

文章编号: 2095-2163(2022)11-0001-09

中图分类号: G642.0

文献标志码: A

在线教学视觉隐私保护学生专注度监测研究

刘信鑫, 韩光, 杨海根, 孙宁

(南京邮电大学 宽带无线通信技术教育部工程研究中心, 南京 210003)

摘要: 新冠疫情下, 在线教学已成为学校教学的重要形式, 这就引发了教学活动方式转变, 特别是学生专注度监测等问题面临隐私保护和数据冗余等挑战。为此, 文章引入视觉隐私保护机制, 提出面向在线教学过程的新型智能化学生专注度监测方案, 其系统源于多层压缩感知、Cascade RCNN、改进LBP以及类字典学习等技术创新融合。对各环节, 文章分别选择配套数据集进行实验并分析论证。结果表明: 相较于传统思路, 所提方案在视觉隐私保护、人脸提取和专注度监测等方面具有明显优势。本研究有利于改善在线教学过程的学习质量监管, 同时也为疫情下教学技术发展提供了探索和积累。

关键词: 在线教学; 专注度监测; 视觉隐私保护

Research on students' engagement monitoring in online teaching based on visual privacy protection mechanism

LIU Jixin, HAN Guang, YANG Haigen, SUN Ning

(Engineering Research Center of Broadband and Wireless Communication Technology, Ministry of Education, Nanjing University of Posts and Telecommunications, Nanjing 210003, China)

[Abstract] In the context of COVID-19, online teaching has gradually become an important form of school teaching. And this has led to a significant change in the teaching activities of the teacher and the student. Especially, there are many challenges (such as privacy protection or data redundancy) in students' engagement monitoring. For these problems, a new intelligent student engagement monitoring solution for online teaching is proposed in this paper. Its core lies in the visual privacy protection mechanism which is one of the previous work of the author's team. The proposed system framework is derived from an innovative fusion of multi-layer compressed sensing, Cascade RCNN, improved LBP, and dictionary-like learning, and so on. In view of the main links of the system, this paper selects the supporting data sets to carry out corresponding experiments. And the experimental data are analyzed and demonstrated. Experimental results show that: compared with the traditional thinking, the proposed system has obvious technical advantages in visual privacy protection, face object extraction and engagement monitoring. In a word, this study is not only helpful for assisting teachers to improve the supervision of students' learning quality in online teaching, but also provides some ideas and data accumulation for the development of new teaching techniques under the epidemic situation.

[Key words] online teaching; engagement monitoring; visual privacy protection

0 引言

新冠大流行的产生与持续导致中国学校教学的授课形式发生了根本改变, 过去作为辅助手段的在线教学(也可称为在线课堂)已经逐渐成为一种常态化的主要教学方式, 特别是在各地区疫情严重的阶段几乎是不可替代的选择。因此, 认识和研究在线教学过程的新矛盾、新问题是必要且紧迫的。在传统课堂上, 教师讲授的同时可以通过观察学生群体的专注程度, 达到评估教学方法是否得当的目的, 进而实时调整教育技术来提高整体授课质量。无疑, 这种师生双向的教学状况实时交互是重要且必须的, 但是该过程在向在线课堂迁移的过程中却产

生了新的困难和挑战。当前主流在线教学的本质是视频会议^[1], 其实施过程天然地将教与学分为了2条途径: 一方面, 相比于传统课堂, 在线教学中从教师到学生方向的讲授过程无疑是有效且通透的, 能够基本达到传统课堂的功能替代; 但是, 另一方面, 在线教学过程中从学生到教师方向的学习状态实时反馈则存在明显的不足、甚至缺失, 而这对于教学质量的负面影响可能很大。对此, 本研究尝试运用团队前期面向居家场景提出的视觉隐私保护模型, 设计一套适于视觉隐私保护视频数据的学生专注度监测系统, 通过兼顾解决学生隐私安全、教师监管智能化、以及网络传输应用效率等现有模式的应用痛点, 达到在线课堂的教学质量提升目的。

基金项目: 国家自然科学基金面上项目(61871445)。

作者简介: 刘信鑫(1982-), 男, 博士, 副教授, 主要研究方向: 视频视觉隐私保护编码及智能应用。

通讯作者: 刘信鑫 Email: liujixin@njupt.edu.cn

收稿日期: 2022-10-05

1 在线教学专注度监测

在线教学的技术本质源于视频会议,但是其对于学生专注度的监管方式则相当于视频监控的一种应用拓展。理论上,视频会议的主体双方都有各自的任务,演讲者侧重于发言,参与者侧重于聆听;而在视频监控中,主体任务则偏重于监视者这边,被监视者相对没有任务要求。但是,对于在线教学的教师,除了要完成教学发言(即视频会议的任务)之外,还需关注学生状态(即视频监控的任务),这相比于传统课堂而言是较为难以兼顾的。

