

视频处理方法及装置、电子设备和存储介质

本公开要求在2019年7月19日提交中国专利局、申请号为201910656059.9、申请名称为“视频处理方法及装置、电子设备和存储介质”的中国专利申请的优先权，其全部内容通过引用结合在本公开中。

5 技术领域

本公开涉及计算机视觉技术领域，尤其涉及一种视频处理方法及装置、电子设备和存储介质。

背景技术

视频由多个视频帧组成，可记录动作、行为等信息，应用场景多样化。但视频不仅帧数较多，处理运算量较大，且视频还具有与时间的关联关系，例如，通过多个视频帧中的内容以及各视频帧对应的时间来表现动作或行为等信息。在相关技术中，可通过光流或3D卷积等处理获得时空特征和运动特征等。

发明内容

本公开提出了一种视频处理方法及装置、电子设备和存储介质。

根据本公开的一方面，提供了一种视频处理方法，包括：通过特征提取网络对待处理视频的多个目标视频帧进行特征提取，获得所述多个目标视频帧的特征图；通过M级动作识别网络对所述多个目标视频帧的特征图进行动作识别处理，获得所述多个目标视频帧的动作识别特征，其中，M为大于或等于1的整数，所述动作识别处理包括基于所述多个目标视频帧的特征图的时空特征提取处理，以及基于所述多个目标视频帧的特征图之间的运动差异信息的运动特征提取处理，所述动作识别特征包括时空特征信息和运动特征信息；根据所述多个目标视频帧的动作识别特征，确定所述待处理视频的分类结果。

根据本公开的实施例的视频处理方法，可通过多级动作识别网络获得目标视频帧的动作识别特征，进而获得待处理视频的分类结果，无需通过光流或3D卷积等处理进行动作识别，减小了运算量，提高了处理效率，能够对待处理视频进行在线实时分类，提高了所述视频处理方法的实用性。

在一种可能的实现方式中，通过M级动作识别网络对所述多个目标视频帧的特征图进行动作识别，获得所述多个目标视频帧的动作识别特征，包括：通过第一级动作识别网络对所述多个目标视频帧的特征图进行处理，获得第一级的动作识别特征；通过第i级动作识别网络对第i-1级的动作识别特征进行处理，获得第i级的动作识别特征，i为整数且 $1 < i < M$ ，其中，各级的动作识别特征分别与所述多个目标视频帧的特征图对应；通过第M级动作识别网络对第M-1级的动作识别特征进行处理，获得所述多个目标视频帧的动作识别特征。

在一种可能的实现方式中，通过第i级动作识别网络对第i-1级的动作识别特征进行处理，获得第i级的动作识别特征，包括：对所述第i-1级的动作识别特征进行第一卷积处理，获得第一特征信息，其中，所述第一特征信息分别与所述多个目标视频帧的特征图对应；对所述第一特征信息进行时空特征提取处理，获得时空特征信息；对所述第一特征信息进行运动特征提取处理，获得运动特征信息；至少根据所述时空特征信息和所述运动特征信息，获得所述第i级的动作识别特征。

在一种可能的实现方式中，至少根据所述时空特征信息和所述运动特征信息，获得所述第i级的动作识别特征，包括：根据所述时空特征信息、所述运动特征信息和所述第i-1级的动作识别特征，获得所述第i级的动作识别特征。

在一种可能的实现方式中，对所述第一特征信息进行时空特征提取处理，获得时空特征信息，包括：分别对与所述多个目标视频帧的特征图对应的第一特征信息进行维度重构处理，获得第二特征信息，所述第二特征信息与所述第一特征信息的维度不同；对所述第二特征信息的各通道分别进行第二卷积处理，获得第三特征信息，其中，所述第三特征信息表示所述多个目标视频帧的特征图的时间特征；对所述第三特征信息进行维度重构处理，获得第四特征信息，所述第四特征信息与所述第一特征信息的维度相同；对所述第四特征信息进行空间特征提取处理，获得所述时空特征信息。

在一种可能的实现方式中，所述第一特征信息包括多个行向量或列向量，分别对与所述多个目标视频帧的特征图对应的第一特征信息进行维度重构处理，包括：对所述第一特征信息的多个行向量或

列向量进行拼接处理,获得所述第二特征信息,其中,所述第二特征信息包括一个行向量或列向量。

通过这种方式,可获得各通道的时空信息,使得时空信息完整,并通过重构处理改变第一特征信息的维度,可通过计算量较小的方式进行卷积处理,例如,通过1D卷积处理的方式进行第二卷积处理,可简化计算,提高处理效率。

- 5 在一种可能的实现方式中,对所述第一特征信息进行运动特征提取处理,获得运动特征信息,包括:对所述第一特征信息的通道进行降维处理,获得第五特征信息,其中,所述第五特征信息分别与
10 所述待处理视频中的各目标视频帧对应;将与第 $k+1$ 个目标视频帧对应的第五特征信息进行第三卷积处理,并与所述第 k 个目标视频帧对应的第五特征信息相减,获得与第 k 个目标视频帧对应的第六特征信息,其中, k 为整数且 $1 \leq k < T$, T 为目标视频帧的数量,且 T 为大于1的整数,所述第六特征信息表示与
15 所述第 $k+1$ 个目标视频帧对应的第五特征信息及与所述第 k 个目标视频帧对应的第五特征信息之间的运动差异信息;将与所述各目标视频帧对应的第六特征信息进行特征提取处理,获得所述运动特征信息。

通过这种方式,可通过对第五特征信息进行第三卷积处理后与前一个第五特征信息相减的处理获得所述运动特征信息,可简化计算,提高处理效率。

- 15 在一种可能的实现方式中,根据所述时空特征信息、所述运动特征信息和所述第 $i-1$ 级的动作识别特征,获得所述第 i 级的动作识别特征,包括:对所述时空特征信息和所述运动特征信息进行求和处理,获得第七特征信息;对所述第七特征信息进行第四卷积处理,并与所述第 $i-1$ 级的动作识别特征进行求和处理,获得所述第 i 级的动作识别特征。

- 20 在一种可能的实现方式中,根据所述多个目标视频帧的动作识别特征,确定所述待处理视频的分类结果,包括:对各目标视频帧的动作识别特征分别进行全连接处理,获得各目标视频帧的分类信息;对各目标视频帧的分类信息进行平均处理,获得所述待处理视频的分类结果。

在一种可能的实现方式中,所述方法还包括:从待处理视频中确定出多个目标视频帧。

- 25 在一种可能的实现方式中,从待处理视频的多个视频帧中确定出多个目标视频帧,包括:将所述待处理视频划分为多个视频片段;从各视频片段中随机确定出至少一个目标视频帧,获得多个目标视频帧。

通过这种方式,可从待处理视频的多个视频帧中确定出目标视频帧,继而可对目标视频帧进行处理,可节约运算资源,提高处理效率。

- 30 在一种可能的实现方式中,所述视频处理方法通过神经网络实现,所述神经网络至少包括所述特征提取网络、所述 M 级动作识别网络,所述方法还包括:通过样本视频及所述样本视频的类别标注,对所述神经网络进行训练。

在一种可能的实现方式中,通过样本视频及所述样本视频的类别标注,对所述神经网络进行训练,包括:从所述样本视频中确定出多个样本视频帧;通过所述神经网络对所述样本视频帧进行处理,确定所述样本视频的分类结果;根据所述样本视频的分类结果及类别标注,确定所述神经网络的网络损失;根据所述网络损失调整所述神经网络的网络参数。

