



(12) 发明专利申请

(10) 申请公布号 CN 114663910 A

(43) 申请公布日 2022.06.24

(21) 申请号 202210027041.4

G06N 3/04 (2006.01)

(22) 申请日 2022.01.11

G06N 3/08 (2006.01)

(71) 申请人 重庆工程学院

G06Q 50/20 (2012.01)

地址 400056 重庆市巴南区南泉街道办事处
白鹤林16号

(72) 发明人 朱世宇 孙令翠 杨红艳 何桢
田菊艳 余玉清 卢政旭 冉程好

(51) Int. Cl.

G06V 40/10 (2022.01)

G06V 40/16 (2022.01)

G06V 10/44 (2022.01)

G06V 40/19 (2022.01)

G06V 10/80 (2022.01)

G06V 10/82 (2022.01)

G06K 9/62 (2022.01)

G06F 16/29 (2019.01)

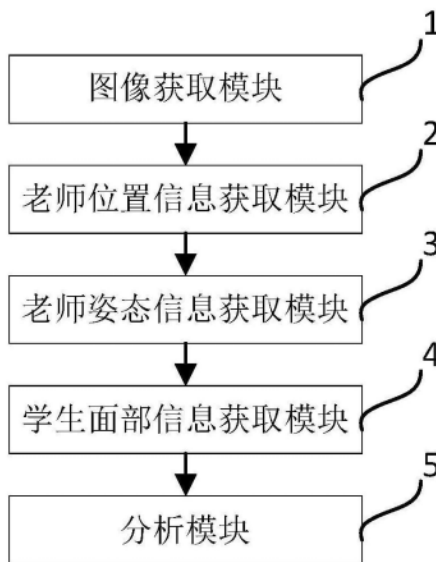
权利要求书1页 说明书6页 附图4页

(54) 发明名称

基于多模态学习状态分析系统

(57) 摘要

本发明涉及数字数据处理领域,具体涉及一种基于多模态学习状态分析系统,包括图像获取模块,用于收集教室内部的图像数据并进行预处理得到处理图像;老师位置信息获取模块,用于基于处理图像和Faster R-CNN目标检测模型获得老师的位置信息;老师姿态信息获取模块,用于基于处理图像和Faster R-CNN目标检测模型获得老师的姿态信息;学生面部信息获取模块,用于基于处理图像和ERT人脸特征点检测方法获取学生面部信息;分析模块,用于基于位置信息、姿态信息和学生面部信息分析学习状态。从而可以更好地对学生的上课情况进行监管,以提高学生的学习效率。