事实上,在线教学的学生专注度研究已经成为国内外相关领域的关注热点。近年来,国际上的代表研究主要有:Monkaresi 等人^[2]通过表情和心率的视频人脸评估尝试解决在线教学学生专注度监测。Bhardwaj 等人^[3]针对电子学习环境下学生专注度运用深度学习进行了相关研究。Hasnine 等人^[4]以学生情绪提取与可视化来实施在线学习的专注度检测。Liu 等人^[5]将情绪与认知专注度检测用于慕课的教学目标达成预判。国内同类研究的代表成果包括:李振华等人^[6]引入模型集成的思路来处理在线学习投入度评测。付长风^[7]基于“课堂在线”微课平台开展了学习投入度的相关研究。左国才等人^[8]选择了深度学习模型进行学生课题参与行为研究。

综上所述,现有在线教学的专注度监测实际就是视频监控和视频会议的综合应用。但是,这种典型的实时视频流交互监管模式与传统教学模式之间是存在矛盾的,概括讲有3个主要方面:

(1)不符合学生的居家听课形式。在线课堂能够替代教师向学生讲授的过程,但是并不能替代传统课堂环境,特别是疫情期间多名学生同在宿舍或与家人共处一室等情况普遍存在,让学生全程开启语音视频有时并不适宜可行。

(2)不符合教师的在线教学形式。即便学生能够全程语音视频,但是教师讲授课程的同时也很难像传统课堂一样通过视觉直观地兼顾大规模学生的听课状况,有必要提供能够替代人眼监测的辅助手段。

(3)不符合国内现有常规网络环境。在线课堂中教师与学生如果想模拟正常情况的视觉交互,其数十或上百人的实时视频码流传输对带宽资源消耗很大,这对于目前手机或电脑设备动辄1080p、甚至4k的摄像头配置也是不小的挑战,尤需指出的是教师人眼监测的低效执行难以匹配这种数据开销的高昂代价。

针对上述问题,本研究通过引入多层压缩感知视觉隐私保护机制用以实现学生居家隐私保护和视频数据降维保真的平衡,进而辅助教师达到视觉隐私保护数据态下在线教学过程学生专注度的高效智能监测,系统架构如图1所示。

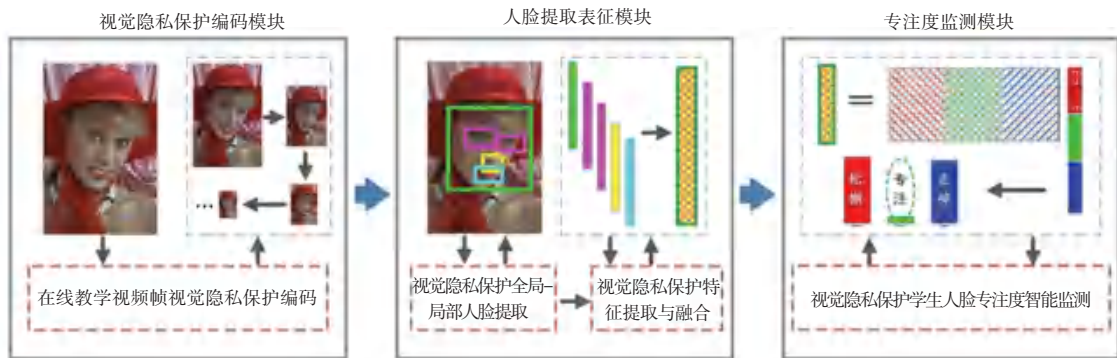


图1 本文系统的总体架构

Fig. 1 General architecture of the proposed system in this paper

2 面向在线教学的视觉隐私保护专注度监测系统

针对在线教学的视觉隐私保护学生专注度监测需求,系统的宏观架构上划分为视觉隐私保护编码、人脸提取表征和专注度监测共3个模块。具体来讲,视觉隐私保护编码模块的任务在于实现在线教学视频帧层面的视觉隐私保护编码;人脸提取表征模块是

针对视觉隐私保护视频帧进行人脸提取和特征融合两方面任务;专注度监测模块则侧重于面向学生人脸视觉隐私保护特征实施智能化专注度监测。

2.1 基于多层压缩感知模型的视觉隐私保护编码

视觉隐私保护^[9]是本团队前期提出的一种针对图像或视频数据的隐私保护新概念,而多层压缩感知模型是自主研究形成的一种实现视觉隐私保护编码的理想工具。以此为基础,本次研究不仅解决

了该理论从定性到定量的成熟优化^[10],而且已将其用于居家老人跌倒检测等应用问题^[11],最新工作获得了相关领域较高的学术认可^[12]。因此,多层压缩感知模型对于解决视觉隐私保护编码的应用需求具有天然的匹配优势。

相比于传统压缩感知采样,其多层化扩展能够在继承数据高保真的同时大幅降低数据规模。经典的压缩感知采样过程是类似如下的一套降维投影运算:

$$y_{M \times 1} = \Phi_{M \times N} x_{N \times 1} = \{ \langle x, \phi^{(i)} \rangle \mid i=1 \}^M \in \mathbb{R}^N \quad (1)$$

