of different deep network models

方法	参数量/ $\times 10^9$
Resnet-3D	0.236
C3D	0.748
S3D	0.029
本文方法	0.006

为了进一步比较不同模型在疲劳判别中的准确率,首先基于面部运动提取与疲劳相关特征,包括眼睑闭合持续时间百分比、眼睛纵横比以及嘴巴纵横比<sup>[6]</sup>。在对特征进行归一化处理后,采用多种分类学习方法进行疲劳判别,结果如表 4 所示。结果表明基于面部运动特征的方法的准确率均低于 70%,其中表现最优的随机森林分类方法的准确率分别为 69%。尽管大多数方法的召回率较高,SVM 分类方法的召回率达到 100%,但其准确率和精确率较低。这表明,尽管模型能够识别出大部分疲劳样本,但也存在将非疲劳样本误判为疲劳样本的情况,显示出模型对训练数据的疲劳样本出现了过拟合现象。在所比较的 3 种深度学习方法中,Resnet-3D 模型表现最佳,疲劳判别准确率达到 76%。相比之下,本文提出的基于面部外观和生理表征的双支路模型实现了 88%的疲劳判别准确率,优于其他对比方法的性能。

所有对比方法和采样时间为 10 s 的双支路网络模型测试结果形成的受试者操作特征(receiver operating characteristic, ROC) 曲线如图 8 所示。双支路模型的 ROC 曲线下面积(area under curve, AUC) 值为最大值 0.88,而其他方法的 AUC 值均小于 0.80,说明双支路模型对疲劳状态和非疲劳状态样本的区分能力最强,可以应用于作业人员的认知疲劳判别。

3.4 基于外观和生理网络结构消融实验

为了探究生理表征支路是否能够从时空图中提取出生理特征,利用该支路进行远程心率估计。实验将训练集和测试集按 4 : 1 的比例随机划分,并训练 100 轮次。结果显示,估计心率与真实心率之间的平均误差为-1.64

每分钟每拍 (beats per minutes, bpm), 平均绝对误差为 9.79 bpm, 均在可接受的范围内。该结果验证了生理表征支路在从时空图中提取生理信息方面的有效性。相比手工设计的特征提取方法, 基于深度学习的生理表征支

路能够自动学习面部时空图中蕴含的微弱生理特征, 减少环境噪声对生理信号的干扰, 提高了特征提取的准确性和泛化能力。

表 4 对比方法结果

Table 4 Results of comparison method ( % )					
特征类别	模型	准确率	查准率	查全率	F1 分数
基于面部运动特征 <sup>[6]</sup>	SVM	66	66	100	80
	XGBoost	60	66	80	73
	决策树	66	72	83	77
	随机森林	69	69	98	81
	梯度提升机	66	67	95	79
	K 近邻	68	72	83	77
	朴素贝叶斯	63	66	93	77
	多层感知机	66	67	98	79
	Resnet-3D <sup>[16]</sup>	76	83	81	82
基于视频帧	C3D <sup>[17]</sup>	61	100	61	76
	S3D <sup>[18]</sup>	71	74	78	76
基于面部外观和生理表征	本文方法 ( $t=10\text{ s}$ )	88	92	88	90

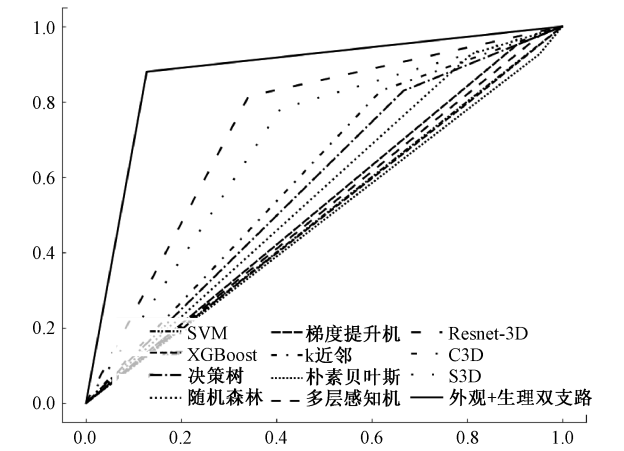


图 8 不同疲劳判别方法 ROC 曲线

Fig. 8 Receiver operating characteristic curves of different methods for fatigue detection

进一步将所提出的双支路策略模型和消融的单支路策略模型在测试集上进行对比测试, 视频帧采样间隔设置为 10 s, 各项评价指标结果如表 5 所示。可以看出, 采用双支路网络模型的疲劳判别整体性能优于单支路模型, 得到的准确率为 88%, 精确率为 92%, 召回率为 88%, F1 分数为 90%。由实验结果表明结合面部外观和生理时空图的双支路网络在疲劳判别任务中具有较高的准确性和可靠性。相较于双支路网络, 仅使用面部外观支路的分类结果表现良好, 准确率为 85%, 而生理表征单支路的性能则明显低于前者, 准确率为 73%。该结果表明, 由于生理信息与疲劳相关性的复杂性和生理信息的多变性导致单独依赖面部视频的生理时空图进行分类存