1. 一种基于多模态学习状态分析系统,其特征在于,
包括图像获取模块、老师位置信息获取模块、老师姿态信息获取模块、学生面部信息获取模块和分析模块,
所述图像获取模块,用于收集教室内部的图像数据并进行预处理得到处理图像;
所述老师位置信息获取模块,用于基于处理图像和Faster R-CNN目标检测模型获得老师的位置信息;
所述老师姿态信息获取模块,用于基于处理图像和Faster R-CNN目标检测模型获得老师的姿态信息;
所述学生面部信息获取模块,用于基于处理图像和ERT人脸特征点检测方法获取学生面部信息;
所述分析模块,用于基于位置信息、姿态信息和学生面部信息分析学习状态。
2. 如权利要求1所述的一种基于多模态学习状态分析系统,其特征在于,
所述图像获取模块包括获取单元和处理单元,所述获取单元用于收集摄像头的图像数据;
所述处理单元,用于将图像数据的尺寸标准化,并进行归一化处理,得到处理图像。
3. 如权利要求2所述的一种基于多模态学习状态分析系统,其特征在于,所述收集摄像头的图像数据的具体方式是,每隔5秒为一帧获取图像。
4. 如权利要求1所述的一种基于多模态学习状态分析系统,其特征在于,
所述老师位置信息获取模块包括特征图提取单元、候选单元、区域特征图生成单元和老师位置获取单元,
所述特征图提取单元,用于基于处理图像提取原始特征图;
所述候选单元,用于将原始特征图输入到候选框提取网络,生成区域候选框;
所述区域特征图生成单元,用于将区域候选框映射到所述原始特征图中,并池化为区域特征图;
所述老师位置获取单元,用于将区域特征图输入Faster R-CNN目标检测模型中获取老师位置信息。
5. 如权利要求4所述的一种基于多模态学习状态分析系统,其特征在于,
所述老师姿态信息获取模块包括关键点获取单元和归一化单元,所述关键点获取单元,用于基于Faster R-CNN目标检测模型,获取人体姿态关键点信息;
所述归一化单元,用于基于人体姿态关键点信息进行姿态归一化模块处理。
6. 如权利要求1所述的一种基于多模态学习状态分析系统,其特征在于,
所述基于位置信息、姿态信息和学生面部信息分析学习状态的具体步骤是:
构建并训练基于全连接层的多模态特征融合网络结构;
将位置信息、姿态信息和学生面部信息映射到特征融合空间进行学习状态分析。
7. 如权利要求1所述的一种基于多模态学习状态分析系统,其特征在于,
所述基于位置信息、姿态信息和学生面部信息分析学习状态的具体步骤是:
采用加权融合方法融合位置信息、姿态信息和学生面部信息得到加权融合特征;
将加权融合特征输入全连接层;
获得加权融合特征的分类概率分布。

基于多模态学习状态分析系统

技术领域

[0001] 本发明涉及数字数据处理领域,尤其涉及一种基于多模态学习状态分析系统。

背景技术

[0002] 在传统课堂教育中,老师通过对学生的面部表情和头部姿态,判断学生对学习状态,但由于老师的精力有限,无法及时观察到每位学生的课堂学习状态情况,进而无法根据每位学生的学习情况调整教学策略。

[0003] 智慧教学是当前我国教育信息化研究的热词,有学者将其称之为教育信息化发展的新形态、新境界、新阶段,将智慧教学研究提到了相当高的高度。现如今关于学生课堂的学情分析的检测方面,可将检测分为基于人脸识别的方法、基于微表情的识别方法和基于脑电波的检测方法。

[0004] 采用上述方法由于只有一个变量,使得检测精度较低。

发明内容

[0005] 本发明的目的在于提供一种基于多模态学习状态分析系统,旨在结合学生的面部表情、老师的位置以及老师的上课语音的数据进行分析,给出可视化的分析结果。可以给出学生专注度的分析,有利于教师制定合适的教学计划,促进老师和学生的课堂交互。

[0006] 为实现上述目的,本发明提供了一种基于多模态学习状态分析系统,包括图像获取模块、老师位置信息获取模块、老师姿态信息获取模块、学生面部信息获取模块和分析模块,

[0007] 所述图像获取模块,用于收集教室内部的图像数据并进行预处理得到处理图像;

[0008] 所述老师位置信息获取模块,用于基于处理图像和Faster R-CNN目标检测模型获得老师的位置信息;

[0009] 所述老师姿态信息获取模块,用于基于处理图像和Faster R-CNN目标检测模型获得老师的姿态信息;

[0010] 所述学生面部信息获取模块,用于基于处理图像和ERT人脸特征点检测方法获取学生面部信息;

[0011] 所述分析模块,用于基于位置信息、姿态信息和学生面部信息分析学习状态。

[0012] 其中,所述图像获取模块包括获取单元和处理单元,所述获取单元用于收集摄像头的图像数据;

[0013] 所述处理单元,用于将图像数据的尺寸标准化,并进行归一化处理,得到处理图像。

[0014] 其中,所述收集摄像头的图像数据的具体方式是,每隔5秒为一帧获取图像。

[0015] 其中,所述老师位置信息获取模块包括特征图提取单元、候选单元、区域特征图生成单元和老师位置获取单元,

[0016] 所述特征图提取单元,用于基于处理图像提取原始特征图;