其中, $x \in \mathbb{R}^N$ 表示源信号; y 表示感知数据; 观测矩阵 Φ 为投影轴集合 $\{ \phi^{(i)} \mid \phi^{(i)} \in \mathbb{R}^N \mid i=1 \}^M$, 这里通过降维条件 $M \ll N$ 实现采样压缩集成。

虽然经典压缩感知采样已具备了视觉隐私保护的可能,但是由于全局化的处理思路,使得数据状态过于紧凑,从而缺乏应对不同智能应用的灵活性。为此,采取了一种宽松化程度更高的分块压缩感知编码,即:

$$y^{(j)} = \langle \phi_{m \times n}^{(j)}, x_{m \times n}^{(j)} \rangle \quad (2)$$

此时 $x = \{ x_{m \times n}^{(j)} \}_{j=1}^{(M \times N)/(m \times n)}$ 而 $\Phi = \{ \phi_{m \times n}^{(j)} \}_{j=1}^{(M \times N)/(m \times n)}$, 其中各 ϕ 需满足压缩感知观测矩阵构造要求,如高斯随机、伯努利随机、托普利兹随机等。

依托上式就形成了分块压缩感知采样的单层模型,这就在理论上具备了传统压缩感知采样的一系列同等属性,并且能够保障多层化扩展的智能化灵活度大幅提高。以压缩感知中最为重要的有限等距约束为例,一种具现化的多层化扩展模型为:

$$Y^{(c+1)} = \begin{cases} \Phi^{(c)} X & c = 0, \Phi^{(0)} = \text{random} \{ \phi \} \\ \Phi^{(c)} Y^{(c)} & c \geq 1, \Phi^{(c)} = F(\Phi^{(c-1)}) \end{cases} \quad (3)$$

此时原始的有限等距约束属性将逐层继承,可以保证多层扩展后的数据保真度几乎等同于初始层采样效果。图 2 给出了多层压缩感知模型的机制示意,同时从人脸图像的编码实例可看出处理后的视觉隐私保护效果。

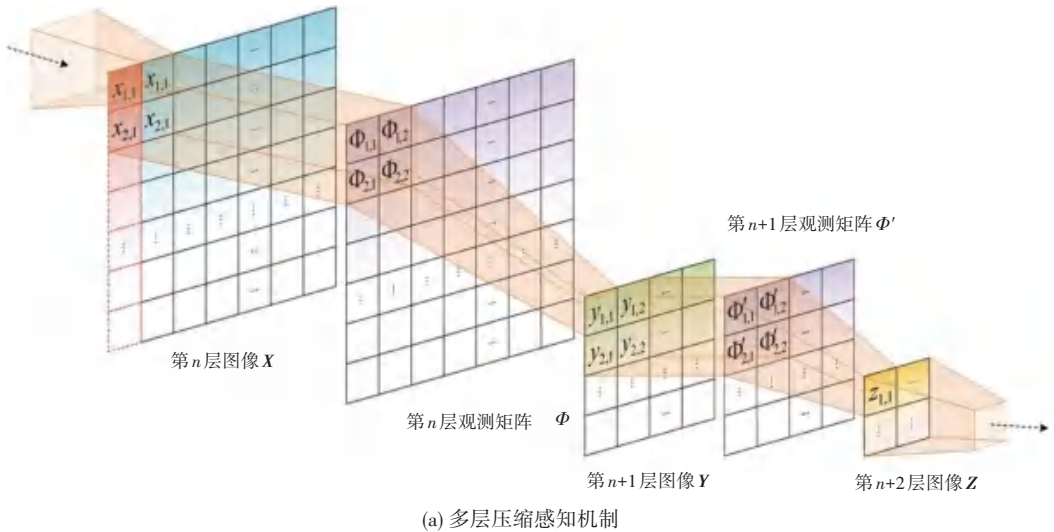


图 2 多层压缩感知机制及视觉隐私保护效果示例

Fig. 2 Multilayer compressed sensing (MCS) mechanism and visual privacy protection (VPP) via MCS

本团队的前期工作^[9-10]表明,多层机制中 3 层及以后的数据形态在主流主客观视觉评价指标下都能达到较高的隐私保护等级,这意味着可将其作为一种优质经验模型用于隐私需求迫切的智能应用场景,事实上在居家跌倒检测等问题上该思路已取得良好验证^[11-12]。因此,本文将在延续上述思路的

基础上用其实现在线教学的视频帧视觉隐私保护编码。

2.2 视觉隐私保护视频帧全局-局部人脸目标提取
视频形态下人脸无疑是学生专注度的关键信息载体,对此相关研究^[2]已经给出了明确的论证和探讨。事实上,相比于人脸整体,主流思路对于局部五

官所蕴含的信息往往更加重视,这在情绪识别等关联领域已经形成了一定的共识^[13-16]。因此,本环节的处理重点不仅要解决全局人脸的实时检测,还需兼顾到局部五官的有效提取。

