文章编号：2095-2163(2020)11-0008-08 中图分类号：TP183 文献标志码：A

# 课堂场景下学习者情感识别研究

苏 超，王国中
（上海工程技术大学 电子电气工程学院，上海 201600）

**摘 要**：课堂上学生的面部表情和姿态是学习状态的一种自然流露，能够反映出学习者当前的学习状态。而传统的情感识别方法存在识别准确率低、特征提取困难以及实时性差等问题。针对上述问题，本文提出了一种基于表情和姿态的双模态情感识别模型。该模型主要由二部分构成：一是针对学习者的表情和姿态识别，在 Tiny_YOLOv3 目标检测算法基础上，通过加入注意力机制 SEBlock，改进原模型的卷积结构，并采用 GIoU loss 改进损失函数，利用 K-means 算法在自主构建的数据集上聚类，得到适合学习者情感识别的 anchor，最终得到适合于学习者情感识别的 ER_Tiny_YOLOv3 模型。二是针对多模态采用决策层融合方法，进行最终结果的判定，提出针对课堂学习者的融合方法。实验结果表明，该模型相比于 Tiny_YOLOv3，mAP@0.5 提升了 17%，Precision 提升了 35%，F1 分数提升了 22.6%。

**关键词**：情感识别；Tiny_YOLOv3；SEnet；GIoU loss；K-means

## Research on learner's emotion recognition in classroom scene

SU Chao，WANG Guozhong
（School of Electronic and Electrical Engineering，Shanghai University of Engineering and Science，Shanghai 201600，China）

**【Abstract】** Facial expressions and postures of students in the classroom are a natural expression of their learning state，which can reflect the student's current learning state. Traditional emotion recognition research methods have problems such as low recognition accuracy，difficult feature extraction and poor real-time performance. In response to the above problems，a dual-modal emotion recognition model based on expressions and gestures is proposed，which is mainly composed of two parts：The first part is aimed at student's expression and gesture recognition. Based on the Tiny_YOLOv3 target detection algorithm，the convolution structure of the original model is improved by adding the attention mechanism SEBlock，and the loss function is improved by GIoU loss. At the same time，the K-means algorithm is used to build independently Clustering on the dataset to get an anchor suitable for student emotion recognition，and finally get a model ER_Tiny_YOLOv3 suitable for student emotion recognition. The second part aims at the multi-modal fusion method，adopts the decision layer fusion method to judge the final result，and proposes a fusion method for classroom students. Experimental results show that compared to Tiny_YOLOv3，the model has improved mAP@0.5 by 17%，Precision by 35%，and F1 score by 22.6%.

**【Key words】** emotion recognition；Tiny_YOLOv3；SEnet；GIoU loss；K-means

## 0 引 言

近年来，随着互联网技术的快速发展以及大数据、云计算、AI(人工智能)等新技术的发展，智能化教育成为一种新的教育趋势。在《计算神经科学前沿》杂志中提到："智能化教育中一直被忽略的词就是情感"[1]，而情感作为一种非智力因素，能够影响和调节认知活动。心理学研究表明：积极的情感有助于激发学习者的学习动力、培养学习兴趣，促进认知过程；而消极的情感则会影响学习者的耐心度、注意力，阻碍认知过程[2-3]。

人对情感的表达是复杂且微妙的，同样对情感的识别和解读也是多通道协同完成的，包括表情、姿态、语言和声调等[4]。当前，针对情感识别，研究者们主要围绕生理信号、心理测量以及外显行为展开研究[5]。

其中，又以基于面部表情的情感识别居多。虽然面部表情能够表达人的大部分情感，但也存在一些不可避免的问题，如：面部遮挡、表情微妙以及姿态改变等，因此，基于面部表情的单模态情感识别方法并不足以准确的识别出情感状态。现实生活中，人们往往也是综合语音、面部表情、肢体动作等多种信息来判断一个人的情感状态，利用信息之间的互补性，从而准确识别出情感状态[6]。

在过去的十几年里，关于课堂中学习者的情感

**基金项目**：国家重点研发计划资助项目(2019YFB1802700)；上海工程技术大学研究生创新计划资助项目(19KY0232)。

**作者简介**：苏 超(1993-)，男，硕士研究生，主要研究方向：图像处理、智能教育、机器视觉等；王国中(1962-)，男，博士，教授，博士生导师，主要研究方向：数字音视频信息处理、智能信息处理、智慧教育等。

**通讯作者**：王国中 Email：wanggz@sues.edu.cn

**收稿日期**：2020-08-12

识别研究逐渐多了起来，如：文献[6]中提出一种基于遗传算法的多模态情感特征融合方法，利用遗传算法对单个模态的情感特征进行选择、交叉以及重组；文献[7]中提出一种基于皮肤电信号与文本信息的双模态情感识别系统；文献[8]中提出了基于双边稀疏偏最小二乘法的表情和姿态双模态情感识别方法。于此同时，国外学者们也进行了相关研究，如：Ray A，Chakrabarti A[9]指出，情绪在人的认知过程中起着重要的作用，因此提出一种新的情感计算模块，采用生物、物理（心率、皮肤电和血容量压）和面部表情方法，用来提取学习者的情感状态；Li C，Bao Z，Li L[10]等人提出了基于生理信号的情感识别方法。通过将原始生理信号转化为光谱图像，利用双向长短期记忆循环神经网络（LSTM-RNNS）学习特征，最后利用深度神经网络（DNN）进行预测。

