

动作识别方法和装置、电子设备、存储介质

本申请要求在 2019 年 3 月 29 日提交中国专利局、申请号为 CN 201910252534.6、发明名称为“动作识别方法和装置、电子设备、存储介质”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

本申请涉及计算机视觉技术，尤其是一种动作识别方法和装置、电子设备、存储介质。

背景技术

在计算机视觉领域，动作识别问题一直是个大家较为关注的问题。对于动作识别，一般研究都聚焦在视频的时序特征，通过人体关键点能够判断的一些动作上。

发明内容

本申请实施例提供了一种动作识别技术。

根据本申请实施例的一个方面，提供一种动作识别方法，包括：

基于人脸图像获得人脸的嘴部关键点；

基于所述嘴部关键点确定第一区域内的图像，所述第一区域内的图像至少包括部分所述嘴部关键点以及与嘴部交互的物体的图像；

基于所述第一区域内的图像确定所述人脸图像中的人是否在吸烟。

根据本申请实施例的另一方面，提供一种动作识别装置，包括：

嘴部关键点单元，用于基于人脸图像获得人脸的嘴部关键点；

第一区域确定单元，用于基于所述嘴部关键点确定第一区域内的图像，所述第一区域内的图像至少包括部分所述嘴部关键点以及与嘴部交互的物体的图像；

吸烟识别单元，用于基于所述第一区域内的图像确定所述人脸图像中的人是否在吸烟。

根据本申请实施例的又一方面，提供一种电子设备，包括处理器，所述处理器包括上述任意一项实施例所述的动作识别装置。

根据本申请实施例的还一方面，提供一种电子设备，包括：存储器，用于存储可执行指令；

以及处理器，用于与所述存储器通信以执行所述可执行指令从而完成上述任意一项实施例所述动作识别方法的操作。

根据本申请实施例的再一方面，提供一种计算机可读存储介质，用于存储计算机可

读取的指令，所述指令被执行时执行上述任意一项实施例所述动作识别方法的操作。

根据本申请实施例的再一方面，提供一种计算机程序产品，包括计算机可读代码，当所述计算机可读代码在设备上运行时，所述设备中的处理器执行用于实现上述任意一项实施例所述动作识别方法的指令。

基于本申请上述实施例提供的一种动作识别方法和装置、电子设备、存储介质，基于人脸图像获得人脸的嘴部关键点；基于嘴部关键点确定第一区域内的图像，第一区域内的图像至少包括部分嘴部关键点以及与嘴部交互的物体的图像；基于第一区域内的图像确定人脸图像中的人是否在吸烟，识别以嘴部关键点确定的第一区域内的图像从而判断人脸图像中的人是否在吸烟，缩小了识别范围，将注意力集中在嘴部和与嘴部交互的物体上，提升了检出率，又降低了误检率，提高了吸烟识别的准确性。

下面通过附图和实施例，对本申请的技术方案做进一步的详细描述。

附图说明

构成说明书的一部分的附图描述了本申请的实施例，并且连同描述一起用于解释本申请的原理。

参照附图，根据下面的详细描述，可以更加清楚地理解本申请，其中：

图 1 为本申请实施例提供的动作识别方法的流程示意图。

图 2 为本申请实施例提供的动作识别方法的另一流程示意图。

图 3a 为本申请实施例提供的动作识别方法中一示例中识别获得的第一关键点示意图。

图 3b 为本申请实施例提供的动作识别方法中另一示例中识别获得的第一关键点示意图。

图 4 为本申请实施例提供的动作识别方法的又一流程示意图。

图 5 为本申请实施例提供的动作识别方法的还一个可选示例对与嘴部交互的物体执行对齐操作的示意图。

图 6a 为本申请实施例提供的动作识别方法中一个示例中采集的原始图像。

图 6b 为本申请实施例提供的动作识别方法中一个示例中检测到人脸框的示意图。

图 6c 为本申请实施例提供的动作识别方法中一个示例中基于关键点确定的第一区域示意图。

图 7 为本申请实施例提供的动作识别装置的一个结构示意图。

图 8 为适于用来实现本申请实施例的终端设备或服务器的电子设备的结构示意图。

具体实施方式

现在将参照附图来详细描述本申请的各种示例性实施例。应注意到：除非另外具体说明，否则在这些实施例中阐述的部件和步骤的相对布置、数字表达式和数值不限制本申请的范围。

同时，应当明白，为了便于描述，附图中所示出的各个部分的尺寸并不是按照实际的比例关系绘制的。

以下对至少一个示例性实施例的描述实际上仅仅是说明性的，决不作为对本申请及其应用或使用的任何限制。

对于相关领域普通技术人员已知的技术、方法和设备可能不作详细讨论，但在适当情况下，所述技术、方法和设备应当被视为说明书的一部分。

应注意到：相似的标号和字母在下面的附图中表示类似项，因此，一旦某一项在一个附图中被定义，则在随后的附图中不需要对其进行进一步讨论。

本申请实施例可以应用于计算机系统/服务器，其可与众多其它通用或专用计算系统环境或配置一起操作。适于与计算机系统/服务器一起使用的众所周知的计算系统、环境和/或配置的例子包括但不限于：个人计算机系统、服务器计算机系统、瘦客户机、厚客户机、手持或膝上设备、基于微处理器的系统、机顶盒、可编程消费电子产品、网络个人电脑、小型计算机系统、大型计算机系统和包括上述任何系统的分布式云计算技术环境，等等。

计算机系统/服务器可以在由计算机系统执行的计算机系统可执行指令（诸如程序模块）的一般语境下描述。通常，程序模块可以包括例程、程序、目标程序、组件、逻辑、数据结构等等，它们执行特定的任务或者实现特定的抽象数据类型。计算机系统/服务器可以在分布式云计算环境中实施，分布式云计算环境中，任务是由通过通信网络链接的远程处理设备执行的。在分布式云计算环境中，程序模块可以位于包括存储设备的本地或远程计算系统存储介质上。

图1为本申请实施例提供的动作识别方法的流程示意图。本实施例可应用在电子设备上，如图1所示，该实施例方法包括：

步骤110，基于人脸图像获得人脸的嘴部关键点。

本申请实施例中的嘴部关键点可以实现将人脸上的嘴部进行标注，可以采用现有技术中任意可实现的人脸关键点识别方法获取，例如，利用深度神经网络识别人脸上的人脸关键点，再从人脸关键点中分离得到嘴部关键点，或者，直接采用深度神经网络识别获得嘴部关键点，本申请实施例不限制具体获得嘴部关键点的方式。