- [0017] 所述候选单元,用于将原始特征图输入到候选框提取网络,生成区域候选框;
- [0018] 所述区域特征图生成单元,用于将区域候选框映射到所述原始特征图中,并池化为区域特征图;
- [0019] 所述老师位置获取单元,用于将区域特征图输入Faster R-CNN目标检测模型中获取老师位置信息。
- [0020] 其中,所述老师姿态信息获取模块包括关键点获取单元和归一化单元,所述关键点获取单元,用于基于Faster R-CNN目标检测模型,获取人体姿态关键点信息;
- [0021] 所述归一化单元,用于基于人体姿态关键点信息进行姿态归一化模块处理。
- [0022] 其中,所述基于位置信息、姿态信息和学生面部信息分析学习状态的具体步骤是:
- [0023] 构建并训练基于全连接层的多模态特征融合网络结构;
- [0024] 将位置信息、姿态信息和学生面部信息映射到特征融合空间进行学习状态分析。
- [0025] 其中,所述基于位置信息、姿态信息和学生面部信息分析学习状态的具体步骤是:
- [0026] 采用加权融合方法融合位置信息、姿态信息和学生面部信息得到加权融合特征;
- [0027] 将加权融合特征输入全连接层;
- [0028] 获得加权融合特征的分类概率分布。
- [0029] 本发明的一种基于多模态学习状态分析系统,可以在教室的前后分别放置摄像头,然后通过所述图像获取模块获取图像数据,之后通过Faster R-CNN目标检测模型分别对图像进行处理以得到老师的位置信息和面部信息,然后通过ERT人脸特征点检测方法提取图像数据中的学生面部信息,之后可以综合对学生的进行学习状态进行评估,从而可以更好地对学生的上课情况进行监管,以提高学生的学习效率。

附图说明

- [0030] 为了更清楚地说明本发明实施例或现有技术中的技术方案,下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本发明的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他的附图。
- [0031] 图1是本发明的池化流程图。
- [0032] 图2是本发明的老师姿态信息计算流程图。
- [0033] 图3是本发明的多模态融合分析学生课堂学习状态图。
- [0034] 图4是本发明的通过三模态卷积神经网络的联合学习可以得到三种基于不同模态的三维模型特征到的流程图。
- [0035] 图5是本发明的一种基于多模态学习状态分析系统的结构图。
- [0036] 图6是本发明的图像获取模块的结构图。
- [0037] 图7是本发明的老师位置信息获取模块的结构图。
- [0038] 图8是本发明的老师姿态信息获取模块的结构图。
- [0039] 1-图像获取模块、2-老师位置信息获取模块、3-老师姿态信息获取模块、4-学生面部信息获取模块、5-分析模块、11-获取单元、12-处理单元、21-特征图提取单元、22-候选单元、23-区域特征图生成单元、24-老师位置获取单元、31-关键点获取单元、32-归一化单元。

具体实施方式

[0040] 下面详细描述本发明的实施例,所述实施例的示例在附图中示出,其中自始至终相同或类似的标号表示相同或类似的元件或具有相同或类似功能的元件。下面通过参考附图描述的实施例是示例性的,旨在用于解释本发明,而不能理解为对本发明的限制。

[0041] 实施例1

[0042] 请参阅图1~图8,本发明提供一种基于多模态学习状态分析系统:

[0043] 包括图像获取模块1、老师位置信息获取模块2、老师姿态信息获取模块3、学生面部信息获取模块4和分析模块5,

[0044] 所述图像获取模块1,用于收集教室内部的图像数据并进行预处理得到处理图像;

[0045] 所述老师位置信息获取模块2,用于基于处理图像和Faster R-CNN目标检测模型获得老师的位置信息;