经研究发现，上述诸多方法本质上并不完全适合于中国课堂上学习者的情感识别，主要有以下原因：

（1）‖数据集。目前关于情感识别研究的数据集都是国外的一些大学或者研究机构采集的，一是不符合课堂场景，其次受地域文化以及肤色人种的影响，国外采集的那些数据集表情与国内的人脸表情相差很大。

（2）情感分类。在情感识别领域，关于情感的分类有很多种，其中最基本的是 Ekman 等[11-12]提出的 6 种基本情感：高兴、愤怒、厌烦、恐惧、悲伤以及惊讶。研究发现，在学习过程中这 6 种基本情感并非全部起到关键作用[12]。因此，针对课堂上学习者的情感分类，需要定义一种符合课堂场景的情感类别。

因此，本文提出一种基于表情和姿态的双模态学习者的情感识别模型，以解决课堂上学习者情感识别存在识别准确率低、特征提取困难等问题。研究中将课堂学习者的面部表情定义为专注、厌烦、困惑、疲惫和走神 5 种表情；将课堂中学习者的上身姿态定义为抬头、低头、趴下和左顾右盼 4 种姿态。通过最终的决策融合，定义课堂上学习者的学习状态主要包括认真听讲、不感兴趣、疑惑/没听懂、犯困/疲劳、开小差、睡觉、交头接耳和不确定等 8 种状态。

## 1　模型设计

### 1.1　方法概述

基于表情和姿态的双模态学习者情感识别模型的整体流程如图 1 所示。主要包含 2 个模块，分别是情感识别模块和决策融合模块。首先，利用自主采集的数据集作为训练集和测试集，通过改进 Tiny_YOLOv3 目标检测算法得到 ER_Tiny_YOLOv3，从而进行面部表情和姿态识别。然后，针对课堂教学环境，将表情识别结果和姿态识别结果在决策层面进行融合，生成课堂评价，判断出学习者当前的学习状态。

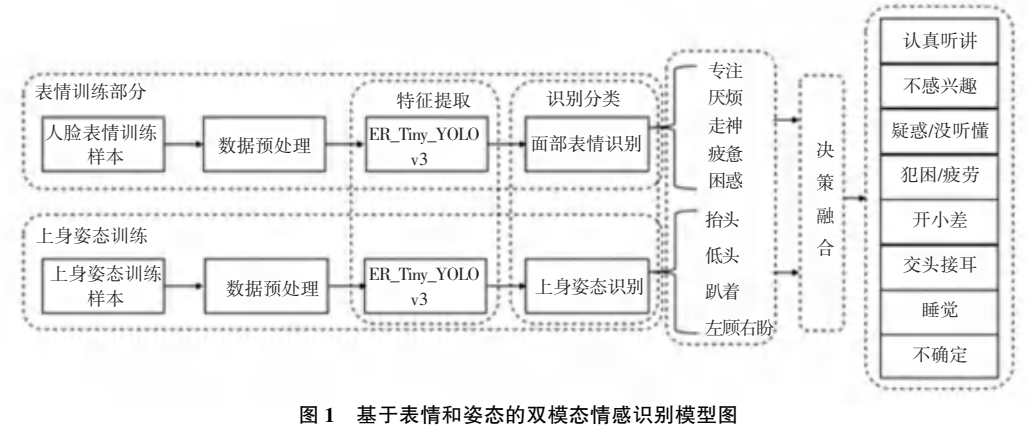



**图 1　基于表情和姿态的双模态情感识别模型图**

**Fig. 1　Diagram of a dual-modal emotion recognition model based on expressions and gestures**

### 1.2　Tiny_YOLOv3

Tiny_YOLOv3 是目标检测算法 YOLOv3 的一种轻量化版本，采用 Tiny Darknet 网络结构，如图 2 所示。

Tiny_YOLOv3 的损失函数主要包括 3 个部分，分别是目标位置损失、目标置信度损失和目标分类损失。Tiny_YOLOv3 的损失函数如式(1)所示：

$$loss = lbox + lobj + lcls. \tag{1}$$

其中：$lbox$ 表示目标位置损失；$lobj$ 表示目标置信度损失；$lcls$ 表示目标分类损失。如式(2)所示：

$$\begin{cases} lbox = \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^{B} l_{i,j}^{obj} (2 - w_i \times h_i) [(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 + (w_i - \hat{w}_i)^2 + (h_i - \hat{h}_i)^2], \\ lobj = \lambda_{noobj} \sum_{i=0}^{S^2} \sum_{j=0}^{B} l_{i,j}^{noobj} (c_i - \hat{c}_i)^2 + \lambda_{obj} \sum_{i=0}^{S^2} \sum_{j=0}^{B} l_{i,j}^{obj} (c_i - \hat{c}_i)^2, \\ lcls = \lambda_{class} \sum_{i=0}^{S^2} \sum_{j=0}^{B} l_{i,j}^{obj} \sum_{c \in classes} p_i(c) \log(\hat{p}_i(c)). \end{cases} \tag{2}$$