- 35 根据本公开的另一方面,提供了一种视频处理装置,包括:特征提取模块,用于通过特征提取网络对待处理视频的多个目标视频帧进行特征提取,获得所述多个目标视频帧的特征图;动作识别模块,用于通过 M 级动作识别网络对所述多个目标视频帧的特征图进行动作识别处理,获得所述多个目标视频帧的动作识别特征,其中, M 为大于或等于1的整数,所述动作识别处理包括基于所述多个目标视频帧的特征图的时空特征提取处理,以及基于所述多个目标视频帧的特征图之间的运动差异信息的运动特征提取处理,所述动作识别特征包括时空特征信息和运动特征信息;分类模块,用于根据所述多个目标视频帧的动作识别特征,确定所述待处理视频的分类结果。

- 40 在一种可能的实现方式中,所述动作识别模块被进一步配置为:通过第一级动作识别网络对所述多个目标视频帧的特征图进行处理,获得第一级的动作识别特征;通过第 i 级动作识别网络对第 $i-1$ 级的动作识别特征进行处理,获得第 i 级的动作识别特征, i 为整数且 $1 < i \leq M$,其中,各级的动作识别特征分别与
45 所述多个目标视频帧的特征图对应;通过第 M 级动作识别网络对第 $M-1$ 级的动作识别特征进

行处理, 获得所述多个目标视频帧的动作识别特征。

在一种可能的实现方式中, 所述动作识别模块被进一步配置为: 对所述第i-1级的动作识别特征进行第一卷积处理, 获得第一特征信息, 其中, 所述第一特征信息分别与所述多个目标视频帧的特征图对应; 对所述第一特征信息进行时空特征提取处理, 获得时空特征信息; 对所述第一特征信息进行运动特征提取处理, 获得运动特征信息; 至少根据所述时空特征信息和所述运动特征信息, 获得所述第i级的动作识别特征。

在一种可能的实现方式中, 所述动作识别模块被进一步配置为: 根据所述时空特征信息、所述运动特征信息和所述第i-1级的动作识别特征, 获得所述第i级的动作识别特征。

在一种可能的实现方式中, 所述动作识别模块被进一步配置为: 分别对与所述多个目标视频帧的特征图对应的第一特征信息进行维度重构处理, 获得第二特征信息, 所述第二特征信息与所述第一特征信息的维度不同; 对所述第二特征信息的各通道分别进行第二卷积处理, 获得第三特征信息, 其中, 所述第三特征信息表示所述多个目标视频帧的特征图的时间特征; 对所述第三特征信息进行维度重构处理, 获得第四特征信息, 所述第四特征信息与所述第一特征信息的维度相同; 对所述第四特征信息进行空间特征提取处理, 获得所述时空特征信息。

在一种可能的实现方式中, 所述第一特征信息包括多个行向量或列向量, 所述动作识别模块被进一步配置为: 对所述第一特征信息的多个行向量或列向量进行拼接处理, 获得所述第二特征信息, 其中, 所述第二特征信息包括一个行向量或列向量。

在一种可能的实现方式中, 所述动作识别模块被进一步配置为: 对所述第一特征信息的通道进行降维处理, 获得第五特征信息, 其中, 所述第五特征信息分别与所述待处理视频中的各目标视频帧对应; 将与第k+1个目标视频帧对应的第五特征信息进行第三卷积处理, 并与所述第k个目标视频帧对应的第五特征信息相减, 获得与第k个目标视频帧对应的第六特征信息, 其中, k为整数且 $1 \leq k < T$, T为目标视频帧的数量, 且T为大于1的整数, 所述第六特征信息表示与所述第k+1个目标视频帧对应的第五特征信息及与所述第k个目标视频帧对应的第五特征信息之间的运动差异信息; 将与所述各目标视频帧对应的第六特征信息进行特征提取处理, 获得所述运动特征信息。

在一种可能的实现方式中, 所述动作识别模块被进一步配置为: 对所述时空特征信息和所述运动特征信息进行求和处理, 获得第七特征信息; 对所述第七特征信息进行第四卷积处理, 并与所述第i-1级的动作识别特征进行求和处理, 获得所述第i级的动作识别特征。

在一种可能的实现方式中, 所述分类模块被进一步配置为: 对各目标视频帧的动作识别特征分别进行全连接处理, 获得各目标视频帧的分类信息; 对各目标视频帧的分类信息进行平均处理, 获得所述待处理视频的分类结果。

在一种可能的实现方式中, 所述装置还包括: 确定模块, 用于从待处理视频中确定出多个目标视频帧。

在一种可能的实现方式中, 所述确定模块被进一步配置为: 将所述待处理视频划分为多个视频片段; 从各视频片段中随机确定出至少一个目标视频帧, 获得多个目标视频帧。

在一种可能的实现方式中, 所述视频处理方法通过神经网络实现, 所述神经网络至少包括所述特征提取网络、所述M级动作识别网络, 所述装置还包括: 训练模块, 用于通过样本视频及所述样本视频的分类标注, 对所述神经网络进行训练。

在一种可能的实现方式中, 所述训练模块被进一步配置为: 从所述样本视频中确定出多个样本视频帧; 通过所述神经网络对所述样本视频帧进行处理, 确定所述样本视频的分类结果; 根据所述样本视频的分类结果及类别标注, 确定所述神经网络的网络损失; 根据所述网络损失调整所述神经网络的网络参数。

根据本公开的一方面, 提供了一种电子设备, 包括:

处理器; 用于存储处理器可执行指令的存储器; 其中, 所述处理器被配置为: 执行上述视频处理方法。

根据本公开的一方面, 提供了一种计算机可读存储介质, 其上存储有计算机程序指令, 所述计

算机程序指令被处理器执行时实现上述视频处理方法。

根据本公开的一方面，提供了一种计算机程序，包括计算机可读代码，当所述计算机可读代码在电子设备中运行时，所述电子设备中的处理器执行用于执行上述的视频处理方法。

应当理解的是，以上的一般描述和后文的细节描述仅是示例性和解释性的，而非限制本公开。

5 根据下面参考附图对示例性实施例的详细说明，本公开的其它特征及方面将变得清楚。

附图说明

此处的附图被并入说明书中并构成本说明书的一部分，这些附图示出了符合本公开的实施例，并与说明书一起用于说明本公开的技术方案。

- 10 图1示出根据本公开实施例的视频处理方法的流程图；
图2示出根据本公开实施例的视频处理方法的流程图；
图3示出根据本公开实施例的动作识别网络的示意图；
图4示出根据本公开实施例的时空特征提取处理的示意图；
图5示出根据本公开实施例的运动特征提取处理的示意图；
图6示出根据本公开实施例的视频处理方法的流程图；
15 图7示出根据本公开实施例的视频处理方法的应用示意图；
图8示出根据本公开实施例的视频处理装置的框图；
图9示出根据本公开实施例的视频处理装置的框图；
图10示出根据本公开实施例的电子设备的框图；
图11示出根据本公开实施例的电子设备的框图。

20 具体实施方式

以下将参考附图详细说明本公开的各种示例性实施例、特征和方面。附图中相同的附图标记表示功能相同或相似的元件。尽管在附图中示出了实施例的各种方面，但是除非特别指出，不必按比例绘制附图。

25 在这里专用的词“示例性”意为“用作例子、实施例或说明性”。这里作为“示例性”所说明的任何实施例不必解释为优于或好于其它实施例。

本文中术语“和/或”，仅仅是一种描述关联对象的关联关系，表示可以存在三种关系，例如，A和/或B，可以表示：单独存在A，同时存在A和B，单独存在B这三种情况。另外，本文中术语“至少一种”表示多种中的任意一种或多种中的至少两种的任意组合，例如，包括A、B、C中的至少一种，可以表示包括从A、B和C构成的集合中选择的任意一个或多个元素。

30 另外，为了更好地说明本公开，在下文的具体实施方式中给出了众多的具体细节。本领域技术人员应当理解，没有某些具体细节，本公开同样可以实施。在一些实例中，对于本领域技术人员熟知的方法、手段、元件和电路未作详细描述，以便于凸显本公开的主旨。