[0046] 所述老师姿态信息获取模块3,用于基于处理图像和Faster R-CNN目标检测模型获得老师的姿态信息;

[0047] 所述学生面部信息获取模块4,用于基于处理图像和ERT人脸特征点检测方法获取学生面部信息;

[0048] 所述分析模块5,用于基于位置信息、姿态信息和学生面部信息分析学习状态。

[0049] 在本实施方式中,可以在教室的前后分别放置摄像头,然后通过所述图像获取模块1获取图像数据,之后通过Faster R-CNN目标检测模型分别对图像进行处理以得到老师的位置信息和面部信息,然后通过ERT人脸特征点检测方法提取图像数据中的学生面部信息,之后可以综合对学生的进行学习状态进行评估,从而可以更好地对学生的上课情况进行监管,以提高学生的学习效率。其中ERT人脸特征点检测方法的公式如下:

[0050] $\xi^{(t+1)} = \xi^{(t)} + r_t(I, \xi^{(t)})$

[0051] 其中:t表示级联序号, $r_t(\cdot)$ 表示当前级的回归器。回归器输入人脸图像I和上一级回归器更新后的面部特征点位置坐标,采用的特征可以是灰度值或者其他特征。当获得图像是,算法生成出事位置,改初始位置是预估面部特征点的位置。采用梯度提升算法来最小化误差,得到每一级联回归银子。

[0052] 进一步的,所述图像获取模块1包括获取单元11和处理单元12,所述获取单元11用于收集摄像头的图像数据;

[0053] 所述处理单元12,用于将图像数据的尺寸标准化,并进行归一化处理,得到处理图像。

[0054] 所述收集摄像头的图像数据的具体方式是,每隔5秒为一帧获取图像。使得采样频率更加合适,在满足实时性的同时也提高了图像的处理效率。

[0055] 进一步的,所述老师位置信息获取模块2包括特征图提取单元21、候选单元22、区域特征图生成单元23和老师位置获取单元24,

[0056] 所述特征图提取单元21,用于基于处理图像提取原始特征图;

[0057] 所述候选单元22,用于将原始特征图输入到候选框提取网络,生成区域候选框;

[0058] 所述区域特征图生成单元23,用于将区域候选框映射到所述原始特征图中,并池化为区域特征图;

[0059] 所述老师位置获取单元24,用于将区域特征图输入Faster R-CNN目标检测模型中

获取老师位置信息。

[0060] 在本实施方式中,提取原始特征图的具体方式是首先将任意大小的上课场景中的老师行为图像缩放置固定大小256*256,然后将256*256大小的图片输入CNN(卷积层)等基础网络,从而特取原始图片的特征图,后续的RPN层和全连接层可以共享该特征图。

[0061] 第二步:将特征图输入到RPN,生成区域候选框。在特征图上应用华东窗口进行目标区域判定和分类,从而生成区域候选框。RPN首先对共享卷积层输入的特征图做一个卷积操作已得到特征向量,然后将特征向量输入连个全连接层,即边界框回归层和边界框分类层。

[0062] 第三步:对每个区域特征图进行池化(POI Pooling)。输入特征图和候选区域,将候选区域映射到特征图中,并池化为统一大小的区域特征图,在将区域特征图输入全连接,得到分类向量和位置坐标

[0063] 第四步:最后将区域特征图输入Faster R-CNN中,根据RPN所提出的候选框,对候选框做进一步地位置精修好额类别回归。

[0064] 对于老师上课行为图片中的每个选定区域 i , f_i 定义为来自改区域的平均池化卷积特征,因此图像特征向量的为维数2048.全连接层将 f_i 转换成 h 维向量。

$$[0065] \quad v_i = W_v f_i + b_v$$

[0066] 因此,老师上课时行为图像的完整表示是一组嵌入向量。

$$[0067] \quad V = \{v_1, \dots, v_k\}, v_i \in R$$

[0068] 其中, v_i 编码一个显著区域 i , k 是区域个数。

[0069] 进一步的,所述老师姿态信息获取模块3包括关键点获取单元31和归一化单元32,所述关键点获取单元31,用于基于Faster R-CNN目标检测模型,获取人体姿态关键点信息;