其中：$S^2$ 表示 13 × 13，26 × 26；$B$ 表示 $box$，$l_{i,j}^{obj}$ 表示如果在 $i,j$ 处的 $box$ 有目标，其值为 1；否则为 0；而 $l_{i,j}^{noobj}$ 反之。$\lambda_{coord}$ 表示 $lbox$ 权重；$x_i,y_i,w_i,h_i$ 为真实框的中心位置和长宽值；$x_i^{'},y_i^{'},w_i^{'},h_i^{'}$ 表示预测框的中心位置和长宽值；$(2-w_i \times h_i)$ 表示根据真实框的大小对 $lbox$ 权重进行修正；$\lambda_{nobj}$ 表示 $lobj$ 权重；$\lambda_{class}$ 表示 $lcls$ 权重；$c_i$ 表示真实框置信；$c_i^{'}$ 表示预测框置信。

Tiny_YOLOv3 的损失函数首先计算预测框和真实框的交并比（Intersection over Union，IoU），示意如图 3 所示。将 IoU 最大的预测框与真实框相匹配，通过匹配的预测框所预测的结果与真实框相比较，得出目标位置损失、目标置信度损失以及目标分类损失。IoU 表达如式（3）所示：

$$IoU = \frac{I}{U} = \frac{|A \cap B|}{|A \cup B|}. \tag{3}$$

其中：$I$ 表示真实框与预测框的交集；$U$ 表示真实框与预测框的并集。

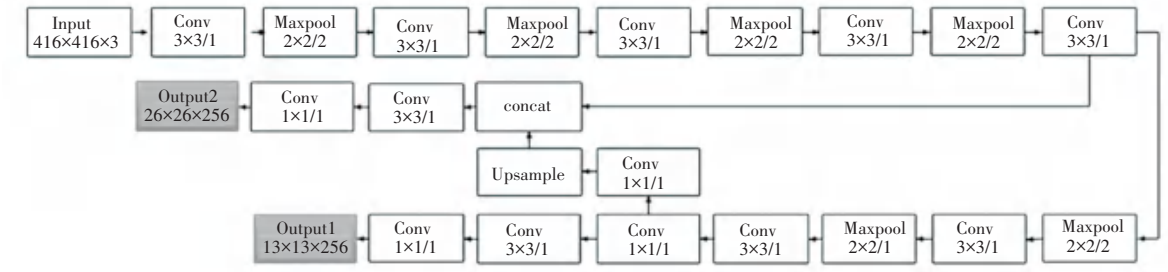



**图 2 Tiny_YOLOv3 网络结构图**

**Fig. 2 Tiny_YOLOv3 network structure diagram**

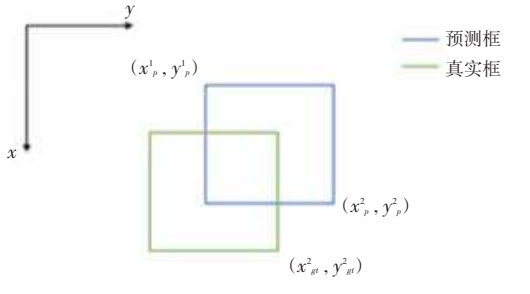



**图 3 IoU 示意图**

**Fig. 3 Schematic diagram of IoU**

## 2 模型改进

虽然 Tiny_YOLOv3 作为轻量化模型，很适用于实际工程项目中，但是检测效果并不是很好。原因在于 Tiny_YOLOv3 的 backbone 是浅层网络 Darknet19，且 Tiny_YOLOv3 只融合了 13x13、26x26 两个尺度上的检测结果。然而，正是由于 Tiny_YOLOv3 的网络结构浅，因而时效性比较好。为了提升课堂中学习者情感识别的准确性，本文从卷积结构、锚框聚类以及损失函数 3 方面对 Tiny_YOLOv3 进行了改进。

### 2.1 卷积结构改进

注意力机制（Attention Mechanism）是一种聚焦于局部信息的机制，目前已广泛应用于计算机视觉领域。注意力机制可以分为：通道注意力机制、空间注意力机制以及混合域注意力机制。SEnet（Squeeze-and-Excitation Networks）[13] 是典型的通道注意力机制，其通过建模各个特征通道的重要程度，针对不同的任务增强或者抑制不同的通道，从而提升精度。SEBlock 结构如图 4 所示。