在一定的局限性。但是通过融合视频帧面部外观和生理时空图信息能够提升疲劳状态判别的准确率。因此, 双支路网络的融合策略显著增强了模型对疲劳状态的判别能力。

表 5 3 种策略模型疲劳判别结果

Table 5 Fatigue discrimination results of three strategy models ( % )				
策略	准确率	查准率	查全率	F1 分数
外观+生理双支路	88	92	88	90
外观单支路	85	87	87	87
生理信号单支路	73	83	71	76

3 种策略模型的混淆矩阵如图 9 所示。基于面部外观的单支路模型预测疲劳状态样本相较于预测非疲劳样本效果更好, 基于生理的单支路模型预测非疲劳样本时相较于预测疲劳样本效果更好。但基于面部外观和生理表征的双支路模型在两类样本上的划分结果都优于单支路策略模型, 表现出更好的疲劳判别效果。

3.5 k 折交叉验证

为了进一步验证模型的稳定性和泛化能力, 研究采用  $k$  折交叉验证法。训练集被划分为  $k$  折, 其中前  $k-1$  折用于模型训练, 最后 1 折作为验证集。研究设置为 5 折交叉验证, 每一折验证集的结果如表 6 所示。通过多次训练和验证, 结果表明模型在不同子数据集上具有较高准确率, 说明本文提出的方法具有稳定性和泛化性, 能够适用于工业场景中的认知疲劳评估。

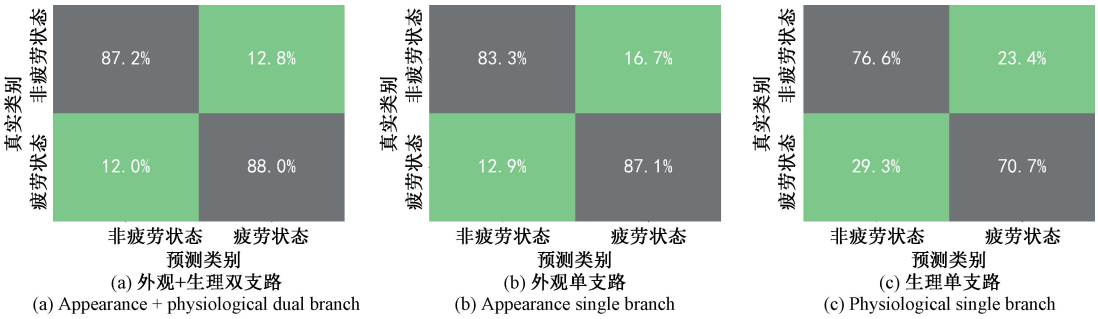


图 9 混淆矩阵

Fig. 9 Confusion matrix

表 6 5 折交叉验证结果

Table 6 5 fold cross validation results

	第 1 折	第 2 折	第 3 折	第 4 折	第 5 折	平均值
准确率/%	81	84	81	94	80	84

4 结 论

本文提出了一种基于面部视频的作业人员疲劳判别方法,将面部外观和生理表征相融合,以提高判别的准确性和适用性。通过双支路网络模型进行多模态特征学习,分别处理视频帧和生理时空图,从而充分提取不同模态的信息。设计的三维卷积网络能够有效捕捉面部视频序列中的时间和空间特征,为疲劳判别提供更全面的特征表示。通过模拟工业生产任务激发认知疲劳试验获取的数据集进行实验验证。结果表明,本文方法在认知疲劳判别中的准确率达到 88%,实现高准确率和无干扰的疲劳判别。与现有先进方法相比,本文研究提高了疲劳判别准确率并节省了计算资源。同时,本文方法无需额外生理传感器设备,为工业场景下的作业人员提供了一种高效、可靠的疲劳监测方案。

然而,研究仍存在一定的局限性。验证数据集主要来源于模拟实验,而非真实现场环境。因此,需要在后续研究进行工业生产场景适配,采集多样化的数据集,以增强模型的适应性和泛化能力。同时,本文仅集中于疲劳与非疲劳状态的二分类判别,未对疲劳程度进行细分评估。未来的研究将致力于探索疲劳程度的判别,以提供更全面的疲劳监测解决方案。

参考文献

[ 1 ] INA V, KIRCHNER C, BOCK O L. Relation between multiple markers of work-related fatigue [J]. Safety and Health at Work, 2016, 7(2):124-129.