[0070] 所述归一化单元32,用于基于人体姿态关键点信息进行姿态归一化模块处理。

[0071] 在本实施方式中,Faster R-CNN目标检测模型,获取人体姿态关键点信息,在进行姿态归一化模块处理。

[0072] 在姿态归一化时,对关键点坐标的估算,从而计算变换矩阵。

[0073] 流程如图2所示。图中,首先基于先前关键点检测网络得到的关键点提取出关键点的坐标。然后选取其中最大值点作为关键点的估计,有了关键点坐标。

[0074] 进一步的,所述基于位置信息、姿态信息和学生面部信息分析学习状态的具体步骤是:

[0075] 构建并训练基于全连接层的多模态特征融合网络结构;

[0076] 将位置信息、姿态信息和学生面部信息映射到特征融合空间进行学习状态分析。

[0077] 为了解决单一模态数据对学生课堂学习状态的评估,本文采用了多模态融合的方法,从3个模态分析学生课堂学习状态。总的流程图如图3所示。

[0078] 通过构建并训练基于全连接层的多模态特征融合网络结构,将多维度的多尺度特征映射到特征融合空间,这种特征融合方式的优势主要在于模型能够在训练阶段学习2个并行网络各自的特征参数,并自主完成协调反馈,实现了模型端到端的训练。

[0079] 本文选自Softmax函数将特征向量映射成概率序列,已保留更多特征的原始信息。Softmax计算输出类别 $y^{(i)}$ 的过程如公式所示。

$$[0080] \quad P(y^{(i)} = k) = \frac{\exp(\eta^k)}{\sum_{j=1}^k \exp(\eta^j)}, y \in [1, k]$$

[0081] 式中： η^i 为融合后的特征值； k 为类别数； P 表示 $y^{(i)}$ 表示类别 k 的概率值。

[0082] 第一阶段：关注每张老师图片中老师位置信息特征与学生眼睛偏移量，得到学生是否看老师

[0083] 第二阶段：将老师行为图像特征区域进行分析，此时应该看黑板还是老师；然后与学生眼睛偏移量进行比较，得出学生此时应该看那里。

[0084] 第三阶段：将学生的面部特征图与老师行为和位置进行分析，得出相似实施例2

[0085] 实施例2和实施例1的区别仅在于所述基于位置信息、姿态信息和学生面部信息分析学习状态的具体步骤是：

[0086] 采用加权融合方法融合位置信息、姿态信息和学生面部信息得到加权融合特征；

[0087] 将加权融合特征输入全连接层；

[0088] 获得加权融合特征的分类概率分布。

[0089] 通过三模态卷积神经网络的联合学习可以得到三种基于不同模态的三维模型特征。区别于传统的采用池化操作的特征融合方法，本文中基于统计方法，采用加权融合方法融合三个特征向量，方法的框架如图4所示。

[0090] 具体公式：

$$[0091] \quad f = \sum_{i=1}^3 \alpha_i \xi(f_{M_i})$$

$$[0092] \quad \sum_{i=1}^3 \alpha_i = 1$$

[0093] 其中： f 表示的是在不同模态表特征提取的特征向量； α_i 表示的是不同模态的权重，对加权融合特征的特征向量输入到全连接层（FC层），全连接层的维度依次为512、256、 C ， C 表示数据集类别的数量。最后通过一个softmax层获得三维模型分类的概率分布。将学习状态分为从容、一般、走神、困难