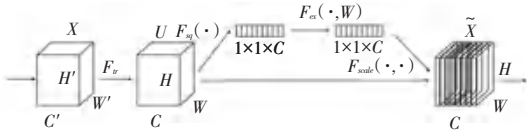



**图 4 SEBlock 结构图**

**Fig. 4 SEBlock structure diagram**

在图 4 中，$X \in R^{H' * W' * C'}$ 表示网络的输入，$F_{tr}$ 表示一系列卷积操作的集合，$V=[v_1,v_2,\cdots,v_c]$ 表示卷积操作，$U \in R^{H * W * C}$ 表示经过一系列卷积操作的输出，$U=[u_1,u_2,\cdots,u_c]$ 其表达如式（4）所示：

$$U = V * X. \tag{4}$$

$F_{sq}(\cdot)$ 操作是将 $U$ 的输出压缩成 $Z \in R^{1*1*C}$。传统的卷积操作大多集中于局部信息，无法提取整体信息。因此，通过 $F_{sq}(\cdot)$ 操作来实现，如式（5）所示：

$$z_c = F_{sq}(u_c) = \frac{1}{H * W}\sum_{i=1}^{H}\sum_{j=1}^{W} u_c(i,j). \tag{5}$$

$F_{ex}(\cdot,W)$ 主要利用非线性的 sigmoid 激活函数，保证非线性的前提下进行通道选择，如式（6）所示：

$$s = F_{ex}(z,W) = \sigma(g(z,w))\sigma(W_2\sigma(W_1 z)). \tag{6}$$

最后，通过 $F_{scale}(\cdot,\cdot)$ 操作将学习到的通道权重应用到原有的特征上，如式（7）所示：

$$\tilde{x} = F_{scale}(u_c, s_c) = u_c s_c, \tilde{X} = [\widetilde{x_1}, \widetilde{x_2}, \cdots \widetilde{x_C}]. \quad (7)$$

SEBlock 可以作为一种子模块插入到不同的卷积结构中，本文通过加入 SEBlock 对 Tiny_YOLOv3 网络结构进行改进，得到一种适用于学习者情感识别的模型 ER_Tiny_YOLOv3，网络结构如图 5 所示。

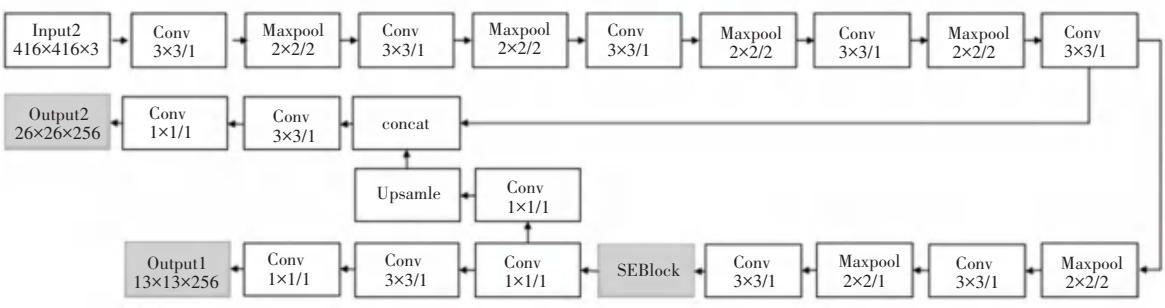



**图 5　ER_Tiny_YOLOv3 网络结构图**

**Fig. 5　ER_Tiny_YOLOv3 network structure diagram**

## 2.2　损失函数改进

在目标检测中，评价一个目标物体是否正确的被检测出，IoU 是一个重要的度量标准[14]。但 IoU 有其不可避免的缺点，如果预测框和真实框二者没有相交，则 IoU 的结果为 0，此时便无法进行学习训练。

针对上述问题，Hamid 等人提出了广义 IoU 的概念(Generalized Intersection over Union，GIoU)[15]，解决当前的问题，如式(8)所示：

$$GIoU = IoU - \frac{|A_c - U|}{|A_c|}. \quad (8)$$

其中：$A_c$ 表示预测框和真实框最小闭包区域面积，$U$ 表示预测框和真实框的并集。根据图 3，$A_c$ 的表达如式(9)所示：

$$A_c = [\max(x_{gt}^2, x_p^2) - \max(x_{gt}^1, x_p^1)] \times [\max(y_{gt}^2, y_p^2) - \max(y_{gt}^1, y_p^1)]. \quad (9)$$

本文在添加注意力机制 SEBlock 改变网络结构的基础上，将 GIoU loss 作为损失函数的一部分，用来改进目标位置损失 $lbox$。改进后的目标位置损失 $lbox'$ 为：

$$lbox' = \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^{B} l_{i,j}^{obj}(1 - GIoU). \quad (10)$$