在一个可选示例中，该步骤110可以由处理器调用存储器存储的相应指令执行，也可以由被处理器运行的嘴部关键点单元71执行。

步骤120，基于嘴部关键点确定第一区域内的图像。

其中，第一区域内的图像至少包括部分嘴部关键点以及与嘴部交互的物体的图像；本申请实施例提供的动作识别主要用于对图像中的人是否吸烟进行识别，由于吸烟的动作是通过嘴部与香烟接触实现的，因此，第一区域内不仅包括部分或全部的嘴部关键点，还可以包括与嘴部交互的物体，当该与嘴部交互的物体是香烟时，即可确定图像中的人在吸烟。可选地，本申请实施例中的第一区域可以是基于嘴部中心位置为中心点确定的矩形或圆形等任意形状的区域，本申请实施例不限制第一区域图像的形状和大小，以该第一区域中可能出现与嘴部接触的香烟、棒棒糖等交互物为准。

在一个可选示例中，该步骤 120 可以由处理器调用存储器存储的相应指令执行，也可以由被处理器运行的第一区域确定单元 72 执行。

步骤 130，基于第一区域内的图像确定人脸图像中的人是否在吸烟。

可选地，本申请实施例通过识别嘴部附近的区域中包括的与嘴部交互的物体是否是香烟来确定图像中的人是否在吸烟，将关注点集中在嘴部附近，降低了其他无关图像对识别结果产生干扰的概率，提高了对吸烟动作识别的准确性。

在一个可选示例中，该步骤 130 可以由处理器调用存储器存储的相应指令执行，也可以由被处理器运行的吸烟识别单元 73 执行。

基于本申请上述实施例提供的一种动作识别方法，基于人脸图像获得人脸的嘴部关键点；基于嘴部关键点确定第一区域内的图像，第一区域内的图像至少包括部分嘴部关键点以及与嘴部交互的物体的图像；基于第一区域内的图像确定人脸图像中的人是否在吸烟，识别以嘴部关键点确定的第一区域内的图像从而判断人脸图像中的人是否在吸烟，缩小了识别范围，将注意力集中在嘴部和与嘴部交互的物体上，提升了检出率，又降低了误检率，提高了吸烟识别的准确性。

图 2 为本申请实施例提供的动作识别方法的另一流程示意图。如图 1 所示，该实施例方法包括：

步骤 210，基于人脸图像获得人脸的嘴部关键点。

步骤 220，基于嘴部关键点确定第一区域内的图像。

步骤 230，基于第一区域内的图像获得与嘴部交互的物体上的至少两个第一关键点。

可选地，可通过神经网络对第一区域内的图像进行关键点提取，以获得与嘴部交互的物体的至少两个第一关键点，这些第一关键点在第一区域中可以表现为一条直线（例如，以香烟中轴线为香烟关键点）或两条直线（例如，以香烟两侧边为香烟关键点）等。

步骤 240，基于至少两个第一关键点对第一区域内的图像进行筛选。

其中，筛选的目的是确定出包含长度不小于预设值的与嘴部交互的物体的第一区域内的图像。

可选地，通过获得的与嘴部交互的物体上的至少两个第一关键点可确定第一区域内与

嘴部交互的物体的长度，当与嘴部交互的物体的长度较小（例如，与嘴部交互的物体长度小于预设值），第一区域中包括的与嘴部交互的物体不一定是香烟，此时可认为第一区域内的图像中不包括香烟；只有当与嘴部交互的物体的长度较大（例如，与嘴部交互的物体长度大于或等于预设值）时，才认为第一区域内的图像中可能包括香烟。

步骤 250，响应于第一区域内的图像通过筛选，基于第一区域内的图像确定人脸图像中的人是否在吸烟。

本申请实施例，上述筛选确定出部分第一区域内的图像，这部分第一区域内的图像中包含了长度达到设定值的与嘴部交互的物体，只有与嘴部交互的物体的长度达到设定值时，才认为该与嘴部交互的物体可能是香烟，本步骤中针对通过筛选的第一区域内的图像确定人脸图像中的人是否在吸烟，即，针对长度大于设定值的与嘴部交互的物体进行判断，判断该与嘴部交互的物体是否是香烟，以确定人脸图像中的人脸是否在吸烟。

可选地，步骤 240 包括：

基于至少两个第一关键点确定在第一区域内的图像中至少两个第一关键点对应的关键点坐标；

基于至少两个第一关键点对应的关键点坐标对第一区域内的图像进行筛选。

在获得与嘴部交互的物体的至少两个第一关键点之后并不能完全确定人脸图像中的人是否在吸烟，有可能只是在嘴部含了其他相似物体（如：棒棒糖或者其他长条形物体等），而香烟通常具有一定长度，为了确定第一区域中是否包括香烟，本申请实施例通过确定第一关键点的坐标，以第一关键点在第一区域中的坐标即可确定与嘴部交互的物体在第一区域图像中的长度，进而确定人脸图像中的人是否在吸烟。

可选地，基于至少两个第一关键点对应的关键点坐标对第一区域内的图像进行筛选，包括：

基于至少两个第一关键点对应的关键点坐标确定第一区域内的图像中的与嘴部交互的物体的长度；

响应于与嘴部交互的物体的长度大于或等于预设值，确定第一区域内的图像通过筛选。

可选地，在得到至少两个第一关键点的坐标之后，为了确定与嘴部交互的物体的长度，至少两个第一关键点中至少包括物体靠近嘴部一端的一个关键点和远离嘴部的一个关键点，例如，与嘴部交互的物体靠近嘴边的关键点分别为 p_1 、 p_2 ，远离嘴边的关键点定义分别为 p_3 、 p_4 。假设 p_1 与 p_2 之间的中点为 p_5 ，而 p_3 与 p_4 之间的中点为 p_6 。此时可以利用 p_5 和 p_6 的坐标来确定香烟的长度。