人脸检测的本质可以理解为目标检测的一种特例,其中公认的经典方法是 Viola 等人^[17]提出的 VJ (Viola-Jones) 检测算法。该算法运用类 Haar 特征结合级联 Adaboost 机制实现了快速高效的人脸检测,凭借其优异性能被广泛应用于手机等智能设备上。但是, VJ 为代表的传统目标检测大多依赖于手工设计特征,这对于自然场景等复杂情况明显是存在缺陷的,因此深度学习等鲁棒性更高的方法逐渐占据主流。在深度学习思路下,目前已形成了单阶段和双阶段两种代表思路。简单讲,单阶段算法对目标的定位和分类一步完成,而双阶段则多了一个候选框生成的步骤。

一方面,考虑到本文的核心任务是专注度监测,这就意味着对于人脸检测的精确性要求是相对更重要的,因此选取深度学习的双阶段目标检测算法更为契合。另一方面,视觉隐私保护编码对视频数据的视觉形态存在一定影响,对此传统算法中弱分类器级联强化的机制优势非常值得借鉴。综合两方面因素, Cascade RCNN^[18] 成为了能够兼顾二者的理想方案。图 3 给出了 Cascade RCNN 的架构示意。图 3 中,“I”表示输入图;“conv”表示主干网络;“pool”表示分区特征提取;“H”表示网络头;“B”表示边界框;“C”表示分类器。由此架构可见, Cascade RCNN 在继承典型双阶段目标检测架构的基础上以级联形式进行了拓展创新。

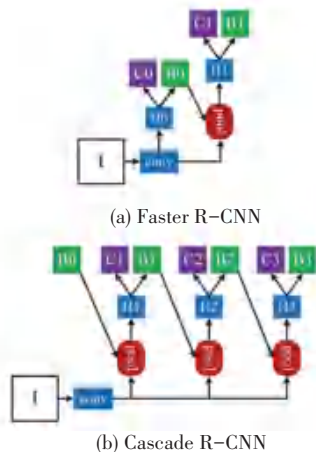


图 3 Cascade R-CNN 与典型双阶段目标检测算法的架构对比^[18]

Fig. 3 Comparison between Cascade R-CNN and typical two-stage target detection algorithm^[18]

现有深度学习架构(包括 Cascade RCNN)在人

脸提取中的有效性基本都是经过实践检验的,但是这些方法的成立大多存在一个潜在前提,即输入的视频帧序列应当具备正常、甚至高清的视觉水平。然而,视觉隐私保护编码在视觉层面产生的影响,会产生一个新的问题:视觉隐私保护视频还能使用常规人脸提取工具吗?事实上,多层压缩感知的高保真优势,确保了视觉隐私保护人脸提取的天然可行性。在多层压缩感知模型中,各层的编码过程依然遵循了压缩感知的所有基本特性,其中数据保真的关键依赖于观测矩阵的有限等距属性。针对公式(2),所谓有限等距约束是指各分块观测矩阵均满足如下关系:

$$(1 - \delta) \|\mathbf{x}_{m \times n}^{(j)}\|_2^2 \leq \|\langle \boldsymbol{\phi}_{m \times n}^{(j)}, \mathbf{x}_{m \times n}^{(j)} \rangle\|_2^2 \leq (1 + \delta) \|\mathbf{x}_{m \times n}^{(j)}\|_2^2 \quad (4)$$

其中, δ 为有限等距参数。

压缩感知理论规定:观测矩阵构造方式是影响压缩感知数据处理质量的决定性要素,而观测矩阵的优劣主要通过有限等距属性来反映。由于本团队的分块压缩感知编码机制^[9-10]天然继承了观测矩阵的理论特性,因此视觉隐私保护视频可视作原始视频数据的一种高保真变体。换言之,面向原始视频的目标提取算法同样可适用于视觉隐私保护视频帧形态。图 4 展示了 Cascade RCNN 在原始态和视觉隐私保护编码(3 层压缩感知)下提取的全局和局部人脸效果。图 4(a)~(d)中,左边图像为原始态,右边图像为视觉隐私保护。图 4 中,图 4(a)和图 4(b)是单人情况,人脸全局保持一致,而局部五官即便在戴墨镜的情况下也绝大部分一致;图 4(c)和图 4(d)是多人情况,在视觉隐私保护下面画面主目标的提取依然有效,其人脸全局和局部也均保持一致。

在获得视觉隐私保护全局-局部人脸的基础上,即可进入到智能数据处理阶段。本质上,该阶段的学术内核是典型的模式识别问题,因此特征描述和分类器设计是本阶段难以回避的 2 项重点任务。前一项任务中,关于视觉隐私保护人脸的特征描述, LBP (Local Binary Pattern) 算子^[19]是非常理想的选择。无论是全局人脸、还是各个五官,相比于整个视频帧来讲都可以算作“局部”的范畴,因此采用 LBP 机制可获得如下特征描述形式:

$$F_{LBP} = \sum_{p=0}^{P-1} s(g_p - g_c) 2^p \quad (5)$$

其中, g_c 为 LBP 算子的窗口中心; g_p 为窗口中心的邻居像素、即 $p = 0, \dots, P - 1$; $s(\cdot)$ 为二值化函数。对于单个图像块窗口半径为 R 的 P 个邻居点,