则改进后的 ER_Tiny_YOLOv3 损失函数为：

$$loss = lbox' + lobj + lcls. \quad (11)$$

## 2.3　k-means 锚框聚类

Tiny_YOLOv3 和 YOLO v3 一样，引入了锚框(anchor)的概念，通过 K-means 聚类算法[16]得到 anchor 的数量和大小。但 Tiny_YOLOv3 模型是通过在 VOC 或 COCO 数据集上聚类而得到 anchor 的数量和大小的，不适合课堂中学习者的情感识别。因此，本文采用 K-means 算法在自制的数据集上进行重新聚类，得到适合于学习者情感识别的 anchor。

k-means 算法采用距离作为相似性指标，其中 K 表示聚类的类别数，算法流程图如图 6 所示。

由图 6 可见，在重新聚类之前，需要先确定 K 值。而在先验知识缺乏的情况下，要想确定 K 值是非常困难的。通常确定 K 值的方法有二种：肘部法和轮廓系数法。本文采用肘部法用来确定 K 值，肘部法的核心指标是通过 SSE(sum of the squared errors)来描述，如式(12)所示：

$$SSE = \sum_{i=1}^{k} \sum_{p \in C_i} |p - m_i|. \quad (12)$$

其中：$C_i$ 表示第 $i$ 个簇；$p$ 表示 $C_i$ 中的样本点；$m_i$ 是 $C_i$ 的质心，即所有样本的均值；$SSE$ 表示所有样本的聚类误差，代表了聚类效果的好坏。实验结果如图 7 所示。

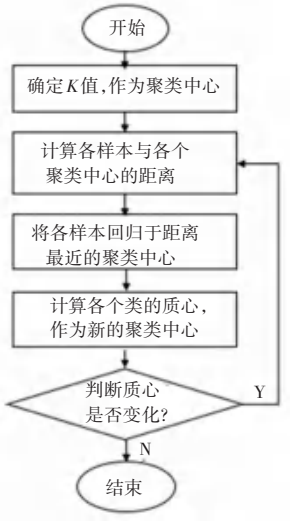



**图 6　K-means 算法聚类流程图**

**Fig. 6　K-means algorithm clustering flowchart**

从图 7 可以看出,在 $K$ 值为 6 时,曲率最高。因此,选取 $K = 6$。

$K$ 值确定后,利用 K-means 算法对自制的数据集进行重新聚类,实验结果如图 8 所示。最终聚类得到的 anchor 值为:10,18、12,23、13,28、16,35、20,45、30,63,用来替换 Tiny_YOLOv3 的 anchor 值:10,14、23,27、37,58、81,82、135,169、344,319。

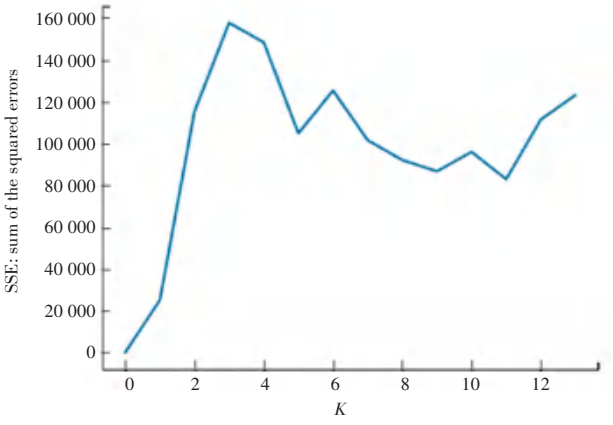



**图 7 利用肘部法确定 K 值实验图**

**Fig. 7 The experimental diagram of using the elbow method to determine the K value**

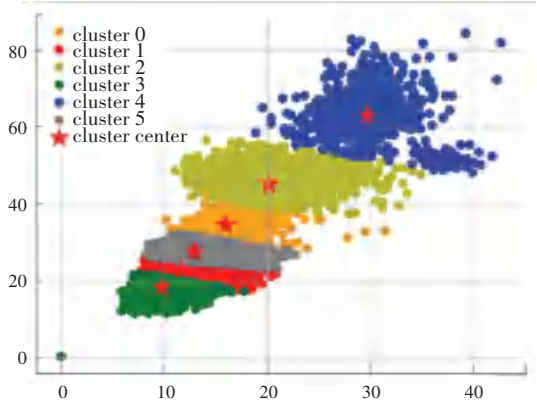
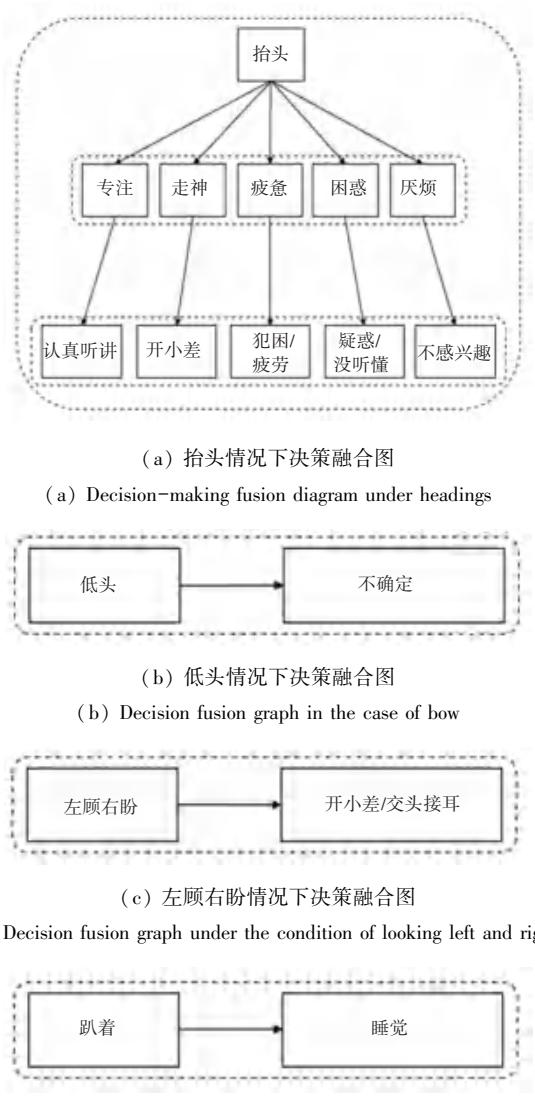