可选地，响应于与嘴部交互的物体的长度小于预设值，确定第一区域内的图像未通过筛选；确定第一区域内的图像中不包括香烟。

由于吸烟动作检测的一大难点在于如何区分香烟在图像上露出很少的一部分（即香烟

基本只露出一个横截面时)和驾驶员不在抽烟的状态,这要求神经网络提取的特征需要捕捉画面中嘴部非常微小的细节。如果要求网络将只露出一个横截面的抽烟图片也较为灵敏地检测出来,势必会引起算法的误检率升高。因此,本申请实施例提出依据与嘴部交互的物体的第一关键点来将与嘴部交互的物体露出部分很少或者驾驶员嘴上没有东西的图片直接在送入分类网络之前过滤掉。通过对训练后的网络进行测试可以发现,在关键点检测算法中,深度网络在利用梯度反向传播算法来更新网络参数后,会重点关注图像上与嘴部交互的物体的边缘信息,在大部分人没有做抽烟动作且嘴部周围没有条形物体会条纹干扰时,关键点的预测会趋于分布在嘴部中心的一个平均位置上(即使这时并没有香烟的存在)。根据上述特性,实现了通过第一关键点对与嘴部交互的物体只露出很少部分或者驾驶员嘴上没有东西的图像进行过滤(即,认为与嘴部交互的物体只露出少部分,接近只露出横截面的情况下,图像上抽烟判断依据不足,认为第一区域中不包括香烟)。

可选地,步骤240还包括:

为至少两个第一关键点中的每个第一关键点分配用于区分每个第一关键点的序号。

通过为至少两个第一关键点中的每个第一关键点分配不同序号,可将每个第一关键点进行区分,并通过不同的第一关键点实现不同的目的,例如:距离嘴部关键点最近的第一关键点和距离嘴部距离最远的第一关键点可确定当前香烟的长度。本申请实施例可按照任意不重复的顺序为第一关键点分配序号,实现区别每个不同的第一关键点即可,本申请实施例不限制具体的分配序号的方式,例如,按照叉乘法则的顺序对至少两个第一关键点中的每个第一关键点分配不同序号。

在一个或多个可选的实施例中,基于至少两个第一关键点确定在第一区域内的图像中至少两个第一关键点对应的关键点坐标,包括:

利用第一神经网络确定第一区域内的图像中的至少两个第一关键点对应的关键点坐标。

其中,第一神经网络经过第一样本图像训练获得。

可选地,第一样本图像包括标注关键点坐标;

训练第一神经网络的过程包括:

将第一样本图像输入第一神经网络,获得至少两个第一关键点对应的预测关键点坐标;

基于预测关键点坐标和标注关键点坐标确定第一网络损失,基于第一网络损失调整第一神经网络的参数。

可选地,第一关键点定位任务,与人脸关键点定位任务类似,也可以看作是一个回归任务,从而得到第一关键点的二维坐标 (x_i, y_i) 的映射函数,算法描述如下:

将第一神经网络第1层的输入记为 x_1 (即输入图像),中间层输出为 x_n ,每一层网络相当于一个非线性函数映射 $F(x)$,假设第一神经网络一共有N层,那么经过第一神经网络的

非线性映射之后，网络的输出可以抽象为公式（1）表达：

$$\hat{y} = F^N(x) \quad \text{公式（1）}$$

其中， $\hat{y}$ 是第一神经网络输出的一维向量，该一维向量中的每个值表示关键点网络最终输出的关键点坐标。

在一个或多个可选的实施例中，步骤 230 包括：

对第一区域内的图像进行与嘴部交互的物体的关键点识别，获得与嘴部交互的物体的中轴线上的至少两个中轴关键点，和/或与嘴部交互的物体的两条边中每条边上的至少两个边关键点。

本申请实施例中定义第一关键点时，可以将图像中与嘴部交互的物体的中轴线上的中轴关键点作为第一关键点，和/或将图像中与嘴部交互的物体的两条边上的边关键点作为第一关键点，可选地，为了进行后续的关键点对齐，选择两条边的关键点定义。图 3a 为本申请实施例提供的动作识别方法中一示例中识别获得的第一关键点示意图。图 3b 为本申请实施例提供的动作识别方法中另一示例中识别获得的第一关键点示意图。如图 3a 和 3b 所示，选择两条边关键点定义第一关键点，为了识别不同的第一关键点，并获得不同第一关键点对应的关键点坐标，还可以为每个第一关键点分配不同的序号。

图 4 为本申请实施例提供的动作识别方法的又一流程示意图。如图 4 所示，该实施例方法包括：

步骤 410，基于人脸图像获得人脸的嘴部关键点。

步骤 420，基于嘴部关键点确定第一区域内的图像。

步骤 430，基于所第一区域内的图像获得与嘴部交互的物体上的至少两个第二关键点。

可选地，本申请实施例中获得的第二关键点与上述实施例中的第一关键点都是与嘴部交互的物体上的关键点，第二关键点可以与第一关键点相同或不同。

步骤 440，基于至少两个第二关键点对与嘴部交互的物体执行对齐操作，使与嘴部交互的物体朝向预设方向，获得包括朝向预设方向的与嘴部交互的物体的第二区域内的图像。

其中，第二区域内的图像至少包括部分嘴部关键点以及与嘴部交互的物体的图像。

本申请实施例中通过获得第二关键点对与嘴部交互的物体进行对齐操作，使与嘴部交互的物体朝向预设方向，获得包括朝向预设方向的与嘴部交互的物体的第二区域，第二区域与上述实施例中的第一区域可存在重叠部分，例如，第二区域包括至少第一区域内的图像中的部分嘴部关键点以及与嘴部交互的物体的图像。本申请实施例提供的动作识别方法可能包括多种实现方式，例如：如果只对第一区域内的图像执行筛选操作，那么，只需确定与嘴部交互的物体的第一关键点，基于至少两个第一关键点对第一区域内的图像进行筛选。如果只与嘴部交互的物体执行对齐操作，那么只需确定与嘴部交互的物体的第二关键点，基于至少两个第二关键点对与嘴部交互的物体执行对齐操作。如果既执行筛选操作，

又执行对齐操作，那么需要确定与嘴部交互的物体的第一关键点和第二关键点，其中，第一关键点和第二关键点可以相同或不同，第二关键点及其坐标的确定方式可以参考第一关键点及其坐标的确定方式，并且，本申请实施例不限制筛选操作和对齐操作的操作顺序。

可选地，步骤 440 可基于至少两个第二关键点获得对应的关键点坐标，基于获得的第二关键点的坐标实现对齐操作，而基于第二关键点获得关键点坐标的过程也可以与基于第一关键点获得关键点坐标类似，通过神经网络获得，本申请实施例不限制基于第二关键点至少对齐操作的具体方式。

可选地，步骤 440 还可以包括为至少两个第二关键点中的每个第二关键点分配用于区分每个第二关键点的序号。分配序号的规则可参照对第一关键点分配序号的方式，在此不再赘述。