[ 2 ] 苏瑞芝,唐巾卜,阿地力·吐合提,等. 基于生理参数的驾驶疲劳检测方法综述[J]. 复旦学报(自然科学版),2023,62(4):419-427.

SU R ZH, TANG J B, TUHATI A, et al. A review of physiological signals-based driving fatigue detection

algorithms [J]. Journal of Fudan University ( Natural Science), 2023, 62(4):419-427.

[ 3 ] 姚贺龙,吕东瀚,张勇,等. 基于傅里叶分解方法的肌肉疲劳状态分类研究[J]. 电子测量与仪器学报, 2023,37(6):48-58.

YAO H L, LYU D H, ZHANG Y, et al. Study of muscle fatigue state classification based on Fourier decomposition method [J]. Journal of Electronic Measurement and Instrumentation, 2023, 37(6):48-58.

[ 4 ] PAN L, YAN C, ZHENG Y, et al. Fatigue detection method for UAV remote pilot based on multi feature fusion [J]. Electronic Research Archive, 2023, 31(1): 442-466.

[ 5 ] 任建新,许锋. 多特征融合的驾驶员疲劳程度量化方法研究[J]. 电子测量技术,2023,46(20):16-23.

REN J X, XU F. Research on the quantification method of driver fatigue level by fusing multiple features [J]. Electronic Measurement Technology, 2023, 46(20): 16-23.

[ 6 ] SAVAS B K, BECERIKLI Y. Real time driver fatigue detection based on SVM algorithm [C]. 2018 6th International Conference on Control Engineering & Information Technology (CEIT), 2018:1-4.

[ 7 ] HOU Z, OU S, XU D. Research on fatigue driving feature detection algorithms of drivers based on machine learning[ J]. System Science and Control Engineering, 2021, 9(1): 167-172.

[ 8 ] GAO Z, LI Y, YANG Y, et al. A recurrence network-based convolutional neural network for fatigue driving detection from EEG [J]. Chaos, 2019, 29(11):113126.

[ 9 ] 冯笑,代少升,黄炼. 基于可解释深度学习的单通道脑电跨被试疲劳驾驶检测[J]. 仪器仪表学报,2023, 44(5):140-149.

FENG X, DAI SH SH, HUANG L. Cross-subject driver fatigue detection from single-channel EEG with an interpretable deep learning model [J]. Chinese Journal of Scientific Instrument, 2023, 44(5):140-149.



- [10] 向洪义. 驾驶疲劳生理特征及识别方法研究 [D]. 重庆: 中国人民解放军陆军军医大学, 2022.  
XIANG H Y. Research on physiological characteristics and identification methods of driving fatigue [D]. Chongqing: Army Medical University, 2022.
- [11] OUYANG Y, LIU M, CHENG C, et al. Monitoring inattention in construction workers caused by physical fatigue using electrocardiograph and galvanic skin response sensors [J]. *Sensors*, 2023, 23(17): 7405.
- [12] DENG P Y, QIU X Y, TANG Z, et al. Detecting fatigue status of pilots based on deep learning network using EEG signals [J]. *IEEE Transactions on Cognitive and Developmental Systems*, 2020, 13(3): 575-585.
- [13] 王宇, 戎舟, 王苏煜. 基于3D残差注意力网络的非接触式心率测量方法 [J]. *国外电子测量技术*, 2023, 42(2): 53-59.  
WANG Y, RONG ZH, SU Y. Non-contact heart rate measurement method based on 3D residual attention network [J]. *Foreign Electronic Measurement Technology*, 2023, 42(2): 53-59.
- [14] NIU X S, SHAN SH G, HAN H, et al. Rhythmnet: End-to-end heart rate estimation from face via spatial-temporal representation [J]. *IEEE Transactions on Image Processing*, 2019, 29: 2409-2423.
- [15] NIU X, HAN H, ZHAO X, et al. Robust remote heart rate estimation from face utilizing spatial-temporal attention [C]. 2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019), 2019: 1-8.
- [16] HARA K, KATAOKA H, SATOH Y. Learning spatio-temporal features with 3D residual networks for action recognition [C]. 2017 IEEE International Conference on Computer Vision Workshops (ICCVW), 2017: 3154-3160.
- [17] TRAN D, BOURDEV L, FERGUS R, et al. Learning spatiotemporal features with 3D convolutional networks [C]. 2015 IEEE International Conference on Computer Vision (ICCV), 2015: 4489-4497.
- [18] XIE S, SUN C, HUANG J, et al. Rethinking spatiotemporal feature learning: Speed-accuracy trade-offs in video classification [C]. 2018 Computer Vision (ECCV 2018). Springer, 2018: 318-335.
- [19] KARTYNNIK Y, ABLAVATSKI A, GRISHCHENKO I, et al. Realtime facial surface geometry from monocular video on mobile GPUs [J]. *ArXiv preprint arXiv. 1907.06724*, 2019.
- [20] SHAHID A, WILKINSON K, MARCU S, et al. Stop, That and one Hundred Other Sleep Scales [M].