[0094] 本文通过相关性损失函数来确保训练过程中多个模态之间可以相互指导，提高网络训练的学习速度，并提高最终特征向量的鲁棒性。

[0095] 具体公式如下：

$$[0096] \quad L_C(M_i, M_j) = \|\xi(f_{M_i}) - \xi(f_{M_j})\|_2$$

[0097] 其中： 2 表示为两个不同特征向量的相关性。 f 表示的是在不同模态表特征提取的特征向量； M 的下标表示第1、2、3个模态的数据； $\xi = \text{sigmoid}(\log(\text{abs}()))$ 表示一个归一化激发函数。在训练过程中，相关性损失的值逐渐减小，表明在训练过程中不同模态特征相互指导，这会加快训练的收敛速度，获得更具有鲁棒性的特征向量。以模态 M_i 为例，基于这种相关性损失函数的设计，不同模态网络的最终损失函数如下：

$$[0098] \quad L_{M_i} = L_{d, M_i} + L_C(M_i, M_2) + L_C(M_i, M_3)$$

[0099] 其中： L_{d, M_i} 表示基于单模态的交叉熵损失； $L_C(M_i, M_2)$ 和 $L_C(M_i, M_3)$ 分别表示模态 M_i 与模态 M_2 和 M_3 的相关性损失。最后通过随机梯度下降的反向传播优化这三个单模态网络。