步骤 450，基于第二区域内的图像确定人脸图像中的人是否在吸烟。

由于卷积神经网络旋转不变性较差，神经网络在物体的不同旋转程度下的特征提取存在一定的差异。而人在吸烟时，香烟的朝向是在各个方向上的，如果直接在原截取图片上直接进行特征提取，可能会出现一定程度的是否吸烟的结果检测性能下降。换句话说，神经网络需要适应香烟在不同角度上的特征提取，从而进行一定程度的解耦。本申请实施例中通过基于第二关键点进行对齐操作，使每个输入的人脸图像中的与嘴部交互的物体都朝向同一方向，可以降低误检的概率。

可选地，对齐操作可以包括：

基于至少两个第二关键点获得关键点坐标，基于至少两个第二关键点对应的关键点坐标获得与嘴部交互的物体；

利用仿射变换基于预设方向对与嘴部交互的物体执行对齐操作，使与嘴部交互的物体朝向预设方向，获得包括朝向预设方向的与嘴部交互的物体的第二区域内的图像。

其中，仿射变换可以包括但不限于以下至少之一：旋转、缩放、平移、翻转、剪切等。

本申请实施例中，通过仿射变换将与嘴部交互的物体的图像上的像素映射到一个新的通过关键点对齐之后的图片上。使得原有的第二关键点和事先设定好的关键点进行对齐。这样可以将图像中与嘴部交互的物体的信号和与嘴部交互的物体的角度信息进行解耦，从而提升后续神经网络的特征提取性能。图 5 为本申请实施例提供的动作识别方法的再一个可选示例对与嘴部交互的物体执行对齐操作的示意图。如图 5 所示，通过利用第二关键点和目标位置进行仿射变换将第一区域图像中的与嘴部交互的物体方向进行转换，本示例中，将与嘴部交互的物体（香烟）的方向转向向下。

关键点对齐是通过仿射变换(Affine Transformation)实现的。仿射变换的功能是从二维坐标到二维坐标之间的线性变换，且保持二维图形的“平直性”和“平行性”。仿射变换可以通过一系列的原子变换的复合来实现，其中，原子变换可以包括但不限于：平移，缩

放，翻转，旋转和剪切等。

仿射变换的其次坐标系表示如公式（2）所示：

$$\begin{bmatrix} x' & y' & 1 \end{bmatrix} = \begin{bmatrix} x & y & 1 \end{bmatrix} \cdot \begin{bmatrix} a_{11} & a_{12} & 0 \\ a_{12} & a_{22} & 0 \\ x_0 & y_0 & 1 \end{bmatrix} \quad \text{公式（2）}$$

其中， $\begin{bmatrix} x' & y' & 1 \end{bmatrix}$ 表示仿射变换之后得到的坐标， $\begin{bmatrix} x & y & 1 \end{bmatrix}$ 表示提取获得的香烟关键点的关键点坐标， $\begin{bmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{bmatrix}$ 表示旋转矩阵， x_0 和 y_0 表示平移向量。

以上的表达式涵盖了旋转、平移、缩放、旋转几个操作。假设模型给出的关键点为 (x_i, y_i) 的集合，设置的目标点位置 (x_i', y_i') （此处的目标点位置可以通过人为实现设定），仿射变换矩阵将源图像进行仿射变换到目标图像，截取之后得到转正之后的图片。

可选地，步骤 130 包括：

利用第二神经网络基于第一区域内的图像确定人脸图像中的人是否在吸烟。

其中，第二神经网络经过第二样本图像训练获得。第二样本图像包括吸烟的样本图像以及非吸烟的样本图像，这样可以训练神经网络将香烟和其他的细长物体区分开，从而识别出到底是在吸烟，还是嘴里叼着别的东西。

本申请实施例中，将获得的关键点坐标输入到第二神经网络（例如，分类卷积神经网络），进行分类，可选地，操作过程也是由卷积神经网络进行特征提取，在最后输出二分类的结果，即拟合出该图像属于抽烟或者不抽烟的图像的概率。

可选地，第二样本图像标注有图像中的人是否在吸烟的标注结果；

训练第二神经网络的过程包括：

将第二样本图像输入第二神经网络，获得第二样本图像中的人是否在吸烟的预测结果；

基于预测结果和标注结果获得第二网络损失，基于第二网络损失调整第二神经网络的参数。

可选地，对第二神经网络的训练中，网络监督可以采用 softmax 损失函数，数学表达式如下：

p_i 为第二神经网络输出的第 i 个第二样本图像的预测结果为实际正确类别（标注结果）的概率， N 为总样本数。

损失函数可以采用以下公式（3）：

$$L_{\text{softmax}} = -\frac{1}{N} \sum_{i=1}^N \log(p_i) \quad \text{公式（3）}$$

定义了网络结构和损失函数之后，训练只需要根据梯度反向传播的计算方式去更新网络参数即可，得到训练之后的第二神经网络的网络参数。

在训练好第二神经网络之后，去掉损失函数并且固定网络参数不变，对预处理好的图像同样输入到卷积神经网络抽取特征和分类，这样就能得到分类模块给出的分类结果。由

此来判断画面中的人是否在抽烟。

在一个或多个可选的实施例中，步骤 110 包括：

对人脸图像进行人脸关键点提取，获得人脸图像中的人脸关键点；

基于人脸关键点获得嘴部关键点。

可选地，通过神经网络对人脸图像进行人脸关键点提取，由于抽烟动作和人的交互方式主要是用嘴和手进行的，抽烟动作在进行的时候基本是在嘴部附近的，因此可以通过人脸检测和人脸关键点定位技术将有效信息区域（第一区域图像）缩小到嘴部附近；可选地，对提取的人脸关键点进行编辑序号，可以通过设定某些序号的关键点为嘴部关键或通过人脸关键点在人脸图像中的位置获得嘴部关键点，基于嘴部关键点确定第一区域图像。

一些可选的示例中，本申请实施例的人脸图像是通过人脸检测获得的，采集的图像经过人脸检测获得人脸图像，人脸检测是整个抽烟动作识别的底层基础模块，由于抽烟人在抽烟的时候画面上一定会出现人脸，因此可以通过人脸检测来粗定位人脸的位置，本申请实施例不限制具体的人脸检测算法。

在通过人脸检测得到人脸框之后，将人脸框内的图像（对应上述实施例中的人脸图像）截取出来并进行人脸关键点提取。可选地，人脸关键点定位任务实际上可以抽象为一个回归任务：给定一幅包含人脸信息的图像，拟合出图像中关键点的二维坐标 (x_i, y_i) 的映射函数：对于一张输入图像，将检测出的人脸位置截取出来，网络的拟合只在一个局部图像的范围内进行，提高了拟合的速度。人脸关键点主要包括人的五官关键点，本申请实施例主要关注嘴部的关键点，例如：嘴角点，嘴唇轮廓关键点等。