图1示出根据本公开实施例的视频处理方法的流程图，如图1所示，所述方法包括：

35 在步骤S11中，通过特征提取网络对待处理视频的多个目标视频帧进行特征提取，获得所述多个目标视频帧的特征图；

在步骤S12中，通过M级动作识别网络对所述多个目标视频帧的特征图进行动作识别处理，获得所述多个目标视频帧的动作识别特征，其中，M为大于或等于1的整数，所述动作识别处理包括基于所述多个目标视频帧的特征图的时空特征提取处理，以及基于所述多个目标视频帧的特征图之间的运动差异信息的运动特征提取处理，所述动作识别特征包括时空特征信息和运动特征信息；

40 在步骤S13中，根据所述多个目标视频帧的动作识别特征，确定所述待处理视频的分类结果。

根据本公开的实施例的视频处理方法，可通过多级动作识别网络获得目标视频帧的动作识别特征，进而获得待处理视频的分类结果，无需通过光流或3D卷积等处理进行动作识别，减小了运算量，提高了处理效率，能够对待处理视频进行在线实时分类，提高了所述视频处理方法的实用性。

45 在一种可能的实现方式中，所述方法可以由终端设备执行，终端设备可以为用户设备（User Equipment, UE）、移动设备、用户终端、终端、蜂窝电话、无绳电话、个人数字处理（Personal Digital

Assistant, PDA)、手持设备、计算设备、车载设备、可穿戴设备等,所述方法可以通过处理器调用存储器中存储的计算机可读指令的方式来实现。或者,所述方法通过服务器执行。

在一种可能的实现方式中,所述待处理视频可以是任意视频获取装置拍摄的视频,所述待处理视频帧可包括一个或多个目标对象(例如,人、车辆和/或茶杯等物品),所述目标对象可正在进行某个动作(例如,拿起水杯、走路等),本公开对待处理视频的内容不做限制。

图2示出根据本公开实施例的视频处理方法的流程图,如图2所示,所述方法包括:

在步骤S14中,从待处理视频中确定出多个目标视频帧。

在一种可能的实现方式中,步骤S14可包括:将所述待处理视频划分为多个视频片段;从各视频片段中随机确定出至少一个目标视频帧,获得多个目标视频帧。

在示例中,待处理视频可包括多个视频帧,可将待处理视频进行划分,例如,可划分为T个视频片段(T为大于1的整数),并可在每个视频片段的多个视频帧中进行抽样,例如,在每个视频片段中抽样出至少一个目标视频帧。例如,可将待处理视频进行等间隔划分,例如可划分为8或16段,并在每个视频片段中进行随机抽样,例如,可在每个视频片段中随机选取1个视频帧作为目标视频帧,即可获得多个目标视频帧。

在示例中,可在待处理视频的所有视频帧中进行随机抽样,获得多个目标视频帧。或者,可等间距地选取多个视频帧,作为目标视频帧,例如,选择第1个视频帧、第11个视频帧、第21个视频帧...又或者,可将待处理视频的所有视频帧均确定为目标视频帧,本公开对选取目标视频帧的方式不做限制。

通过这种方式,可从待处理视频的多个视频帧中确定出目标视频帧,继而可对目标视频帧进行处理,可节约运算资源,提高处理效率。

在一种可能的实现方式中,在步骤S11中,可对待处理视频的多个目标视频帧进行特征提取,获得所述多个目标视频帧的特征图。可通过神经网络的特征提取网络来进行所述特征提取处理,所述特征提取网络可以是所述神经网络的一部分(例如,子网络或某个层级的神经网络),在示例中,所述特征提取网络可包括一个或多个卷积层,可对多个目标视频帧进行特征提取,获得多个目标视频帧的特征图。

在示例中,可通过特征提取网络对T(T为大于1的整数)个目标视频帧进行特征提取处理,每个目标视频帧可分为C(C为正整数)个通道输入至特征提取网络,例如,目标视频帧为RGB图像,可分别通过R、G和B三个通道输入特征提取网络。每个目标视频帧的尺寸为H×W(H为图像的高度,可表示为图像在高度方向的像素点的数量,W为图像的宽度,可表示为图像在宽度方向的像素点的数量),因此,输入特征提取网络的目标视频帧的维度为T×C×H×W。例如,T可以是16,C可以是3,H和W可均为224,则输入特征提取网络的目标视频帧的维度为16×3×224×224。

在示例中,所述神经网络可对多个待处理视频进行批处理,例如,特征提取网络可对N个待处理视频的目标视频帧进行特征提取处理,输入特征提取网络的目标视频帧的目标视频帧维度的为N×T×C×H×W。

在示例中,特征提取网络可对维度为T×C×H×W的目标视频帧进行特征提取处理,获得与T个目标视频帧分别对应的T组特征图。例如,在特征提取处理中,目标视频帧的特征图尺寸可小于目标视频帧,但通道数可多于目标视频帧,可增大对目标视频帧的感受野,即,C的值可增大,H和W的值可减小。例如,输入特征提取网络的目标视频帧的维度为16×3×224×224,目标视频帧的通道数可扩大16倍,即,C的值可增大至48,目标视频帧的特征图尺寸可缩小4倍,即H和W的值均可缩小至56,与各目标视频帧对应的特征图的通道数量为48,每个特征图的尺寸为56×56,特征图的维度可以是16×48×56×56。以上数据仅为示例,本公开对目标视频帧与特征图的维度不做限制。

在一种可能的实现方式中,在步骤S12中,可对T个目标视频帧的特征图进行动作识别,分别获得各目标视频帧的动作识别特征。可通过神经网络的M级动作识别网络对所述多个目标视频帧的特征图进行动作识别处理,所述M级动作识别网络可以是级联的M个动作识别网络,每个动作识别网络可以是所述神经网络的一部分。

在一种可能的实现方式中，步骤S12可包括：通过第一级动作识别网络对所述多个目标视频帧的特征图进行处理，获得第一级的动作识别特征；通过第i级动作识别网络对第i-1级的动作识别特征进行处理，获得第i级的动作识别特征，i为整数且 $1 < i < M$ ，其中，各级的动作识别特征分别与所述多个目标视频帧的特征图对应；通过第M级动作识别网络对第M-1级的动作识别特征进行处理，获得所述多个目标视频帧的动作识别特征。

在一种可能的实现方式中，M级动作识别网络是级联的，每一级动作识别网络的输出信息（即，该级动作识别网络的动作识别特征）可作为下一级动作识别网络的输入信息。第一级动作识别网络可对目标视频帧的特征图进行处理，并输出第一级的动作识别特征，第一级的动作识别特征可作为第二级动作识别特征的输入信息，即，第二级动作识别网络可对第一级动作识别特征进行处理，获得第二级动作识别特征，并可将第二级动作识别特征作为第三级动作识别网络的输入信息……

在一种可能的实现方式中，以第i级动作识别网络为例，第i级动作识别网络可将第i-1级的动作识别特征作为输入信息进行处理，通过第i级动作识别网络对第i-1级的动作识别特征进行处理，获得第i级的动作识别特征，包括：对所述第i-1级的动作识别特征进行第一卷积处理，获得第一特征信息；对所述第一特征信息进行时空特征提取处理，获得时空特征信息；对所述第一特征信息进行运动特征提取处理，获得运动特征信息；至少根据所述时空特征信息和所述运动特征信息，获得所述第i级的动作识别特征。

图3示出根据本公开实施例的动作识别网络的示意图，第一级动作识别网络至第M级动作识别网络的结构均如图3所示。以第i级动作识别网络为例，第i级动作识别网络可将第i-1级的动作识别特征作为输入信息进行处理。在示例中，第i级动作识别网络可通过卷积核为 1×1 的2D卷积层对第i-1级的动作识别特征进行第一卷积处理，可对第i-1级的动作识别特征进行降维，在示例中，所述卷积核为 1×1 的2D卷积层可降低第i-1级的动作识别特征的通道数，例如，可将通道数C的数量减少16倍，获得所述第一特征信息。本公开对降低的倍数不做限制。