[0100] 以上所揭露的仅为本发明一种较佳实施例而已，当然不能以此来限定本发明之权

利范围,本领域普通技术人员可以理解实现上述实施例的全部或部分流程,并依本发明权利要求所作的等同变化,仍属于发明所涵盖的范围。

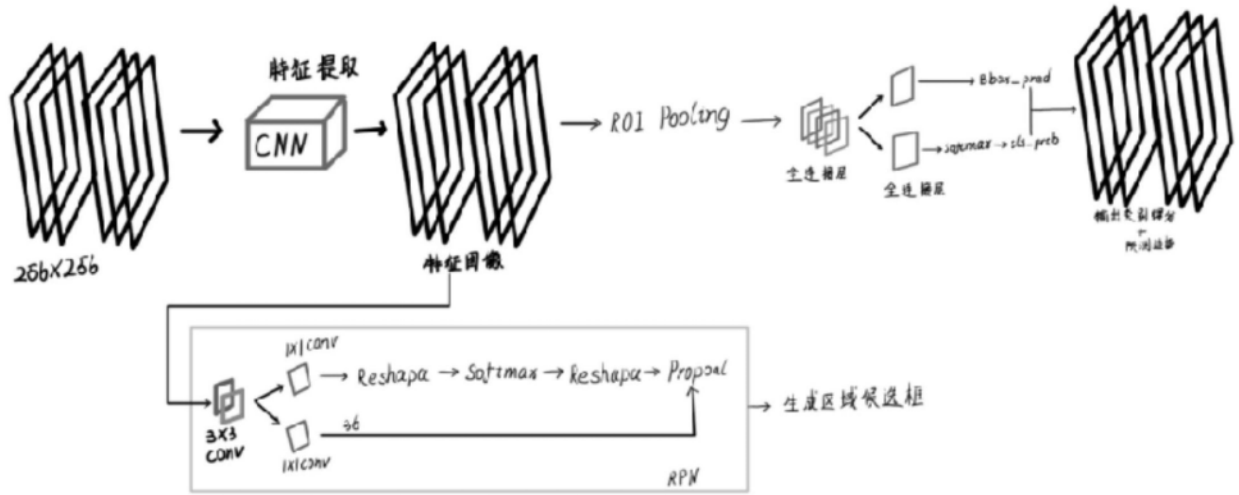


图1

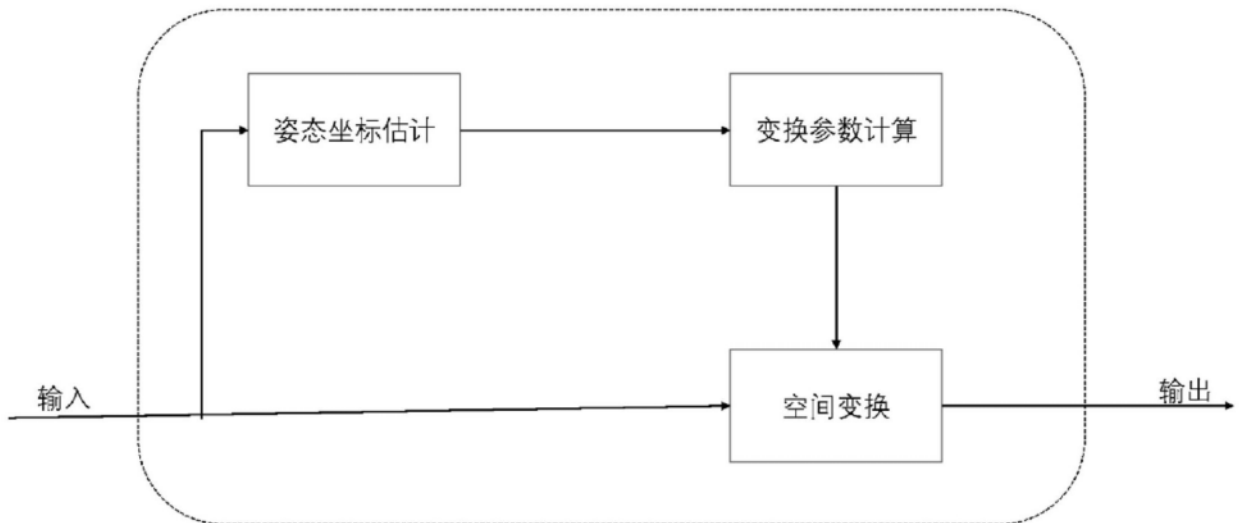


图2

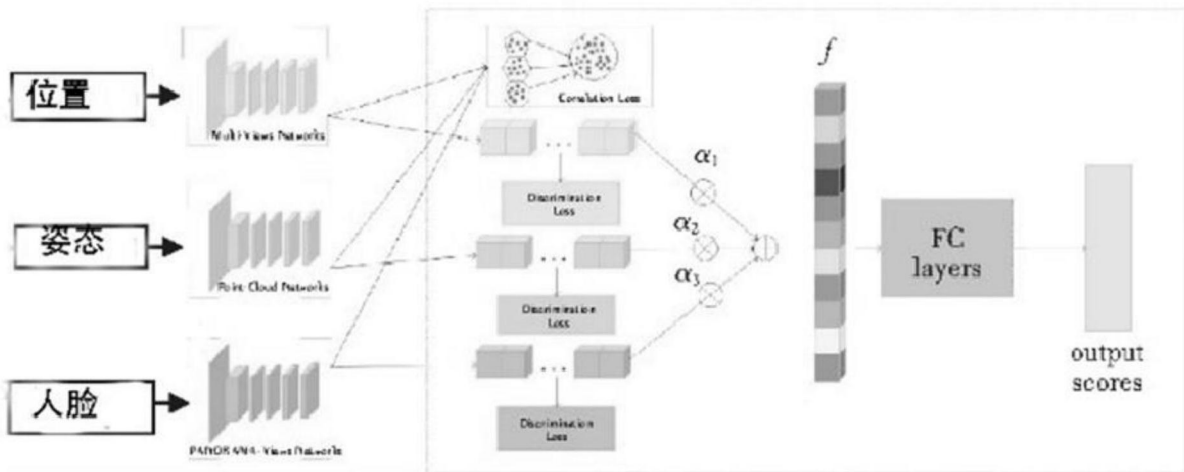