LBP 算子的模式水平可达 2^P 。显然, 对于多图像块融合的情况, 特别是串联融合的形式, 经典 LBP 的特征维度将大幅增加, 这还会导致数据的冗余度过高, 从而影响智能监测的性能。对此, 本团队前期工作^[9-10]中已经针对视觉隐私保护人脸引入了基于等价模式 (Uniform Pattern) 的 LBP 改进。简单讲, 针对经典 LBP 模式中 0-1 或 1-0 的跳变规律, 将最多 2 次跳变的情况定义为等价模式, 从而形成特征维度的极大约减。新的特征描述形式如下:

$$F_{UP-LBP} = \begin{cases} \sum_{p=0}^{P-1} s(g_p - g_c) & \text{若 } U(F_{LBP}) \leq 2 \\ P + 1 & \text{其余情况} \end{cases} \quad (6)$$



图 4 人脸目标提取示例

Fig. 4 Face object extraction examples

这里函数 $U(\cdot)$ 满足:

$$U(F_{LBP}) = |s(g_{p-1} - g_c) - s(g_0 - g_c)| + \sum_{p=1}^{P-1} |s(g_p - g_c) - s(g_{p-1} - g_c)| \quad (7)$$

由此可将模式水平从 2^P 降低至 $P(P-1) + 1$, 从而在保持表征质量的前提下达到维度约减的

效果。

2.3 视觉隐私保护时序人脸特征融合的专注度监测

作为智能数据处理阶段的另一个重要任务, 分类器设计也使得本研究进入到专注度智能监测的后一项、数据处理关键环节。面向视觉降质数据的分类任务, 稀疏识别方法具有独特优势。视觉隐私保护人脸在隐私保护的同时, 其典型的视觉特征往往存在不同程度的损失, 如角点或边缘等几何特征。因此, 不仅是特征提取需要采用纹理等鲁棒性高的描述形式, 分类器设计也应以鲁棒性优势为重点选取依据。在诸多分类器方案中, 稀疏识别在人脸应用中的高鲁棒性特点是学术界较为公认的^[20]。在相关领域, 本团队前期也有一定的积累^[21-23], 特别是针对特征融合的情况提出了鲁棒性能较好的类字典学习方法。在本文研究的相关领域, 目前可公开获得的数据集中最符合要求的是由 Kamath 等人^[24]构建的慕课学习者数据集, 其数据标注可理解为“专注”、“松懈”和“走神”共 3 类。因此, 本研究涉及的专注度监测即可等价于模式识别的经典多分类问题, 相应地, 本团队的前期研究^[21]已归纳出可选的稀疏分类机制, 主要有如下 3 种代表形式。

(1) 基于稀疏表示的分类器。假设类别数为 k 的训练样本集 $T = \{T_i\}_{i=1}^k$, 则稀疏识别的分类计算过程如下:

$$\hat{\alpha} = \arg \min_{\alpha} \|\alpha\|_1 \quad \text{s.t.} \quad T\alpha = F \quad (8)$$

其中, α 为特征 F 在训练集 T 上的稀疏表示, $\hat{\alpha}$ 为上述优化问题的最优解。由此, 分类器的判定依据如下:

$$\text{classification}(F) = \underset{i=1, \dots, k}{\operatorname{argmin}} \|F - T\delta_i(\hat{\alpha})\|_2 \quad (9)$$

其中, δ_i 为第 i 类的稀疏取值。

(2) 基于字典学习的分类器。字典学习的稀疏分类则将训练集按类拆分后进行逐个训练, 其过程如下:

$$\min_{D, \Gamma} \|T_i - D\Gamma\|_F \quad \text{s.t.} \quad \|\Gamma_j\|_{0, j=1, \dots, N_{T_i}} \leq S \quad (10)$$

其中, D 为第 i 类的稀疏字典, Γ 为 T_i 在该字典下的稀疏表示。在此基础上, 分类器判据调整如下:

$$\text{classification}(F) = \underset{i=1, \dots, k}{\operatorname{argmin}} \|F - D_{T_i}(D'_{T_i}D_{T_i})^{-1}D'_{T_i}F\|_2 \quad (11)$$

其中, D'_{T_i} 为 D_{T_i} 的转置。

(3) 自主提出的类字典学习分类器。类字典学

习是本团队自主提出的一种新型稀疏分类方法,其有效性已经在人脸识别^[21, 23]和场景识别^[22]等智能领域得到了一定验证。本质上,该方法相当于稀疏表示和字典学习的一种综合创新,其核心分类机制如下:

$$\hat{\alpha} = \arg \min_{\alpha} \|\alpha\|_1 \quad \text{s.t.} \quad D_T \alpha = F \quad (12)$$