在示例中，在第一级动作识别网络中，第一级动作识别网络可将目标视频帧的特征图作为输入信息进行处理。第一级动作识别网络可通过卷积核为 1×1 的2D卷积层对所述目标视频帧的特征图进行第一卷积处理，可对特征图进行降维，获得第一特征信息。

在一种可能的实现方式中，所述第i级动作识别网络可将第一特征信息分别进行时空特征提取处理和运动特征提取处理，可将第一特征信息通过两个分支（时空特征提取分支和运动特征提取分支）分别进行处理，分别获得时空特征信息和运动特征信息。

在一种可能的实现方式中，根据时空特征信息、运动特征信息和第i-1级的动作识别特征，获得第i级的动作识别特征，可包括：根据所述时空特征信息、所述运动特征信息和所述第i-1级的动作识别特征，获得所述第i级的动作识别特征。例如，可对时空特征信息和运动特征信息进行求和，并对求和结果进行卷积处理，进一步地，可将卷积处理结果与第i-1级的动作识别特征进行求和，获得第i级的动作识别特征。

图4示出根据本公开实施例的时空特征提取处理的示意图，对所述第一特征信息进行时空特征提取处理，获得时空特征信息，包括：分别对与所述多个目标视频帧的特征图对应的第一特征信息进行维度重构处理，获得第二特征信息，所述第二特征信息与所述第一特征信息的维度不同；对所述第二特征信息的各通道分别进行第二卷积处理，获得第三特征信息，其中，所述第三特征信息表示所述多个目标视频帧的特征图的时间特征；对所述第三特征信息进行维度重构处理，获得第四特征信息，所述第四特征信息与所述第一特征信息的维度相同；对所述第四特征信息进行空间特征提取处理，获得所述时空特征信息。

在一种可能的实现方式中，第一特征信息的维度为 $T \times C \times H \times W$ ，其中，参数C、H和W的取值与目标视频帧的特征图可不同，所述第一特征信息可通过特征矩阵来表示，所述特征矩阵可表示为多个行向量或列向量。所述第一特征信息包括多个行向量或列向量，分别对与所述多个目标视频帧的特征图对应的第一特征信息进行维度重构处理，包括：对所述第一特征信息的多个行向量或列向量进行拼接处理，获得所述第二特征信息，其中，所述第二特征信息包括一个行向量或列向量。可对第一特

征信息（特征矩阵）进行重构处理，将所述特征矩阵的维度变换为 $HW \times C \times T$ ，获得与第一特征信息维度不同的第二特征信息，例如，第一特征信息包括T组特征矩阵，每组特征矩阵的通道数量为C（例如，每组特征矩阵的数量为C），每个特征矩阵的尺寸为 $H \times W$ ，可对每个特征矩阵分别进行拼接，例如，可将特征矩阵看作H个行向量或W个列向量，并对H个行向量或W个列向量进行拼接，形成一个行向量或一个列向量，所述行向量或列向量即为所述第二特征信息，HW的值可等于H和W的乘积。
5 本公开对重构处理的方式不做限制。

在一种可能的实现方式中，可对第二特征信息的各通道分别进行第二卷积处理，获得第三特征信息。在示例中，可通过卷积核为 3×1 的1D深度分离卷积层对第二特征信息的各通道分别进行第二卷积处理。例如，T组第二特征信息各包括C个通道，例如，每组第二特征信息的数量为C，可分别对每组的C个第二特征信息进行第二卷积处理，获得T组第三特征信息，T组第三特征信息可表示所述多个目标视频帧的特征图的时间特征，即，所述第三特征信息具有各目标视频帧的时间信息。在示例中，每个通道的第二特征信息包含的时空信息可互不相同，分别对每个通道的第二特征信息进行第二卷积处理，可获取每个通道的第三特征信息，且通过卷积核为 3×1 的1D卷积层分别对每个通道对重构后获得的第二特征信息进行第二卷积处理的运算量较小，即，对行向量或列向量进行1D卷积处理，相对于对特征图进行2D卷积或3D卷积的运算量较小，可提升处理效率。在示例中，第三特征信息的维度为 $HW \times C \times T$ ，即，每个第三特征信息可以是行向量或列向量。
10
15

在一种可能的实现方式中，可对第三特征信息进行重构，例如，可将每个第三特征信息（行向量或列向量形式）重构为矩阵，获得第四特征信息，第四特征信息的维度与第一特征信息的维度相同，例如，每个第三特征信息均为长度为HW的行向量或列向量，可将第三特征信息分为W个长度为H的列向量或者H个长度为W的行向量，并将所述行向量或列向量进行合并，获得特征矩阵（即，第四特征信息），第四特征信息的维度为 $T \times C \times H \times W$ 。本公开对第四特征信息的参数不做限制。
20

在一种可能的实现方式中，可通过卷积核为 3×3 的2D卷积层对第四特征信息进行卷积处理，可提取第四特征信息的空间特征，获得时空特征信息，即，提取表示第四特征信息中目标对象的位置的特征信息，并与时间信息融合，可表示所述时空特征信息。所述时空特征信息可为维度为 $T \times C \times H \times W$ 的特征矩阵，时空特征信息的H和W与第四特征信息可不同。
25

通过这种方式，可获得各通道的时空信息，使得时空信息完整，并通过重构处理改变第一特征信息的维度，可通过计算量较小的方式进行卷积处理，例如，通过1D卷积处理的方式进行第二卷积处理，可简化计算，提高处理效率。

图5示出根据本公开实施例的运动特征提取处理的示意图，对所述第一特征信息进行运动特征提取处理，获得运动特征信息，可包括：对所述第一特征信息的通道进行降维处理，获得第五特征信息，其中，所述第五特征信息分别与所述待处理视频中的各目标视频帧对应；将与第k+1个目标视频帧对应的第五特征信息进行第三卷积处理，并与所述第k个目标视频帧对应的第五特征信息相减，获得与第k个目标视频帧对应的第六特征信息，其中，k为整数且 $1 \leq k < T$ ，T为目标视频帧的数量，且T为大于1的整数，所述第六特征信息表示与所述第k+1个目标视频帧对应的第五特征信息及与所述第k个目标视频帧对应的第五特征信息之间的运动差异信息；将与所述各目标视频帧对应的第六特征信息进行特征提取处理，获得所述运动特征信息。
30
35

在一种可能的实现方式中，可对第一特征信息的通道进行降维处理，获得第五特征信息，例如，可通过卷积核为 1×1 的2D卷积层对第一特征信息的通道进行降维处理，即，可降低通道的数量。在示例中，可将维度为 $T \times C \times H \times W$ 的第一特征信息的通道数C降为C/16。获得与各目标视频帧对应的第五特征信息，所述第五特征信息的维度为 $T \times C/16 \times H \times W$ ，即，包括分别与T个目标视频帧对应的T组第五特征信息，每组第五特征信息的维度为 $C/16 \times H \times W$ 。
40

在一种可能的实现方式中，以与第k个目标视频帧对应的第五特征信息（简称为第五特征信息k）为例，可将与第k+1个目标视频帧对应的第五特征信息（简称为第五特征信息k+1）进行的各通道第三卷积处理，例如，可通过卷积核为 3×3 的2D深度分离卷积层对第五特征信息k+1进行第三卷积处理，并将第三卷积处理获得的结果与第五特征信息k相减，获得与第k个目标视频帧对应的第六特征信息，
45