可选地，基于嘴部关键点确定第一区域内的图像，包括：

基于嘴部关键点确定人脸中嘴部的中心位置；

以嘴部的中心位置作为第一区域的中心点，以设定长度为边长或半径确定第一区域。

本申请实施例中，为了将可能出现香烟的区域包括在第一区域中，将嘴部的中心位置确定为第一区域图像的中心点，以设定长度为半径或边长确定一个矩形或圆形的第一区域，可选地，设定长度可以事先设定，或根据嘴部中心位置与人脸中某个关键点的距离确定。例如：可基于嘴部关键点与眉部关键点之间的距离确定设定长度。

可选地，基于人脸关键点获得眉部关键点；

以嘴部的中心位置作为第一区域中心点，以设定长度为边长或半径确定第一区域，包括：

以嘴部的中心位置作为中心点，以嘴部的中心位置到眉心的垂直距离作为边长或半径，确定第一区域。

其中，眉心是基于眉部关键点确定的。

例如，定位出人脸关键点之后，计算嘴部中心和眉心的垂直距离 d ，然后得到以嘴部

中心为中心， $2d$ 为边长的正方形区域 R ，将 R 区域图像作为本申请实施例的第一区域。

图 6a 为本申请实施例提供的动作识别方法中一个示例中采集的原始图像。图 6b 为本申请实施例提供的动作识别方法中一个示例中检测到人脸框的示意图。图 6c 为本申请实施例提供的动作识别方法中一个示例中基于关键点确定的第一区域示意图。在一个可选示例中，通过图 6a、6b 和 6c，实现了基于采集的原始图像获得第一区域的过程。

本领域普通技术人员可以理解：实现上述方法实施例的全部或部分步骤可以通过程序指令相关的硬件来完成，前述的程序可以存储于一计算机可读取存储介质中，该程序在执行时，执行包括上述方法实施例的步骤；而前述的存储介质包括：ROM、RAM、磁碟或者光盘等各种可以存储程序代码的介质。

图 7 为本申请实施例提供的动作识别装置的一个结构示意图。该实施例的装置可用于实现本申请上述各方法实施例。如图 7 所示，该实施例的装置包括：

嘴部关键点单元 71，用于基于人脸图像获得人脸的嘴部关键点。

第一区域确定单元 72，用于基于嘴部关键点确定第一区域内的图像。

其中，第一区域内的图像至少包括部分嘴部关键点以及与嘴部交互的物体的图像。

吸烟识别单元 73，用于基于第一区域内的图像确定人脸图像中的人是否在吸烟。

基于本申请上述实施例提供的一种动作识别装置，基于人脸图像获得人脸的嘴部关键点；基于嘴部关键点确定第一区域内的图像，第一区域内的图像至少包括部分嘴部关键点以及与嘴部交互的物体的图像；基于第一区域内的图像确定人脸图像中的人是否在吸烟，以嘴部关键点确定的第一区域识别是否在吸烟，缩小了识别范围，将注意力集中在嘴部和与嘴部交互的物体上，提升了检出率，又降低了误检率，提高了吸烟识别的准确性。

在一个或多个可选的实施例中，装置还包括：

第一关键点单元，用于基于第一区域内的图像获得与嘴部交互的物体上的至少两个第一关键点；

图像筛选单元，用于基于至少两个第一关键点对第一区域内的图像进行筛选，筛选用于确定第一区域内的与嘴部交互物的长度；其中，对所述第一区域内的图像进行筛选是确定出包含长度不小于预设值的与嘴部交互的物体的图像的第一区域内的图像；

吸烟识别单元 73，用于响应于第一区域内的图像通过筛选，基于第一区域内的图像确定人脸图像中的人是否在吸烟。

可选地，图像筛选单元，用于基于至少两个第一关键点确定在第一区域内的图像中至少两个第一关键点对应的关键点坐标；基于至少两个第一关键点对应的关键点坐标对第一区域内的图像进行筛选。

可选地，图像筛选单元在基于至少两个第一关键点对应的关键点坐标对第一区域内的图像进行筛选时，用于基于至少两个第一关键点对应的关键点坐标确定第一区域内的图像

中与嘴部交互的物体的长度；响应于与嘴部交互的物体的长度大于或等于预设值，确定第一区域内的图像通过筛选。

可选地，图像筛选单元在基于至少两个第一关键点对应的关键点坐标对第一区域内的图像进行筛选时，还用于响应于与嘴部交互的物体的长度小于预设值，确定第一区域内的图像未通过筛选；确定第一区域内的图像中不包括香烟。

可选地，图像筛选单元，还用于为至少两个第一关键点中的每个第一关键点分配用于区分每个第一关键点的序号。

可选地，图像筛选单元在基于至少两个第一关键点确定在第一区域内的图像中至少两个第一关键点对应的关键点坐标时，用于利用第一神经网络确定第一区域内的图像中的至少两个第一关键点对应的关键点坐标，第一神经网络经过第一样本图像训练获得。

可选地，第一样本图像包括标注关键点坐标；训练第一神经网络的过程包括：

将第一样本图像输入第一神经网络，获得至少两个第一关键点对应的预测关键点坐标；

基于预测关键点坐标和标注关键点坐标确定第一网络损失，基于第一网络损失调整第一神经网络的参数。

可选地，第一关键点单元，用于对第一区域内的图像进行与嘴部交互的物体的关键点识别，获得与嘴部交互的物体的中轴线上的至少两个中轴关键点，和/或与嘴部交互的物体的两条边中每条边上的至少两个边关键点。

在一个或多个可选的实施例中，本申请实施例提供的装置还包括：

第二关键点单元，用于基于第一区域内的图像获得与嘴部交互的物体上的至少两个第二关键点；

图像对齐单元，用于基于至少两个第二关键点对与嘴部交互的物体执行对齐操作，使与嘴部交互的物体朝向预设方向，获得包括朝向预设方向的与嘴部交互的物体的第二区域内的图像，第二区域内的图像至少包括部分嘴部关键点以及与嘴部交互的物体的图像；