**图 8 K-means 算法重新聚类结果图**

**Fig. 8 K-means algorithm re-clustering result graph**

## 3 多模态融合方法

本文针对课堂中学习者的情感识别,采用决策层融合的方法。定义课堂上学习者的表情有:“专注”、“走神”、“疲惫”、“困惑”和“厌烦”5 种表情,而上身姿态有:“抬头”、“低头”、“左顾右盼”和“趴下”4 种行为。在实际场景中,只有当学习者处于“抬头”状态下才能完整观察到学习者的面部表情。因此,本文只针对“抬头”这种情况进行最后的决策融合。而其它 3 种行为对于课堂学习者来说,很容易判别出学习状态。4 种上身姿态对应的决策融合图分别对应图 9 中(a)、(b)、(c)、(d)图。




(a) 抬头情况下决策融合图

(a) Decision-making fusion diagram under headings

(b) 低头情况下决策融合图

(b) Decision fusion graph in the case of bow

(c) 左顾右盼情况下决策融合图

(c) Decision fusion graph under the condition of looking left and right

(d) 趴着情况下决策融合图

(d) Decision-making fusion graph under the condition of lying on the stomach

**图 9 4 种不同上身姿态对应的决策融合图**

**Fig.9 Decision fusion diagram corresponding to 4 different upper body postures**

由图 9 得知,当学习者处于抬头情况下,此时才能比较完整的检测出学习者的面部表情,从而进行面部表情识别,进而结合上身姿态和面部表情判定此时学习者的情感状态。而当学习者处于低头时,由于检测不到完整的面部表情,所以决策融合归结为不确定状态。同理,针对趴下和左顾右盼的上身姿态,在决策融合时不必考虑面部表情,直接归结为睡觉/犯困和开小差/交头接耳状态。

## 4 实验结果及分析

实验环境为:

软件环境:Windows 10 下的 Pytorch 平台;

硬件环境:处理器是 Intel(R) Xeon(R) W-

2123 CPU@3.60GHz；

运行内存：16.0GB；

### 4.1　数据集

当前，人脸表情数据库的种类有很多，最常用的主要有日本女性人脸表情数据库（JAFFE）、卡内基梅隆大学的 CK（Cohn-Kanade）人脸表情数据库及其扩展数据库 CK+人脸表情数据库等[17]。而目前唯一公开的表情和姿态双模态情感数据库只有 FABO 数据库[18]。但是，由于 FABO 数据库 ground truth label 很繁琐，且每个人的样本和情感类别数目不一致，同时外国人的脸部表情特征和中国人的人脸表情特征明显不同，且数据库的采集环境并不是针对课堂环境，因此训练出来的模型并不适合课堂学习者的情感识别。因此，本文以安徽省池州市东至三中高一某班的全体学生为采集对象，自主构建数据集。

### 4.2　网络训练和超参数设置

ER_Tiny_YOLOv3 在训练时采用 Tiny_YOLOv3 提供的权重参数作为网络训练的初始化参数，通过在自主构建的数据集上进行训练，并进行相应的网络参数微调，使得检测效果达到最优。实验参数见表 1。

**表 1　实验参数说明**

**Tab. 1　Description of experimental parameters**

| 参数名称 | 参数值 |
|---|---|
| epoch（轮回次数） | 200 |
| batch size（批量大小） | 32 |
| learning rate（学习率） | 0.001 |
| momentum（动量参数） | 0.9 |
| decay（权重衰减） | 0.000 5 |
| lr_factor（学习率衰减因子） | 0.01 |
| reduction（SEBlock 中的缩放参数） | 16 |

### 4.3　评价指标

目标检测模型一般采用准确率（Precision）、召回率（Recall）以及均值平均精度（Mean Average Precision，mAP）等指标来评价模型的效果。其中，准确率表示所有检测出的目标中，正确检测出的目标所占的比例；召回率表示所有待检测目标中正确检测出的目标所占的比例。