第六特征信息的维度与第五特征信息相同，为 $C/16 \times H \times W$ 。可分别对每个第五特征信息进行第三卷积处理，并与前一个第五特征信息相减，获得第六特征信息，第六特征信息可表示相邻的两个目标视频帧对应的第五特征信息之间的运动差异信息，即，可用于表示两个目标视频帧中的目标对象的动作差异，以确定目标对象的动作。在示例中，所述相减的过程可获得 $T-1$ 个第六特征信息，可将与第 T 个目标视频帧对应的第五特征信息与参数全为0的矩阵经过第三卷积处理后的处理结果相减或者与参数全为0的矩阵直接相减，获得与第 T 个目标视频帧对应的第六特征信息，或者，可将全为0的矩阵作为第六特征信息，获得与第 T 个目标视频帧对应的第六特征信息，即，共可获得与 T 个目标视频帧分别对应的 T 个第六特征信息。进一步地，可将 T 个第六特征信息进行合并，即可获得维度为 $T \times C/16 \times H \times W$ 的第六特征信息。

在一种可能的实现方式中，可对维度为 $T \times C/16 \times H \times W$ 的第六特征信息进行特征提取处理，例如，可通过卷积核为 1×1 的2D卷积层对第六特征信息进行升维，例如，可对通道数进行升维，将通道数 $C/16$ 升维至 C ，获得所述运动特征信息，运动特征信息的维度与时空特征信息的维度一致，均为 $T \times C \times H \times W$ 。

在一种可能的实现方式中，如图3所示，可根据所述时空特征信息、所述运动特征信息和所述第 $i-1$ 级的动作识别特征，获得第 i 级的动作识别特征。在示例中，该步骤可包括：对所述时空特征信息和所述运动特征信息进行求和处理，获得第七特征信息；对所述第七特征信息进行第四卷积处理，并与所述第 $i-1$ 级的动作识别特征进行求和处理，获得所述第 i 级的动作识别特征。

在一种可能的实现方式中，时空特征信息和运动特征信息的维度相同，均为 $T \times C \times H \times W$ ，可将时空特征信息和运动特征信息的多个特征信息（例如，各特征图或特征矩阵）分别求和，获得第七特征信息，第七特征信息的维度为 $T \times C \times H \times W$ 。

在一种可能的实现方式中，可对第七特征信息进行第四卷积处理，例如，可通过卷积核为 1×1 的2D卷积层对第七特征信息进行第四卷积处理，可对第七特征信息进行升维，将第七特征信息的维度变换为与第 $i-1$ 级的动作识别特征相同的维度，例如，可将通道数增大16倍。进一步地，可将第四卷积处理的处理结果与第 $i-1$ 级的动作识别特征进行求和，获得第 i 级动作识别特征。

在一种可能的实现方式中，第一级动作识别网络可将目标视频帧的特征图与第四卷积处理的处理结果进行求和，获得第一级动作识别特征，第一级动作识别特征可作为第二级动作识别网络的输入信息。

通过这种方式，可通过对第五特征信息进行第三卷积处理后与前一个第五特征信息相减的处理获得所述运动特征信息，可简化计算，提高处理效率。

在一种可能的实现方式中，可通过上述方式逐级获得动作识别特征，并可通过上述方式通过第 M 级动作识别网络对第 $M-1$ 级的动作识别特征进行处理，获得所述多个目标视频帧的动作识别特征，即，将第 M 级动作识别特征作为所述目标视频帧的动作识别特征。

在一种可能的实现方式中，在步骤S13中，可根据多个目标视频帧的动作识别特征，获得待处理视频帧的分类结果。步骤S13可包括：对各目标视频帧的动作识别特征分别进行全连接处理，获得各目标视频帧的分类信息；对各目标视频帧的分类信息进行平均处理，获得所述待处理视频的分类结果。

在一种可能的实现方式中，可通过所述神经网络的全连接层对各目标视频帧的动作识别特征进行全连接处理，获得各目标视频帧的分类信息，在示例中，各目标视频帧的分类信息可以是特征向量，即，所述全连接层可输出 T 个特征向量。进一步地，可对所述 T 个特征向量进行平均处理，获得待处理视频的分类结果。所述分类结果也可以是特征向量，可表示所述待处理视频所属类别的概率。

在示例中，所述分类结果可以是400维的向量，该向量包括400个参数，分别表示待处理视频属于400种类别的概率。所述类别可以是待处理视频中的目标对象的动作的类别，例如，行走、举杯、饮食等动作。例如，在该向量中，第2个参数的值最大，即，待处理视频属于第2个种类的概率最大，可判断待处理视频属于第2个种类，例如，可判断待处理视频中的目标对象正在行走。本公开对分类结果的类型和维度不做限制。

根据本公开的实施例的视频处理方法，可从待处理视频的多个视频帧中确定出目标视频帧，继而可对目标视频帧进行处理，可节约运算资源，提高处理效率。每级动作识别网络可获得各通道的时空

信息,使得时空信息完整,并通过重构处理改变第一特征信息的维度,可通过计算量较小的方式进行卷积处理,还可通过对第五特征信息进行第三卷积处理后与前一个第五特征信息相减的处理获得所述运动特征信息,可简化计算。进一步地,可获得每级动作识别网络的动作识别结果,进而获得待处理视频的分类结果,无需通过光流或3D卷积等处理进行动作识别,通过输入的目标视频帧(RGB图像)即可获得时空特征信息和运动特征信息,减少了输入参数,并减小了运算量,提高了处理效率,能够对待处理视频进行在线实时分类,提高了所述视频处理方法的实用性。

在一种可能的实现方式中,所述视频处理方法可通过神经网络实现,所述神经网络至少包括所述特征提取网络、所述M级动作识别网络。所述神经网络还可包括所述全连接层,以对动作识别特征进行全连接处理。

图6示出根据本公开实施例的视频处理方法的流程图,如图6所示,所述方法还包括:

在步骤S15中,通过样本视频及所述样本视频的类别标注,对所述神经网络进行训练。

在一种可能的实现方式中,步骤S15可包括:从所述样本视频中确定出多个样本视频帧;通过所述神经网络对所述样本视频帧进行处理,确定所述样本视频的分类结果;根据所述样本视频的分类结果及类别标注,确定所述神经网络的网络损失;根据所述网络损失调整所述神经网络的网络参数。

在一种可能的实现方式中,样本视频可包括多个视频帧,可从所述样本视频的多个视频帧中确定出样本视频帧,例如,可进行随机抽样或将样本视频划分为多个视频片段,并在每个视频片段中进行抽样,获得所述样本视频帧。

在一种可能的实现方式中,可将样本视频帧输入所述神经网络,并由所述特征提取网络进行特征提取处理,并由M级动作识别网络进行动作识别处理,进一步地,由全连接层进行全连接处理后,可获得各样本视频帧的分类信息,对各样本视频帧的分类信息进行平均处理,获得所述样本视频的分类结果。

在一种可能的实现方式中,所述分类结果可以是表示样本视频的分类的多维向量(可能具有误差)。所述样本视频可具有类别标注,可表示样本视频的实际类别(无误差)。可根据分类结果和类别标注确定神经网络的网络损失,例如,可确定分类结果和类别标注之间的余弦距离或欧氏距离,并根据所述余弦距离或欧氏距离与0之间的差异确定网络损失。本公开对网络损失的确定方式不做限制。

在一种可能的实现方式中,可根据所述网络损失调整神经网络的网络参数,例如,可确定所述网络损失对于神经网络的各参数的梯度,并按照使网络损失最小化的方向,通过梯度下降法来调节各网络参数。可通过上述方式对网络参数进行多次调节(即,通过多个样本视频进行多个训练周期的训练),并在满足训练条件时,获得训练后的神经网络。所述训练条件可包括训练次数(即,训练周期数),例如,当训练次数达到预设次数时,满足训练条件。或者,所述训练条件可包括网络损失的大小或收敛性,例如,当所述网络损失小于或等于损失阈值或收敛于预设区间内时,满足训练条件。本公开对训练条件不做限制。

图7示出根据本公开实施例的视频处理方法的应用示意图。如图6所示,待处理视频可以是包括一个或多个目标对象的任意视频,可通过采样等方式在待处理视频的多个视频帧中确定出T个目标视频帧。例如,可将待处理视频划分为T(例如,T为8或16)个视频片段,并在每个视频片段中随机采样出一个视频帧作为目标视频帧。