吸烟识别单元 73，用于基于第二区域内的图像确定人脸图像中的人是否在吸烟。

在一个或多个可选的实施例中，吸烟识别单元 73，用于利用第二神经网络基于第一区域内的图像确定人脸图像中的人是否在吸烟，第二神经网络经过第二样本图像训练获得。

可选地，第二样本图像标注有图像中的人是否在吸烟的标注结果；训练第二神经网络的过程包括：

将第二样本图像输入第二神经网络，获得第二样本图像中的人是否在吸烟的预测结果；

基于预测结果和标注结果获得第二网络损失，基于第二网络损失调整第二神经网络的参数。

在一个或多个可选的实施例中，嘴部关键点单元 71，用于对人脸图像进行人脸关键点提取，获得人脸图像中的人脸关键点；基于人脸关键点获得嘴部关键点。

可选地，第一区域确定单元 72，用于基于嘴部关键点确定人脸中嘴部的中心位置；以嘴部的中心位置作为第一区域的中心点，以设定长度为边长或半径确定第一区域。

可选地，本申请实施例提供的装置还包括：

眉部关键点单元，用于基于人脸关键点获得眉部关键点；

第一区域确定单元 72，用于以嘴部的中心位置作为中心点，以嘴部的中心位置到眉心的垂直距离作为边长或半径，确定第一区域，眉心基于眉部关键点确定。

本公开实施例提供的动作识别装置任一实施例的工作过程、设置方式及相应技术效果，均可以参照本公开上述相应方法实施例的具体描述，限于篇幅，在此不再赘述。

根据本申请实施例的又一个方面，提供一种电子设备，包括处理器，该处理器包括如上任意一实施例提供的动作识别装置。

根据本申请实施例的还一个方面，提供一种电子设备，包括：存储器，用于存储可执行指令；

以及处理器，用于与存储器通信以执行可执行指令从而完成如上任意一实施例提供的动作识别方法的操作。

根据本申请实施例的再一个方面，提供一种计算机可读存储介质，用于存储计算机可读取的指令，指令被执行时执行如上任意一实施例提供的动作识别方法的操作。

根据本申请实施例的又一个方面，提供一种计算机程序产品，包括计算机可读代码，当计算机可读代码在设备上运行时，设备中的处理器执行用于实现如上任意一实施例提供的动作识别方法的指令。

本申请实施例还提供了一种电子设备，例如可以是移动终端、个人计算机（PC）、平板电脑、服务器等。下面参考图 8，其示出了适于用来实现本申请实施例的终端设备或服务器的电子设备 800 的结构示意图：如图 8 所示，电子设备 800 包括一个或多个处理器、通信部等，所述一个或多个处理器例如：一个或多个中央处理单元（CPU）801，和/或一个或多个图像处理器（加速单元）813 等，处理器可以根据存储在只读存储器（ROM）802 中的可执行指令或者从存储部分 808 加载到随机访问存储器（RAM）803 中的可执行指令而执行各种适当的动作和处理。通信部 812 可包括但不限于网卡，所述网卡可包括但不限于 IB（Infiniband）网卡。

处理器可与只读存储器 802 和/或随机访问存储器 803 中通信以执行可执行指令，通过总线 804 与通信部 812 相连、并经通信部 812 与其他目标设备通信，从而完成本申请实施例提供的任一项方法对应的操作，例如，基于人脸图像获得人脸的嘴部关键点；基于嘴部关键点确定第一区域内的图像，第一区域内的图像至少包括部分嘴部关键点以及与嘴部交互的物体的图像；基于第一区域内的图像确定人脸图像中的人是否在吸烟。

此外,在 RAM 803 中,还可存储有装置操作所需的各种程序和数据。CPU801、ROM802 以及 RAM803 通过总线 804 彼此相连。在有 RAM803 的情况下,ROM802 为可选模块。RAM803 存储可执行指令,或在运行时向 ROM802 中写入可执行指令,可执行指令使中央处理单元 801 执行上述通信方法对应的操作。输入/输出(I/O)接口 805 也连接至总线 804。通信部 812 可以集成设置,也可以设置为具有多个子模块(例如多个 IB 网卡),并在总线链接上。

以下部件连接至 I/O 接口 805:包括键盘、鼠标等的输入部分 806;包括诸如阴极射线管(CRT)、液晶显示器(LCD)等以及扬声器等的输出部分 807;包括硬盘等的存储部分 808;以及包括诸如 LAN 卡、调制解调器等网络接口卡的通信部分 809。通信部分 809 经由诸如因特网的网络执行通信处理。驱动器 810 也根据需要连接至 I/O 接口 805。可拆卸介质 811,诸如磁盘、光盘、磁光盘、半导体存储器等等,根据需要安装在驱动器 810 上,以便于从其上读出的计算机程序根据需要被安装入存储部分 808。

需要说明的,如图 8 所示的架构仅为一种可选实现方式,在具体实践过程中,可根据实际需要对上述图 8 的部件数量和类型进行选择、删减、增加或替换;在不同功能部件设置上,也可采用分离设置或集成设置等实现方式,例如加速单元 813 和 CPU801 可分离设置或者可将加速单元 813 集成在 CPU801 上,通信部可分离设置,也可集成设置在 CPU801 或加速单元 813 上,等等。这些可替换的实施方式均落入本申请公开的保护范围。

特别地,根据本申请的实施例,上文参考流程图描述的过程可以被实现为计算机软件程序。例如,本申请的实施例包括一种计算机程序产品,其包括有形地包含在机器可读介质上的计算机程序,计算机程序包含用于执行流程图所示的方法的程序代码,程序代码可包括对应执行本申请实施例提供的方法步骤对应的指令,例如,基于人脸图像获得人脸的嘴部关键点;基于嘴部关键点确定第一区域内的图像,第一区域内的图像至少包括部分嘴部关键点以及与嘴部交互的物体的图像;基于第一区域内的图像确定人脸图像中的人是否在吸烟。在这样的实施例中,该计算机程序可以通过通信部分 809 从网络上被下载和安装,和/或从可拆卸介质 811 被安装。在该计算机程序被中央处理单元(CPU)801 执行时,执行本申请的方法中限定的上述功能的操作。

本说明书中各个实施例均采用递进的方式描述,每个实施例重点说明的都是与其它实施例的不同之处,各个实施例之间相同或相似的部分相互参见即可。对于系统实施例而言,由于其与方法实施例基本对应,所以描述的比较简单,相关之处参见方法实施例的部分说明即可。

可能以许多方式来实现本申请的方法和装置。例如,可通过软件、硬件、固件或者软件、硬件、固件的任何组合来实现本申请的方法和装置。用于所述方法的步骤的上述顺序仅是为了进行说明,本申请的方法的步骤不限于以上具体描述的顺序,除非以其它方式特

别说明。此外，在一些实施例中，还可将本申请实施为记录在记录介质中的程序，这些程序包括用于实现根据本申请的方法的机器可读指令。因而，本申请还覆盖存储用于执行根据本申请的方法的程序的记录介质。