这里,公式(8)的训练集 T 被替换为相应的学习字典。而字典 $D_T = \{D_{T_i}\}_{i=1}^k$ 的学习可参照公式(10)的相关过程,由此分类器的判据调整为:

$$\text{classification}(F) = \arg \min_{i=1, \dots, k} \|F - D_T \delta_i(\hat{\alpha})\|_2 \quad (13)$$

根据分类器判据的结果,可形成当前视频帧内视觉隐私保护人脸的专注度等级归类,从而实现在线教学视觉隐私保护条件下学生学习专注程度的实时监测和量化评估。

3 实验验证及分析

3.1 实验数据集

本研究讨论的重点聚焦在3个问题,即:视觉隐

私保护的处理效果、视觉隐私保护人脸目标的提取效果、以及视觉隐私保护专注度的监测效果。因此,实验的设计和数据集的选择都是围绕上述问题进行的:针对视觉隐私保护方法,选取面向视觉质量研究的代表性数据集 LIVE(Laboratory for Image & Video Engineering)^[25],包含29幅参考图、5种失真,共779幅图像;针对视觉隐私保护人脸提取,选择非受限人脸识别研究的代表性数据集 LFW(Labeled Faces in the Wild)^[26],共有13 233张人脸图像,每张尺寸为250×250,共5 749人;针对视觉隐私保护专注度监测,选择的 Kamath 等人^[24]构建的慕课学习者数据集,包含23人,其专注度等级“Very engaged”、“Nominally engaged”和“Not engaged”,分别对应于本文的“专注”、“松懈”和“走神”三类。图5展示了上述数据集的一些样本情况。实验的软硬件条件为:处理器 Intel i9-11900K,内存64 G,显卡 NVIDIA RTX3090,操作系统 Ubuntu 和深度学习架构 PyTorch。



图5 不同数据集的样本示例

Fig. 5 Some samples of different datasets

3.2 实验过程

3.2.1 视觉隐私保护效果的验证与分析

为验证多层压缩感知模型的视觉隐私保护效果,本环节选取4种经典视觉降质方法和本研究方案进行对比实验,具体包括:离焦模糊、运动模糊、高

斯噪声、椒盐噪声和压缩感知(本文方案)。相关方法的详细内容参见本团队前期工作^[9]。以人脸数据为例,图6展示了上述方法的视觉隐私保护的直观效果。

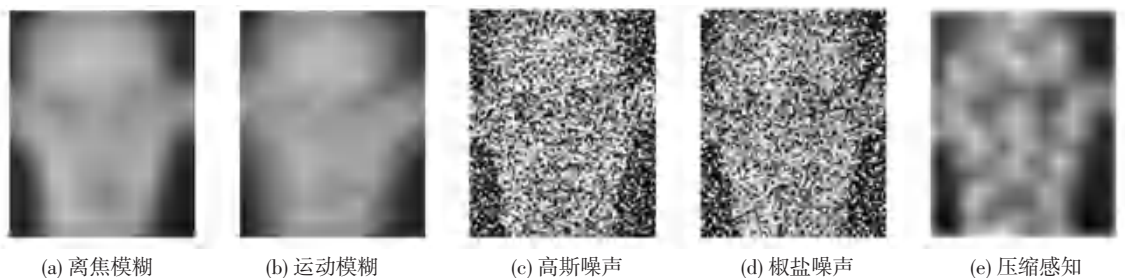


图6 本研究所用视觉隐私保护方法的直观示例

Fig. 6 Visual effect examples of the proposed VPP method

本团队前期研究^[9]表明,无参考图像质量评价方法中的 SFA(semantic feature aggregation)、视觉安全性评估方法中的 LE(Local Entropy)、以及本团队针对多层压缩感知模型自主提出的 VPLE(Visual Privacy-preserving Level Evaluation)是当前相关技术中较为适合视觉隐私保护编码的质量评价工具。依据前期研究经验,关于这些数值的有效性和可信度等主要从单调性、一致性和准确性三方面衡量,对应的指标有: *SROCC* (Spearman Rank Order Correlation Coefficient)、*RMSE* (Root Mean Square Error) 和 *PLCC* (Pearson Linear Correlation Coefficient)。其中,针对单调性还有另一个指标 *KROCC* (Kendall Rank Order Correlation Coefficient) 可选,但其作用与 *SROCC* 基本等效,因此这里只用其一即可。

表 1~3 给出了所提出的 5 种视觉隐私保护方法在 3 种视觉隐私保护评价方法下的 *SROCC*、*RMSE* 和 *PLCC* 结果。单调性方面,*SROCC* 的取值在图像质量评价时通常在 0 - 1 之间,越大表示一致性越好,而在视觉隐私保护中则越小越好;一致性方面,*RMSE* 的取值在图像质量评价时越接近 0 越好,在视觉隐私保护中也遵循同样规律;准确性方面,*PLCC* 的取值在图像质量评价时从 -1 到 1,无论正负越远离 0 越好,而视觉隐私保护中则相反。结果表明,本研究采用的压缩感知方案在单调性和一致性方面有着显著优势,而准确性方面除 SFA 外也同样具有较好的性能,即便在 SFA 的 *PLCC* 中也是能达到最接近最优水平的次优结果。因此,针对自然场景随机成像的条件下,可以认为多层压缩感知能够较好地满足在线课堂的视觉隐私保护需求。