对于课堂中学习者的情感识别，如果对于一类表情或行为能正确的被检测出，则为真正类（True Positive，TP），相反，如果对于一个既不是表情又不是行为的位置检测为某类面部表情或行为，则为假正类（False Positive，FP）。假设学习者的情感识别中表情和行为的总数为 N，则：

$$\begin{cases} \mathrm{Re}call = \dfrac{TP}{N}, \\ \mathrm{Pr}ecision = \dfrac{TP}{TP + FP}. \end{cases} \tag{13}$$

但是，准确率和召回率是相互影响的，一般情况下准确率高、召回率就低，而准确率低，召回率就高。因此，需要在准确率和召回率之间进行权衡。一种方式是画出准确率-召回率曲线，计算 AP 值，另一种方式是计算 $F_\beta$ 分数。如式（14）所示：

$$F_\beta = (1 + \beta^2) \cdot \frac{\mathrm{Pr}ecision \cdot \mathrm{Re}call}{\beta^2 \cdot (\mathrm{Pr}ecision + \mathrm{Re}call)}. \tag{14}$$

其中，当 $\beta = 1$ 时，称为 $F_1$ 分数，是最常用的指标之一。

### 4.4　模型对比

本文通过将改进后的模型 ER_Tiny_YOLOv3 与 Tiny_YOLOv3 以及 YOLO 系列最新算法 YOLOv4 进行对比，以此来说明此模型的有效性，实验结果见表 2。

**表 2　各模型实验结果对比**

**Tab. 2　Comparison of test results of various models**

| 网络模型 | mAP@0.5 | Recall | Precision | F1 | Time/s | Params |
|---|---|---|---|---|---|---|
| YOLOv4 | 0.60 | 0.62 | 0.48 | 0.541 | 0.1617 | 6.39808e+07 |
| Tiny_YOLOv3 | 0.62 | 0.78 | 0.40 | 0.529 | 0.0079 | 8.68836e+06 |
| ER_Tiny_YOLOv3（本文） | 0.79 | 0.76 | 0.75 | 0.755 | 0.0088 | 8.81943e+06 |

从表 2 的结果可以得知：改进后的模型 ER_Tiny_YOLOv3 相比于原模型 Tiny_YOLOv3，mAP@0.5 提升了 17%，Precision 提升了 35%，而 F1 分数提升了 22.6%。相比于 YOLOv4，mAP@0.5 提升了 19%，Precision 提升了 27%，F1 分数提升了 21.4%。但是由于 ER_Tiny_YOLOv3 在 Tiny_YOLOv3 的基础上加入了注意力机制 SEBlock，所以参数比 Tiny_YOLOv3 多了 1/10，检测时间比 Tiny_YOLOv3 慢了约 1/100，二者几乎没有区别。但相比于深层网络 YOLOv4，检测时间大幅缩短，只有 YOLOv4 检测时

间的 1/20。同时,ER_Tiny_YOLOv3 的参数只有 YOLOv4 的 1/10,可见,网络参数大幅度减少。

实验结果表明,改进后的 ER_Tiny_YOLOv3 模型是一种兼顾速度和精度的模型。YOLOv4、Tiny_YOLOv3 和改进后的模型 ER_Tiny_YOLOv3 的检测结果对比如图 10 所示。

(a) 原图

(b) YOLOv4 检测图

(c) Tiny_YOLOv3 检测图

(d) ER_Tiny_YOLOv3 检测图

**图 10 各模型检测结果对比图**

**Fig. 10 Comparison of test results of various models**

从图 10 可以看出,对于课堂上学习者的情感识别,检测效果最好的是 ER_Tiny_YOLOv3 模型。而作为 YOLO 系列最新的检测算法 YOLOv4,检测效果反而一般。虽然 Tiny_YOLOv3 检测速度最快,但是检测效果明显不如 ER_Tiny_YOLOv3,漏检的情况比较多。ER_Tiny_YOLOv3 在检测速度上与 Tiny_YOLOv3 相差无几的情况下,检测效果明显好于 Tiny_YOLOv3,且识别效果比 YOLOv4 效果更好。可见,本文提出的学习者情感识别模型是一种兼顾速度和精度的模型,适用于课堂场景下学习者的情感识别。

## 5 结束语

针对当前智能化教育环境中的“情感缺失”问题,本文提出了一个快速、准确、轻量的学习者情感识别模型。通过对 Tiny_YOLOv3 的卷积结构、损失函数以及锚框值进行改进,经过训练得到一个适合于课堂中学习者情感识别的模型 ER_Tiny_YOLOv3。同时,针对最终的课堂评价,采用决策层融合方法用来判断学习者的学习状态。实验结果表明,相比于 Tiny_YOLOv3 和 YOLOv4,识别效果更

好。当然，还有很多问题需要进一步研究，比如：决策层融合的方法不具有完整的代表性。如学习者在低头的情况下，也有可能在思考问题，而本文章将其归结于不确定状态。同理，针对趴下和左顾右盼的情况，也有同样的问题。因此，下一步将重点研究更具代表性的融合方法。

## 参考文献


[1] Imbernón Cuadrado L E, Manjarrés Riesco Á, De La Paz López F. ARTIE: An integrated environment for the development of affective robot tutors[J]. Frontiers in computational neuroscience, 2016, 10: 77.