Newyork: Springer, 2011: 209-210.

- [21] LI F, YI C, JIANG Y, et al. Different contexts in the oddball paradigm induce distinct brain networks in generating the P300 [J]. *Frontiers in Human Neuroscience*, 2019, 12(520): 1-10.

## 作者简介



**颜文琴**, 2022年于武汉理工大学获得学士学位, 现为四川大学硕士研究生, 主要研究方向为非接触生理信号测量和疲劳判别。

E-mail: yanwenqin@stu.scu.edu.cn

**Yan Wenqin** received her B. Sc. degree from Wuhan University of Technology in 2022.

She is now a M. Sc. candidate at Sichuan University. Her main research interests include non-contact physiological signal measurement and fatigue identification.



**郑秀娟** (通信作者), 2004年于西北工业大学获得学士学位, 2006年于浙江大学获得硕士学位, 2011年于香港理工大学获得博士学位, 现为四川大学电气工程学院自动化系副教授, 主要研究方向为人体多模态信息协同感知与智能分析、情感计算、自然人机交互等。

E-mail: xiujuanzheng@scu.edu.cn

**Zheng Xiujuan** (Corresponding author) received her B. Sc. degree from Northwestern Polytechnical University in 2004, her M. Sc. degree from Zhejiang University in 2006, and her Ph. D. degree from The Hong Kong Polytechnic University in 2011. Now she is an associate professor in the Department of Automation, School of Electrical Engineering, Sichuan University. Her main research interests include multi-modal human information collaborative perception and intelligent analysis, affective computing, and natural human-computer interaction.



**殷中平**, 2005年获于华北电力大学获得学士学位, 现为深圳中广核工程设计有限公司高级工程师, 主要研究方向是人因工程、人机界面设计等。

E-mail: zhongping011@126.com

**Yin Zhongping** received his B. Sc. degree from North China Electric Power University in 2005. He is now a senior engineer at China Nuclear Power Design Company (Shenzhen). His main research interests include human factors engineering and human-machine interface design.



**张学刚**, 2000年于华北电力大学获得学士学位, 2003年于华北电力大学获得硕士学位, 现为中广核工程有限公司核电安全技术与装备全国重点实验室正高级工程师, 主要研究方向为智能人机交互、人因工程与

人因安全和测控技术等。

E-mail: 24457037@ qq. com

**Zhang Xuegang** received his B. Sc. degree from North China Electric Power University in 2000 and his M. Sc. degree from North China Electric Power University in 2003. He is now a senior engineer at the National Key Laboratory of Nuclear Power Safety Technology and Equipment, China Nuclear Power Engineering Co., Ltd. His main research interests include intelligent human-computer interaction, human factors engineering and safety, and measurement and control technology.



**贾明**,2006 年于南京航空航天大学获得学士学位,2009 年于南京航空航天大学获得硕士学位,现为深圳中广核工程设计有限公司研究员级高级工程师,主要研究方向为人因工程。

E-mail: sony\_jm@ 163. com

**Jia Ming** received his B. Sc. degree from Nanjing University of Aeronautics and Astronautics in 2006 and his M. Sc. degree from Nanjing University of Aeronautics and Astronautics in 2009. He is now a senior research engineer at China Nuclear Power Design Company (Shenzhen). His main research interest includes human factors engineering.



**刘伯相**,2018 年于西华大学获得学士学位,2022 年于西华大学获得硕士学位,现为四川大学电气工程学院博士研究生,主要研究方向为非接触式生理信息测量。

E-mail: liubx@ stu. scu. edu. cn

**Liu Boxiang** received his B. Sc. degree from Xihua University in 2018 and his M. Sc. degree from Xihua University in 2022. He is now a Ph. D. candidate at the School of Electrical Engineering, Sichuan University. His main research interest includes non-contact physiological signal measurement.



**涂海燕**,1993 年于成都科技大学获得学士学位,2000 年于四川大学获得硕士学位,2012 年于四川大学获得博士学位,现为四川大学副教授,主要研究方向为生物医学工程,智能控制。

E-mail: haiyantu@ 163. com

**Tu Haiyan** received her B. Sc. degree from Chengdu University of Science and Technology in 1993, received her M. Sc. degree from Sichuan University in 2000 and received her Ph. D. degree from Sichuan University in 2015, She is now an associate professor at Sichuan University. Her main research interests include biomedical engineering and intelligent control.