表 1 不同方法评价的 *SROCC* 结果

Tab. 1 *SROCC* results of different VPP methods

| | 离焦 模糊 | 运动 模糊 | 高斯 噪声 | 椒盐 噪声 (本团队) | 压缩感知 |
|-----------|----------|----------|----------|----------------|----------------|
| SFA | 0.516 6 | 0.525 3 | 0.496 8 | 0.496 6 | 0.490 8 |
| LE | 0.377 4 | 0.377 1 | 0.372 7 | 0.371 1 | 0.360 9 |
| VPLE(本团队) | 0.495 3 | 0.510 1 | 0.432 9 | 0.433 3 | 0.427 2 |

表 2 不同方法评价的 *RMSE* 结果

Tab. 2 *RMSE* results of different VPP methods

| | 离焦 模糊 | 运动 模糊 | 高斯 噪声 | 椒盐 噪声 (本团队) | 压缩感知 |
|-----------|----------|----------|----------|----------------|----------------|
| SFA | 0.659 6 | 0.817 1 | 0.926 9 | 0.971 1 | 0.579 6 |
| LE | 1.993 9 | 2.112 3 | 2.796 3 | 2.841 2 | 1.713 3 |
| VPLE(本团队) | 0.424 8 | 0.445 2 | 0.762 8 | 0.733 3 | 0.386 2 |

表 3 不同方法评价的 *PLCC* 结果

Tab. 3 *PLCC* results of different VPP methods

| | 离焦 模糊 | 运动 模糊 | 高斯 噪声 | 椒盐 噪声 (本团队) | 压缩感知 |
|-----------|----------|----------------|----------|----------------|----------------|
| SFA | 0.618 2 | 0.603 7 | 0.599 2 | 0.598 6 | 0.607 9 |
| LE | 0.567 6 | 0.527 3 | 0.512 6 | 0.520 3 | 0.512 4 |
| VPLE(本团队) | 0.661 7 | 0.686 9 | 0.616 9 | 0.621 1 | 0.612 2 |

3.2.2 视觉隐私保护人脸目标提取的验证与分析

为验证视觉隐私保护数据的人脸目标提取效果,本环节选取一些最具代表性的人脸检测算法和 Cascade R-CNN 进行对比实验,主要借鉴文献[18]的思路,选取了单阶段代表算法 YOLO 和双阶段代表算法 Faster R-CNN,而算法对比的衡量指标主要选择了平均精度和提取时间两项。考虑到对比的公平性,以 Cascade RCNN 提出的时间阶段为参照,YOLO 并非最新的 v5 版、而是同时期的 v3 版,Faster R-CNN 也采用的是同时期基于 FPN (Feature Pyramid Network) 的版本。数据选取策略参照文献[21],从 LFW 中样本个数超 25 的类别中随机选取 30 个,并进行压缩感知编码。同时,结合 LFW 的特点,以各样本中心点在横纵坐标约 70% 范围内的 VJ 人脸提取的同尺寸下采样为基准。

表 4 展示了几种代表性人脸检测算法在视觉隐私保护 LFW 数据上的提取效果,具体来看:就深度学习架构的对比,双阶段思路基本优于单阶段思路;就主干网络的对比,ResNet 系列的性能也大多好于 Darknet;就参数选择的对比, *Batchsize* 的增大有利于精度方面的性能提升,但提取速度(这里即指测试速度)的差异并不明显。因此,实验结果表明本研究所提以 Cascade R-CNN 为基础实施视觉隐私保护人脸目标提取的方案在可行性和实用性等方面得到了一定支撑。

表 4 几种代表性算法的视觉隐私保护人脸提取结果

Tab. 4 VPP face object extraction results via some representative algorithms

| | 主干网络 | <i>Batchsize</i> | 平均精度/% | 提取时间/s |
|---------------|------------|------------------|-------------|--------------|
| YOLO | DarkNet-53 | 4 | 70.4 | 0.026 |
| | DarkNet-53 | 16 | 74.6 | 0.028 |
| Faster R-CNN | ResNet-101 | 4 | 81.2 | 0.095 |
| | ResNet-101 | 16 | 84.3 | 0.102 |
| Cascade R-CNN | ResNet-101 | 4 | 83.4 | 0.042 |
| | ResNet-101 | 16 | 86.6 | 0.047 |

3.2.3 视觉隐私保护专注度监测的验证与分析

为验证视觉隐私保护改进 LBP 特征下不同分

类器的专注度监测效果,本环节选取几种代表性算法进行对比实验,具体包括:最近邻(Nearest Neighbor, NN)、支持向量机(Support Vector Machines, SVM)、稀疏表示、字典学习、以及本团队提出的类字典学习。由于在线课堂专注度研究的特殊性,Kamath数据集的规模相对较小(视觉隐私保护编码的数据就更少),这导致深度学习相关的分类器性能难以发挥,因此本实验主要采用了非深度学习方法。为保证实验公平性,训练和测试样本在交叉验证过程的随机选取规模借鉴了文献[24],并且每次实验中各类别采用one-vs-all的统计形式,分别记录并计算原始态和视觉隐私保护的500次平均。