在一种可能的实现方式中,可通过神经网络的特征提取网络对多个目标视频帧进行特征提取,特征提取网络可包括一个或多个卷积层,可对多个目标视频帧进行卷积处理,获得多个目标视频帧的特征图。例如,在T个目标视频帧中,每个目标视频帧可分为C个通道(例如,R、G和B三个通道)输入至特征提取网络,目标视频帧的尺寸为H×W(例如,224×224),在特征提取处理后,C、H和W的数值均可发生变化。

在一种可能的实现方式中,可通过M级动作识别网络对特征图进行处理,M级动作识别网络可以是级联的M个动作识别网络,每个动作识别网络的网络结构相同,且均为所述神经网络的一部分。如图6所示,M级动作识别网络可组成多个组,每个组之间可具有卷积层或激活层等神经网络层级,或者,组之间可不具有神经网络层级,各组动作识别网络可直接级联,各组动作识别网络的数量总和为

M。

在一种可能的实现方式中，第一级动作识别网络可对T组特征图进行处理，获得第一级的动作识别特征，第一级动作识别特征可作为第二级动作识别网络的输入信息，第二级动作识别网络可对第一级动作识别特征进行处理，获得第二级动作识别特征，并可将第二级动作识别特征作为第三级动作识别网络的输入信息.....

在一种可能的实现方式中，以第i级动作识别网络为例，第i级动作识别网络可将第i-1级的动作识别特征作为输入信息进行处理，可通过卷积核为 1×1 的2D卷积层对第i-1级的动作识别特征进行第一卷积处理，可对第i-1级的动作识别特征进行降维，获得第一特征信息。

在一种可能的实现方式中，第i级动作识别网络可将第一特征信息分别进行时空特征提取处理和运动特征提取处理，例如，可分为时空特征提取分支和运动特征提取分支分别进行处理。

在一种可能的实现方式中，所述时空特征提取分支可首先对第一特征信息进行重构，例如，可将第一特征信息的特征矩阵重构为行向量或列向量，获得第二特征信息，并通过卷积核为 3×1 的1D卷积层对第二特征信息的各通道分别进行第二卷积处理，在运算量较小的情况下获得第三特征信息。进一步地，可对第三特征信息进行重构，获得矩阵形式的第四特征信息，并可通过卷积核为 3×3 的2D卷积层对第四特征信息进行卷积处理，获得所述时空特征信息。

在一种可能的实现方式中，所述运动特征提取分支可首先通过卷积核为 1×1 的2D卷积层对第一特征信息的通道进行降维处理，例如，可将第一特征信息的通道数C降为C/16，获得与各目标视频帧对应的第五特征信息。以与第k个目标视频帧对应的第五特征信息为例，可通过卷积核为 3×3 的2D卷积层对与第k+1个目标视频帧对应的第五特征信息进行的各通道第三卷积处理，并将第三卷积处理获得的结果与第五特征信息k相减，获得与第k个目标视频帧对应的第六特征信息，可通过上述方式，获得与前T-1个目标视频帧对应的第六特征信息，并可将与第T个目标视频帧对应的第五特征信息与参数全为0的矩阵经过第三卷积处理后的处理结果相减，获得与第T个目标视频帧对应的第六特征信息，即，可获得T个第六特征信息。进一步地，可将T个第六特征信息进行合并，并通过卷积核为 1×1 的2D卷积层对第六特征信息进行升维，获得运动特征信息。

在一种可能的实现方式中，可将时空特征信息和运动特征信息进行求和处理，获得第七特征信息，并通过卷积核为 1×1 的2D卷积层对第七特征信息进行第四卷积处理，可对第七特征信息进行升维，将第七特征信息的维度变换为与第i-1级的动作识别特征相同的维度，并与第i-1级的动作识别特征进行求和，获得第i级动作识别特征。

在一种可能的实现方式中，可将第M级动作识别网络输出的动作识别特征确定为目标视频帧的动作识别特征，并将目标视频帧的动作识别特征输入神经网络的全连接层进行处理，获得与各目标视频帧对应的分类信息，例如，分类信息1，分类信息2.....在示例中，所述分类信息可以是向量，可将T个目标视频帧对应的分类信息进行平均处理，获得待处理视频的分类结果。所述分类结果也为向量，可表示待处理视频所属类别的概率。例如，分类结果可以是400维的向量，该向量包括400个参数，分别表示待处理视频属于400种类别的概率。所述类别可以是待处理视频中的目标对象的动作的类别，例如，行走，举杯、饮食等动作。例如，在该向量中，第2个参数的值最大，表示待处理视频属于第2个种类的概率最大，可判断待处理视频属于第2个种类。

在一种可能的实现方式中，所述视频处理方法可通过时空特征信息和动作特征信息识别相似的动作，例如，关门和开门的动作，日落和日出的动作等，且所述视频处理方法运算量较小，处理效率高，可用于对视频进行实时分类中，例如，可用于监狱的监控，可实时判别犯罪嫌疑人是否有越狱动作；可用于地铁的监控，可实时判别地铁车辆的运行状态可客流的状态；可用于安防领域，可实时判别监控区域是否有人进行危险的动作等。本公开对所述视频处理方法的应用领域不做限制。

可以理解，本公开提及的上述各个方法实施例，在不违背原理逻辑的情况下，均可以彼此相互结合形成结合后的实施例，限于篇幅，本公开不再赘述。

图8示出根据本公开实施例的视频处理装置的框图，如图8所示，所述视频处理装置包括：

特征提取模块11，用于通过特征提取网络对待处理视频的多个目标视频帧进行特征提取，获得所

述多个目标视频帧的特征图；

动作识别模块12，用于通过M级动作识别网络对所述多个目标视频帧的特征图进行动作识别处理，获得所述多个目标视频帧的动作识别特征，其中，M为大于或等于1的整数，所述动作识别处理包括基于所述多个目标视频帧的特征图的时空特征提取处理，以及基于所述多个目标视频帧的特征图之间的运动差异信息的运动特征提取处理，所述动作识别特征包括时空特征信息和运动特征信息；

分类模块13，用于根据所述多个目标视频帧的动作识别特征，确定所述待处理视频的分类结果。

在一种可能的实现方式中，所述动作识别模块被进一步配置为：通过第一级动作识别网络对所述多个目标视频帧的特征图进行处理，获得第一级的动作识别特征；通过第i级动作识别网络对第i-1级的动作识别特征进行处理，获得第i级的动作识别特征，i为整数且 $1 < i < M$ ，其中，各级的动作识别特征分别与所述多个目标视频帧的特征图对应；通过第M级动作识别网络对第M-1级的动作识别特征进行处理，获得所述多个目标视频帧的动作识别特征。

在一种可能的实现方式中，所述动作识别模块被进一步配置为：对所述第i-1级的动作识别特征进行第一卷积处理，获得第一特征信息，其中，所述第一特征信息分别与所述多个目标视频帧的特征图对应；对所述第一特征信息进行时空特征提取处理，获得时空特征信息；对所述第一特征信息进行运动特征提取处理，获得运动特征信息；至少根据所述时空特征信息和所述运动特征信息，获得所述第i级的动作识别特征。

在一种可能的实现方式中，所述动作识别模块被进一步配置为：根据所述时空特征信息、所述运动特征信息和所述第i-1级的动作识别特征，获得所述第i级的动作识别特征。

在一种可能的实现方式中，所述动作识别模块被进一步配置为：分别对与所述多个目标视频帧的特征图对应的第一特征信息进行维度重构处理，获得第二特征信息，所述第二特征信息与所述第一特征信息的维度不同；对所述第二特征信息的各通道分别进行第二卷积处理，获得第三特征信息，其中，所述第三特征信息表示所述多个目标视频帧的特征图的时间特征；对所述第三特征信息进行维度重构处理，获得第四特征信息，所述第四特征信息与所述第一特征信息的维度相同；对所述第四特征信息进行空间特征提取处理，获得所述时空特征信息。