本申请的描述是为了示例和描述起见而给出的，而并不是无遗漏的或者将本申请限于所公开的形式。很多修改和变化对于本领域的普通技术人员而言是显然的。选择和描述实施例是为了更好说明本申请的原理和实际应用，并且使本领域的普通技术人员能够理解本申请从而设计适于特定用途的带有各种修改的各种实施例。

权 利 要 求

1、一种动作识别方法，其特征在于，包括：

基于人脸图像获得人脸的嘴部关键点；

基于所述嘴部关键点确定第一区域内的图像，所述第一区域内的图像至少包括部分所述嘴部关键点以及与嘴部交互的物体的图像；

基于所述第一区域内的图像确定所述人脸图像中的人是否在吸烟。

2、根据权利要求 1 所述的方法，其特征在于，在基于所述第一区域内的图像确定所述人脸图像中的人是否在吸烟之前，所述方法还包括：

基于所述第一区域内的图像获得与嘴部交互的物体上的至少两个第一关键点；

基于所述至少两个第一关键点对所述第一区域内的图像进行筛选；其中，对所述第一区域内的图像进行筛选是确定出包含长度不小于预设值的与嘴部交互的物体的第一区域内的图像；

基于所述第一区域内的图像确定所述人脸图像中的人是否在吸烟，包括：

响应于所述第一区域内的图像通过筛选，基于所述第一区域内的图像确定所述人脸图像中的人是否在吸烟。

3、根据权利要求 2 所述的方法，其特征在于，所述基于所述至少两个第一关键点对所述第一区域内的图像进行筛选，包括：

基于所述至少两个第一关键点确定在所述第一区域内的图像中所述至少两个第一关键点对应的关键点坐标；

基于所述至少两个第一关键点对应的关键点坐标对所述第一区域内的图像进行筛选。

4、根据权利要求 3 所述的方法，其特征在于，所述基于所述至少两个第一关键点对应的关键点坐标对所述第一区域内的图像进行筛选，包括：

基于所述至少两个第一关键点对应的关键点坐标确定所述第一区域内的图像中与嘴部交互的物体的长度；

响应于所述与嘴部交互的物体的长度大于或等于预设值，确定所述第一区域内的图像通过筛选。

5、根据权利要求 4 所述的方法，其特征在于，所述方法还包括：

响应于所述与嘴部交互的物体的长度小于预设值，确定所述第一区域内的图像未通过筛选；确定所述第一区域内的图像中不包括香烟。

6、根据权利要求 3-5 任一所述的方法，其特征在于，所述基于所述至少两个第一关键点确定在所述第一区域内的图像中所述至少两个第一关键点对应的关键点坐标之前，还包括：

为所述至少两个第一关键点中的每个所述第一关键点分配用于区分每个所述第一关键点的序号。

7、根据权利要求 3-6 任一所述的方法，其特征在于，所述基于所述至少两个第一关键点确定在所述第一区域内的图像中所述至少两个第一关键点对应的关键点坐标，包括：

利用第一神经网络确定所述第一区域内的图像中的所述至少两个第一关键点对应的关键点坐标，所述第一神经网络经过第一样本图像训练获得。

8、根据权利要求 7 所述的方法，其特征在于，所述第一样本图像包括标注关键点坐标；

所述训练第一神经网络的过程包括：

将所述第一样本图像输入所述第一神经网络，获得至少两个第一关键点对应的预测关键点坐标；

基于所述预测关键点坐标和所述标注关键点坐标确定第一网络损失，基于所述第一网络损失调整所述第一神经网络的参数。

9、根据权利要求 2-8 任一所述的方法，其特征在于，所述基于所述第一区域内的图像获得与嘴部交互的物体上的至少两个第一关键点，包括：

对所述第一区域内的图像进行与嘴部交互的物体的关键点识别，获得所述与嘴部交互的物体的中轴线上的至少两个中轴关键点，和/或所述与嘴部交互的物体的两条边中每条边上的至少两个边关键点。

10、根据权利要求 1-9 任一所述的方法，其特征在于，在基于所述第一区域内的图像确定所述人脸图像中的人是否在吸烟之前，所述方法还包括：

基于所述第一区域内的图像获得与嘴部交互的物体上的至少两个第二关键点；

基于所述至少两个第二关键点对所述与嘴部交互的物体执行对齐操作，使所述与嘴部交互的物体朝向预设方向，获得包括所述朝向预设方向的与嘴部交互的物体的第二区域内的图像，所述第二区域内的图像至少包括部分所述嘴部关键点以及与嘴部交互的物体的图像；

所述基于所述第一区域内的图像确定所述人脸图像中的人是否在吸烟，包括：基于所述第二区域内的图像确定所述人脸图像中的人是否在吸烟。

11、根据权利要求 1-10 任一所述的方法，其特征在于，所述基于所述第一区域内的图像确定所述人脸图像中的人是否在吸烟，包括：

利用第二神经网络基于所述第一区域内的图像确定所述人脸图像中的人是否在吸烟，所述第二神经网络经过第二样本图像训练获得。

12、根据权利要求 11 所述的方法，其特征在于，所述第二样本图像标注有图像中的人是否在吸烟的标注结果；

所述训练第二神经网络的过程包括：

将所述第二样本图像输入所述第二神经网络，获得所述第二样本图像中的人是否在吸烟的预测结果；

基于所述预测结果和所述标注结果获得第二网络损失，基于所述第二网络损失调整所述第二神经网络的参数。

13、根据权利要求 1-12 任一所述的方法，其特征在于，所述基于人脸图像获得人脸的嘴部关键点，包括：

对所述人脸图像进行人脸关键点提取，获得所述人脸图像中的人脸关键点；

基于所述人脸关键点获得所述嘴部关键点。

14、根据权利要求 13 所述的方法，其特征在于，所述基于所述嘴部关键点确定所述第一区域内的图像，包括：

基于所述嘴部关键点确定所述人脸中嘴部的中心位置；

以所述嘴部的中心位置作为所述第一区域的中心点，以设定长度为边长或半径确定所述第一区域。

15、根据权利要求 14 所述的方法，其特征在于，所述基于所述嘴部关键点确定所述第一区域内的图像之前，还包括：

基于所述人脸关键点获得眉部关键点；

所述以所述嘴部的中心位置作为所述第一区域的中心点，以设定长度为边长或半径确定所述第一区域，包括：

以所述嘴部的中心位置作为中心点，以所述嘴部的中心位置到眉心的垂直距离作为边长或半径，确定所述第一区域，所述眉心基于所述眉部关键点确定。

16、一种动作识别装置，其特征在于，包括：

嘴部关键点单元，用于基于人脸图像获得人脸的嘴部关键点；

第一区域确定单元，用于基于所述嘴部关键点确定所述第一区域内的图像，所述第一区域内的图像至少包括部分所述嘴部关键点以及与嘴部交互的物体的图像；

吸烟识别单元，用于基于所述第一区域内的图像确定所述人脸图像中的人是否在吸烟。

17、根据权利要求 16 所述的装置，其特征在于，所述装置还包括：

第一关键点单元，用于基于所述第一区域内的图像获得与嘴部交互的物体上的至少两个第一关键点；

图像筛选单元，用于基于所述至少两个第一关键点对所述第一区域内的图像进行筛选；其中，对所述第一区域内的图像进行筛选是确定出包含长度不小于预设值的与嘴部交互的物体的图像的第一区域内的图像；