[2] 孟昭兰. 情绪心理学[M]. 北京:北京大学出版社,2005.

[3] 徐振国,张冠文,孟祥增,等. 基于深度学习的学习者情感识别与应用[J]. 电化教育研究,2019 (2):87-93.

[4] 余梓彤,李晓白,赵国英. 情感识别与教育[J]. 人工智能,2019 (3):29-36.

[5] 刘永娜. 学习环境中基于面部表情的情感识别[D]. 北京:北京师范大学,2015.

[6] 卢官明,程晓,李霞,等. 基于遗传算法的多模态情感特征融合方法[J]. 南京邮电大学学报,2019,39(5):41-47.

[7] 张力行,叶宁,黄海平,等. 基于皮肤电信号与文本信息的双模态情感识别系统[J]. 计算机系统应用, 2018, 27(11):103-108.

[8] 闫静杰,郑文明,辛明海,等. 表情和姿态的双模态情感识别[J]. 中国图象图形学报,2013,18(9):1101-1106.

[9] RAY A, CHAKRABARTI A. Design and implementation of technology enabled affective learning using fusion of bio-physical and facial expression[J]. Journal of Educational Technology & Society, 2016, 19(4): 112-125.

[10] LI C, BAO Z, LI L, et al. Exploring temporal representations by leveraging attention-based bidirectional lstm-rnns for multi-modal emotion recognition[J]. Information Processing & Management, 2020, 57(3): 102-185.

[11] EKMAN P, FRIESEN W V. Unmasking the face Englewood Cliffs[J]. Spectrum-Prentice Hall, New Jersey, 1975.

[12] 孙波,刘永娜,陈玖冰,等. 智慧学习环境中基于面部表情的情感分析[J]. 现代远程教育研究,2015 (2):96-103.

[13] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018, 1: 7132-7141.

[14] 何智成,王振兴. 基于改进 YOLO v2 的白车身焊点检测方法[J]. 计算机工程,2020.

[15] REZATOFIGHI H, TSOI N, GWAK J Y, et al. Generalized intersection over union: A metric and a loss for bounding box regression[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019, 1: 658-666.

[16] Hartigan J A, Wong M A. Algorithm AS 136: A K-Means Clustering Algorithm[J]. Journal of the Royal Statistical Society, 1979, 28(1):100-108.

[17] 刘永娜,孙波,陈玖冰等. BNU 学习情感数据库的设计与实现[J].现代教育技术, 2015, 25(10):99-105.

[18] GUNES H, PICCARDI M. A bimodal face and body gesture database for automatic analysis of human nonverbal affective behavior[C]// Pattern Recognition, 2006. ICPR 2006. 18th International Conference on IEEE, 2006, 1: 1148-1153.


(上接第 7 页)


[19] KIM J, LEE J K, LEE K M. Deeply-recursive convolutional network for image super-resolution[C]//IEEE Conference on Computer Vision and Pattern Recognition, 2016:1637-1645.

[20] AHN N, KANG B, SOHN K. Fast, accurate, and lightweight super-resolution with Cascading residual network[C]//European Conference on Computer Vision, 2018:256-272.

[21] ZHANG Z, WANG X, JUNG C. DCSR: Dilated convolutions for single image super-resolution[J]. IEEE Trans. Image Process. 2019,28 (4) :1625-1635.

[22] WANG C, LI Z, SHI J. Lightweight image super-resolution with adaptive weighted learning network, 2019, arXiv preprint arXiv: 1904.02358.

[23] QIN J, XIE Z, SHI Y, et al. Difficulty-aware image super resolution via deep adaptive dual-network[C]//2019 IEEE International Conference on Multimedia and Expo (ICME), 2019:586-591.

[24] ZHAO X, HU X, LIAO Y, et al. Accurate MR image super-resolution via lightweight lateral inhibition network[J]. Computer Vision and Image Understanding, 2020, 201:103075.

[25] Manjón, José V, Coupé, Pierrick, Buades A, et al. MRI Superresolution Using Self-Similarity and Image Priors[J]. International Journal of Biomedical Imaging, 2010, (2010-10-11), 2010, 2010(1687-4188).

[26] HYUN CM, KIM H P, LEE S M, et al., 2017. Deep learning for undersampled MRI reconstruction. ArXiv preprint arXiv: 1709.02576.

[27] SHI J, LI Z, YING S, et al. MR image super-resolution via wide residual networks with fixed skip connection[J]. IEEE J. Biomed. Health Inform. 2019,23 (3): 1129-1140.