在一种可能的实现方式中，所述第一特征信息包括多个行向量或列向量，所述动作识别模块被进一步配置为：对所述第一特征信息的多个行向量或列向量进行拼接处理，获得所述第二特征信息，其中，所述第二特征信息包括一个行向量或列向量。

在一种可能的实现方式中，所述动作识别模块被进一步配置为：对所述第一特征信息的通道进行降维处理，获得第五特征信息，其中，所述第五特征信息分别与所述待处理视频中的各目标视频帧对应；将与第k+1个目标视频帧对应的第五特征信息进行第三卷积处理，并与所述第k个目标视频帧对应的第五特征信息相减，获得与第k个目标视频帧对应的第六特征信息，其中，k为整数且 $1 \leq k < T$ ，T为目标视频帧的数量，且T为大于1的整数，所述第六特征信息表示与所述第k+1个目标视频帧对应的第五特征信息及与所述第k个目标视频帧对应的第五特征信息之间的运动差异信息；将与所述各目标视频帧对应的第六特征信息进行特征提取处理，获得所述运动特征信息。

在一种可能的实现方式中，所述动作识别模块被进一步配置为：对所述时空特征信息和所述运动特征信息进行求和处理，获得第七特征信息；对所述第七特征信息进行第四卷积处理，并与所述第i-1级的动作识别特征进行求和处理，获得所述第i级的动作识别特征。

在一种可能的实现方式中，所述分类模块被进一步配置为：对各目标视频帧的动作识别特征分别进行全连接处理，获得各目标视频帧的分类信息；对各目标视频帧的分类信息进行平均处理，获得所述待处理视频的分类结果。

图9示出根据本公开实施例的视频处理装置的框图，如图9所示，所述视频处理装置还包括：

确定模块14，用于从待处理视频中确定出多个目标视频帧。

在一种可能的实现方式中，所述确定模块被进一步配置为：将所述待处理视频划分为多个视频片段；从各视频片段中随机确定出至少一个目标视频帧，获得多个目标视频帧。

在一种可能的实现方式中，所述视频处理方法通过神经网络实现，所述神经网络至少包括所述特

征提取网络、所述M级动作识别网络，所述装置还包括：训练模块15，用于通过样本视频及所述样本视频的类别标注，对所述神经网络进行训练。

在一种可能的实现方式中，所述训练模块被进一步配置为：从所述样本视频中确定出多个样本视频帧；通过所述神经网络对所述样本视频帧进行处理，确定所述样本视频的分类结果；根据所述样本视频的分类结果及类别标注，确定所述神经网络的网络损失；根据所述网络损失调整所述神经网络的网络参数。

此外，本公开还提供了视频处理装置、电子设备、计算机可读存储介质、程序，上述均可用来实现本公开提供的任一种视频处理方法，相应技术方案和描述和参见方法部分的相应记载，不再赘述。

本领域技术人员可以理解，在具体实施方式的上述方法中，各步骤的撰写顺序并不意味着严格的执行顺序而对实施过程构成任何限定，各步骤的具体执行顺序应当以其功能和可能的内在逻辑确定。

在一些实施例中，本公开实施例提供的装置具有的功能或包含的模块可以用于执行上文方法实施例描述的方法，其具体实现可以参照上文方法实施例的描述，为了简洁，这里不再赘述

本公开实施例还提出一种计算机可读存储介质，其上存储有计算机程序指令，所述计算机程序指令被处理器执行时实现上述方法。计算机可读存储介质可以是非易失性计算机可读存储介质。

本公开实施例还提出一种电子设备，包括：处理器；用于存储处理器可执行指令的存储器；其中，所述处理器被配置为上述方法。

电子设备可以被提供为终端、服务器或其它形态的设备。

图10是根据一示例性实施例示出的一种电子设备800的框图。例如，电子设备800可以是移动电话，计算机，数字广播终端，消息收发设备，游戏控制台，平板设备，医疗设备，健身设备，个人数字助理等终端。

参照图10，电子设备800可以包括以下一个或多个组件：处理组件802，存储器804，电源组件806，多媒体组件808，音频组件810，输入/输出（I/O）的接口812，传感器组件814，以及通信组件816。

处理组件802通常控制电子设备800的整体操作，诸如与显示，电话呼叫，数据通信，相机操作和记录操作相关联的操作。处理组件802可以包括一个或多个处理器820来执行指令，以完成上述的方法的全部或部分步骤。此外，处理组件802可以包括一个或多个模块，便于处理组件802和其他组件之间的交互。例如，处理组件802可以包括多媒体模块，以方便多媒体组件808和处理组件802之间的交互。

存储器804被配置为存储各种类型的数据以支持在电子设备800的操作。这些数据的示例包括用于在电子设备800上操作的任何应用程序或方法的指令，联系人数据，电话簿数据，消息，图片，视频等。存储器804可以由任何类型的易失性或非易失性存储设备或者它们的组合实现，如静态随机存取存储器（SRAM），电可擦除可编程只读存储器（EEPROM），可擦除可编程只读存储器（EPROM），可编程只读存储器（PROM），只读存储器（ROM），磁存储器，快闪存储器，磁盘或光盘。

电源组件806为电子设备800的各种组件提供电力。电源组件806可以包括电源管理系统，一个或多个电源，及其他与为电子设备800生成、管理和分配电力相关联的组件。

多媒体组件808包括在所述电子设备800和用户之间提供一个输出接口的屏幕。在一些实施例中，屏幕可以包括液晶显示器（LCD）和触摸面板（TP）。如果屏幕包括触摸面板，屏幕可以被实现为触摸屏，以接收来自用户的输入信号。触摸面板包括一个或多个触摸传感器以感测触摸、滑动和触摸面板上的手势。所述触摸传感器可以不仅感测触摸或滑动动作的边界，而且还检测与所述触摸或滑动操作相关的持续时间和压力。在一些实施例中，多媒体组件808包括一个前置摄像头和/或后置摄像头。当电子设备800处于操作模式，如拍摄模式或视频模式时，前置摄像头和/或后置摄像头可以接收外部的多媒体数据。每个前置摄像头和后置摄像头可以是一个固定的光学透镜系统或具有焦距和光学变焦能力。

音频组件810被配置为输出和/或输入音频信号。例如，音频组件810包括一个麦克风（MIC），当电子设备800处于操作模式，如呼叫模式、记录模式和语音识别模式时，麦克风被配置为接收外部音频信号。所接收的音频信号可以被进一步存储在存储器804或经由通信组件816发送。在一些实施例中，

音频组件810还包括一个扬声器，用于输出音频信号。

I/O接口812为处理组件802和外围接口模块之间提供接口，上述外围接口模块可以是键盘，点击轮，按钮等。这些按钮可包括但不限于：主页按钮、音量按钮、启动按钮和锁定按钮。

5 传感器组件814包括一个或多个传感器，用于为电子设备800提供各个方面的状态评估。例如，传感器组件814可以检测到电子设备800的打开/关闭状态，组件的相对定位，例如所述组件为电子设备800的显示器和小键盘，传感器组件814还可以检测电子设备800或电子设备800一个组件的位置改变，用户与电子设备800接触的存在或不存在，电子设备800方位或加速/减速和电子设备800的温度变化。传感器组件814可以包括接近传感器，被配置用来在没有任何的物理接触时检测附近物体的存在。传感器组件814还可以包括光传感器，如CMOS或CCD图像传感器，用于在成像应用中使用。在一些实
10 施例中，该传感器组件814还可以包括加速度传感器，陀螺仪传感器，磁传感器，压力传感器或温度传感器。