表5 几种代表性算法的视觉隐私保护专注度监测准确率

Tab. 5 VPP engagement monitoring accuracy via some representative algorithms

| | | NN | SVM | 稀疏表示 | 字典学习 | 类字典学习(本团队) | % |
|-------|----------|-------|-------|-------|-------|--------------|---|
| “专注”类 | 原始视频编码 | 62.14 | 63.97 | 81.01 | 85.86 | 91.15 | |
| | 视觉隐私保护编码 | 59.59 | 60.02 | 78.12 | 82.18 | 89.01 | |
| “松懈”类 | 原始视频编码 | 54.91 | 55.26 | 66.64 | 71.01 | 72.11 | |
| | 视觉隐私保护编码 | 51.22 | 50.87 | 62.96 | 69.70 | 69.97 | |
| “走神”类 | 原始视频编码 | 78.92 | 83.05 | 93.17 | 91.33 | 94.86 | |
| | 视觉隐私保护编码 | 77.60 | 81.46 | 92.03 | 90.24 | 93.25 | |

4 结束语

专注度监测技术有利于帮助老师掌握在线教学的学生学习质量,而视觉隐私保护处理则可有效平衡学生的隐私保护诉求和视频流的数据冗余困境。因此,本研究能够较好契合疫情条件下线上教学的技术发展需求。从实验结果看,多层压缩感知编码、Cascade R-CNN、改进LBP以及类字典学习等自研为主的数据处理技术,能够有效满足宏观和局部层面等各方面的系统构建具体需要,从而为在线课堂的视觉隐私保护专注度监测提供了一种较为可行的方案探索。当然,由于相关研究及应用领域较为前沿,目前国内外在数据储备及研发经验等方面普遍存在一定的不足或缺,后续将针对数据集、评价体系以及验证标准等方面开展更多的攻关和突破,以便该技术能够尽早应用于实际的线上教学场景。

参考文献

- [1] CAMILLERI M A, CAMILLERI A C. Remote learning via video conferencing technologies Implications for research and practice [J]. *Technology in Society*, 2022, 68: 101881.
- [2] MONKARESI H, BOSCH N, CALVO R A, et al. Automated detection of engagement using video-based estimation of facial

表5作为几种算法的监测统计结果,可以提供3方面的解读:从代表性方法看,稀疏类3种方法普遍优于经典思路,即NN和SVM,而类字典方法由于集成了稀疏表示和字典学习的优势,其正确率在稀疏类方法中为最佳;从数据集类别划分看,“专注”和“走神”类相对较容易监测,而“松懈”类可能由于数据标签的主观标定方式,其监测效果还有一定的提升空间;从数据形态看,视觉隐私保护的监测结果略低于原始视频,但其微弱损失相对于隐私保护的增强来讲是能够接受的。概括起来,类字典学习对于视觉隐私保护的专注度监测具有较好的效果,而针对“松懈”类或者是类别标定方式的改进可能会有利于系统整体性能的进一步提高。

expressions and heart rate [J]. *IEEE Transactions on Affective Computing*, 2017, 8(1): 15-28.

- [3] BHARDWAJ P, GUPTA P K, PANWAR H, et al. Application of deep learning on student engagement in e-learning environments [J]. *Computers and Electrical Engineering*, 2021, 93: 107277.
- [4] HASNINE M N, BUI H T T, TRAN T T T, et al. Students' emotion extraction and visualization for engagement detection in online learning [J]. *Procedia Computer Science*, 2021, 192: 3423-3431.
- [5] LIU S, LIU Shiqi, LIU Zhi, et al. Automated detection of emotional and cognitive engagement in MOOC discussions to predict learning achievement [J]. *Computers & Education*, 2022, 181: 104461.
- [6] 李振华, 张昭理, 刘海. 基于模型集成的在线学习投入评测方法研究 [J]. *中国远程教育*, 2020, 10: 9-16.
- [7] 付长风. “课堂在线”教学平台在课堂教学中的应用与实践 [J]. *智能计算机与应用*, 2020, 10(10): 194-195, 199.
- [8] 左国才, 苏秀芝, 王海东, 等. 基于CNN深度学习模型的大学生课堂行为检测研究 [J]. *智能计算机与应用*, 2020, 10(2): 158-160.
- [9] LIU Jixin, TANG Zheng, SUN Ning, et al. Visual privacy-preserving level evaluation for multilayer compressed sensing model using contrast and salient structural features [J]. *Signal Processing: Image Communication*, 2020, 89: 115996.
- [10] 汤正, 刘佳鑫, 孙宁, 等. 面向多层压缩感知图像的内容隐私保护度评价方法 [J]. *陕西师范大学学报(自然科学版)*, 2020, 48(02): 58-68.

(下转第17页)