所述吸烟识别单元，用于响应于所述第一区域内的图像通过筛选，基于所述第一区域

内的图像确定所述人脸图像中的人是否在吸烟。

18、根据权利要求 17 所述的装置，其特征在于，所述图像筛选单元，用于基于所述至少两个第一关键点确定在所述第一区域内的图像中所述至少两个第一关键点对应的关键点坐标；基于所述至少两个第一关键点对应的关键点坐标对所述第一区域内的图像进行筛选。

19、根据权利要求 18 所述的装置，其特征在于，所述图像筛选单元在基于所述至少两个第一关键点对应的关键点坐标对所述第一区域内的图像进行筛选时，用于基于所述至少两个第一关键点对应的关键点坐标确定所述第一区域内的图像中与嘴部交互的物体的长度；响应于所述与嘴部交互的物体的长度大于或等于预设值，确定所述第一区域内的图像通过筛选。

20、根据权利要求 19 所述的装置，其特征在于，所述图像筛选单元在基于所述至少两个第一关键点对应的关键点坐标对所述第一区域内的图像进行筛选时，还用于响应于所述与嘴部交互的物体的长度小于预设值，确定所述第一区域内的图像未通过筛选；确定所述第一区域内的图像中不包括香烟。

21、根据权利要求 18-20 任一所述的装置，其特征在于，所述图像筛选单元，还用于为所述至少两个第一关键点中的每个所述第一关键点分配用于区分每个所述第一关键点的序号。

22、根据权利要求 18-21 任一所述的装置，其特征在于，所述图像筛选单元在基于所述至少两个第一关键点确定在所述第一区域内的图像中所述至少两个第一关键点对应的关键点坐标时，用于利用第一神经网络确定所述第一区域内的图像中的所述至少两个第一关键点对应的关键点坐标，所述第一神经网络经过第一样本图像训练获得。

23、根据权利要求 22 所述的装置，其特征在于，所述第一样本图像包括标注关键点坐标；

所述训练第一神经网络的过程包括：

将所述第一样本图像输入所述第一神经网络，获得至少两个第一关键点对应的预测关键点坐标；

基于所述预测关键点坐标和所述标注关键点坐标确定第一网络损失，基于所述第一网络损失调整所述第一神经网络的参数。

24、根据权利要求 17-23 任一所述的装置，其特征在于，所述第一关键点单元，用于对所述第一区域内的图像进行与嘴部交互的物体的关键点识别，获得所述与嘴部交互的物体的中轴线上的至少两个中轴关键点，和/或所述与嘴部交互的物体的两条边中每条边上的至少两个边关键点。

25、根据权利要求 16-24 任一所述的装置，其特征在于，所述装置还包括：

第二关键点单元，用于基于所述第一区域内的图像获得与嘴部交互的物体上的至少两个第二关键点；

图像对齐单元，用于基于所述至少两个第二关键点对所述与嘴部交互的物体执行对齐操作，使所述与嘴部交互的物体朝向预设方向，获得包括所述朝向预设方向的与嘴部交互的物体的第二区域内的图像，所述第二区域内的图像至少包括部分所述嘴部关键点以及与嘴部交互的物体的图像；

所述吸烟识别单元，用于基于所述第二区域内的图像确定所述人脸图像中的人是否在吸烟。

26、根据权利要求 16-25 任一所述的装置，其特征在于，所述吸烟识别单元，用于利用第二神经网络基于所述第一区域内的图像确定所述人脸图像中的人是否在吸烟，所述第二神经网络经过第二样本图像训练获得。

27、根据权利要求 26 所述的装置，其特征在于，所述第二样本图像标注有图像中的人是否在吸烟的标注结果；

所述训练第二神经网络的过程包括：

将所述第二样本图像输入所述第二神经网络，获得所述第二样本图像中的人是否在吸烟的预测结果；

基于所述预测结果和所述标注结果获得第二网络损失，基于所述第二网络损失调整所述第二神经网络的参数。

28、根据权利要求 16-27 任一所述的装置，其特征在于，所述嘴部关键点单元，用于对所述人脸图像进行人脸关键点提取，获得所述人脸图像中的人脸关键点；基于所述人脸关键点获得所述嘴部关键点。

29、根据权利要求 28 所述的装置，其特征在于，所述第一区域确定单元，用于基于所述嘴部关键点确定所述人脸中嘴部的中心位置；以所述嘴部的中心位置作为所述第一区域的中心点，以设定长度为边长或半径确定所述第一区域。

30、根据权利要求 29 所述的装置，其特征在于，所述装置还包括：

眉部关键点单元，用于基于所述人脸关键点获得眉部关键点；

所述第一区域确定单元，用于以所述嘴部的中心位置作为中心点，以所述嘴部的中心位置到眉心的垂直距离作为边长或半径，确定所述第一区域，所述眉心基于所述眉部关键点确定。

31、一种电子设备，其特征在于，包括处理器，所述处理器包括权利要求 16 至 30 任意一项所述的动作识别装置。

32、一种电子设备，其特征在于，包括：存储器，用于存储可执行指令；

以及处理器，用于与所述存储器通信以执行所述可执行指令从而完成权利要求 1 至 15

任意一项所述动作识别方法的操作。

33、一种计算机可读存储介质，用于存储计算机可读取的指令，其特征在于，所述指令被执行时执行权利要求 1 至 15 任意一项所述动作识别方法的操作。

34、一种计算机程序产品，包括计算机可读代码，其特征在于，当所述计算机可读代码在设备上运行时，所述设备中的处理器执行用于实现权利要求 1 至 15 任意一项所述动作识别方法的指令。