通信组件816被配置为便于电子设备800和其他设备之间有线或无线方式的通信。电子设备800可以接入基于通信标准的无线网络，如WiFi，2G或3G，或它们的组合。在一个示例性实施例中，通信组件816经由广播信道接收来自外部广播管理系统的广播信号或广播相关信息。在一个示例性实施
15 例中，所述通信组件816还包括近场通信（NFC）模块，以促进短程通信。例如，在NFC模块可基于射频识别（RFID）技术，红外数据协会（IrDA）技术，超宽带（UWB）技术，蓝牙（BT）技术和其他技术来实现。

在示例性实施例中，电子设备800可以被一个或多个应用专用集成电路（ASIC）、数字信号处理器（DSP）、数字信号处理设备（DSPD）、可编程逻辑器件（PLD）、现场可编程门阵列（FPGA）、控
20 制器、微控制器、微处理器或其他电子元件实现，用于执行上述方法。

在示例性实施例中，还提供了一种非易失性计算机可读存储介质，例如包括计算机程序指令的存储器804，上述计算机程序指令可由电子设备800的处理器820执行以完成上述方法。

本公开实施例还提供了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一实施例提供的方法的指令。

25 该计算机程序产品可以具体通过硬件、软件或其结合的方式实现。在一个可选实施例中，所述计算机程序产品具体体现为计算机存储介质，在另一个可选实施例中，计算机程序产品具体体现为软件产品，例如软件开发包（Software Development Kit，SDK）等等。

图11是根据一示例性实施例示出的一种电子设备1900的框图。例如，电子设备1900可以被提供为一服务器。参照图11，电子设备1900包括处理组件1922，其进一步包括一个或多个处理器，以及由存储器1932所代表的存储器资源，用于存储可由处理组件1922的执行的指令，例如应用程序。存储器1932
30 中存储的应用程序可以包括一个或一个以上的每一个对应于一组指令的模块。此外，处理组件1922被配置为执行指令，以执行上述方法。

电子设备1900还可以包括一个电源组件1926被配置为执行电子设备1900的电源管理，一个有线或无线网络接口1950被配置为将电子设备1900连接到网络，和一个输入输出（I/O）接口1958。电子设备
35 1900可以操作基于存储在存储器1932的操作系统，例如Windows Server™，Mac OS X™，Unix™，Linux™，FreeBSD™或类似。

在示例性实施例中，还提供了一种非易失性计算机可读存储介质，例如包括计算机程序指令的存储器1932，上述计算机程序指令可由电子设备1900的处理组件1922执行以完成上述方法。

40 本公开可以是系统、方法和/或计算机程序产品。计算机程序产品可以包括计算机可读存储介质，其上载有用于使处理器实现本公开的各个方面的计算机可读程序指令。

计算机可读存储介质可以是可以保持和存储由指令执行设备使用的指令的有形设备。计算机可读存储介质例如可以是一—但不限于——电存储设备、磁存储设备、光存储设备、电磁存储设备、半导体存储设备或者上述的任意合适的组合。计算机可读存储介质的更具体的例子（非穷举的列表）包括：
45 便携式计算机盘、硬盘、随机存取存储器（RAM）、只读存储器（ROM）、可擦式可编程只读存储器（EPROM或闪存）、静态随机存取存储器（SRAM）、便携式压缩盘只读存储器（CD-ROM）、数字多

功能盘（DVD）、记忆棒、软盘、机械编码设备、例如其上存储有指令的打孔卡或凹槽内凸起结构、以及上述的任意合适的组合。这里所使用的计算机可读存储介质不被解释为瞬时信号本身，诸如无线电波或者其他自由传播的电磁波、通过波导或其他传输媒介传播的电磁波（例如，通过光纤电缆的光脉冲）、或者通过电线传输的电信号。

5 这里所描述的计算机可读程序指令可以从计算机可读存储介质下载到各个计算/处理设备，或者通过网络、例如因特网、局域网、广域网和/或无线网下载到外部计算机或外部存储设备。网络可以包括铜传输电缆、光纤传输、无线传输、路由器、防火墙、交换机、网关计算机和/或边缘服务器。每个计算/处理设备中的网络适配卡或者网络接口从网络接收计算机可读程序指令，并转发该计算机可读程序指令，以供存储在各个计算/处理设备中的计算机可读存储介质中。

10 用于执行本公开操作的计算机程序指令可以是汇编指令、指令集架构（ISA）指令、机器指令、机器相关指令、微代码、固件指令、状态设置数据、或者以一种或多种编程语言的任意组合编写的源代码或目标代码，所述编程语言包括面向对象的编程语言—诸如Smalltalk、C++等，以及常规的过程式编程语言—诸如“C”语言或类似的编程语言。计算机可读程序指令可以完全地在用户计算机上执行、部分地在用户计算机上执行、作为一个独立的软件包执行、部分在用户计算机上部分在远程计算机上执行、或者完全在远程计算机或服务器上执行。在涉及远程计算机的情形中，远程计算机可以通过任意种类的网络—包括局域网（LAN）或广域网（WAN）—连接到用户计算机，或者，可以连接到外部计算机（例如利用因特网服务提供商来通过因特网连接）。在一些实施例中，通过利用计算机可读程序指令的状态信息来个性化定制电子电路，例如可编程逻辑电路、现场可编程门阵列（FPGA）或可编程逻辑阵列（PLA），该电子电路可以执行计算机可读程序指令，从而实现本公开的各个方面。

20 这里参照根据本公开实施例的方法、装置（系统）和计算机程序产品的流程图和/或框图描述了本公开的各个方面。应当理解，流程图和/或框图的每个方框以及流程图和/或框图中各方框的组合，都可以由计算机可读程序指令实现。

这些计算机可读程序指令可以提供给通用计算机、专用计算机或其它可编程数据处理装置的处理装置，从而生产出一种机器，使得这些指令在通过计算机或其它可编程数据处理装置的处理装置执行时，产生了实现流程图和/或框图中的一个或多个方框中规定的功能/动作的装置。也可以把这些计算机可读程序指令存储在计算机可读存储介质中，这些指令使得计算机、可编程数据处理装置和/或其他设备以特定方式工作，从而，存储有指令的计算机可读介质则包括一个制造品，其包括实现流程图和/或框图中的一个或多个方框中规定的功能/动作的各个方面的指令。

也可以把计算机可读程序指令加载到计算机、其它可编程数据处理装置、或其它设备上，使得在计算机、其它可编程数据处理装置或其它设备上执行一系列操作步骤，以产生计算机实现的过程，从而使得在计算机、其它可编程数据处理装置、或其它设备上执行的指令实现流程图和/或框图中的一个或多个方框中规定的功能/动作。

附图中的流程图和框图显示了根据本公开的多个实施例的系统、方法和计算机程序产品的可能实现的体系架构、功能和操作。在这点上，流程图或框图中的每个方框可以代表一个模块、程序段或指令的一部分，所述模块、程序段或指令的一部分包含一个或多个用于实现规定的逻辑功能的可执行指令。在有些作为替换的实现中，方框中所标注的功能也可以以不同于附图中所标注的顺序发生。例如，两个连续的方框实际上可以基本并行地执行，它们有时也可以按相反的顺序执行，这依所涉及的功能而定。也要注意的，框图和/或流程图中的每个方框、以及框图和/或流程图中的方框的组合，可以用执行规定的功能或动作的专用的基于硬件的系统来实现，或者可以用专用硬件与计算机指令的组合来实现。

45 以上已经描述了本公开的各实施例，上述说明是示例性的，并非穷尽性的，并且也不限于所披露的各实施例。在不偏离所说明的各实施例的范围和精神的情况下，对于本技术领域的普通技术人员来说许多修改和变更都是显而易见的。本文中所用术语的选择，旨在最好地解释各实施例的原理、实际应用或对市场中的技术的技术改进，或者使本技术领域的其它普通技术人员能理解本文披露的各实施例。