图3

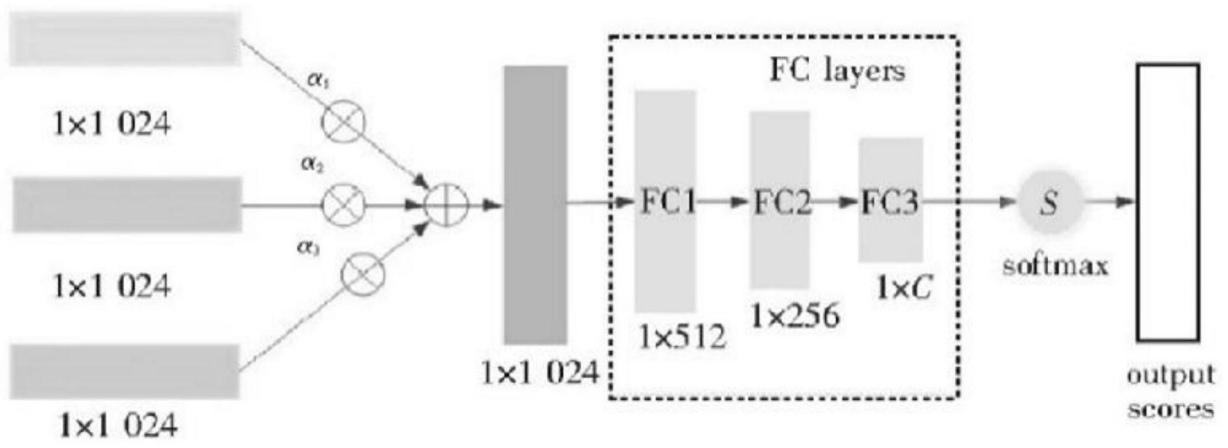


图4

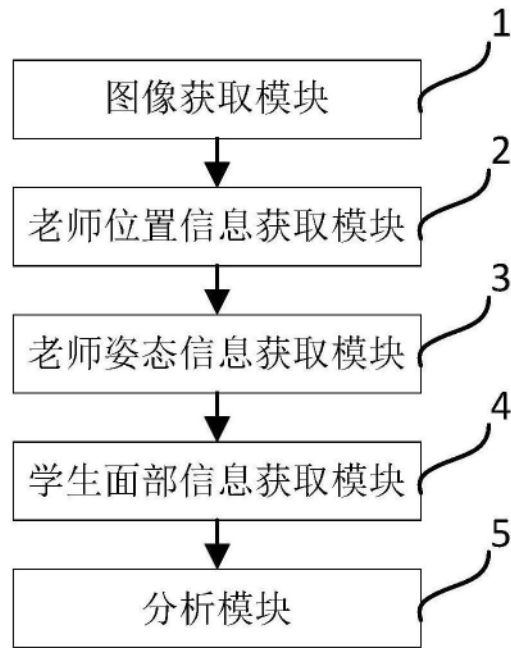


图5

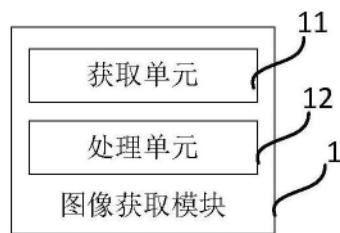


图6

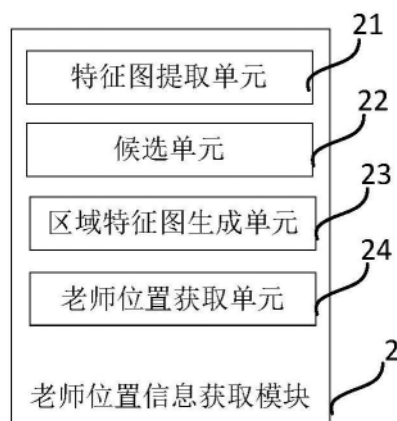


图7

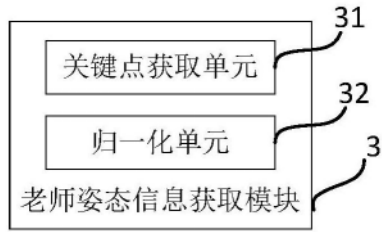


图8