

权 利 要 求 书

1、一种视频处理方法，其特征在于，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；

5 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。

2、根据权利要求1所述的方法，其特征在于，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，获得所述多个目标视频帧的动作识别特征，包括：

10 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；

15 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。

3、根据权利要求2所述的方法，其特征在于，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：

20 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。

4、根据权利要求3所述的方法，其特征在于，所述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：

25 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。

5、根据权利要求3所述的方法，其特征在于，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：

30 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；

35 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。

6、根据权利要求5所述的方法，其特征在于，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：

40 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。

7、根据权利要求3-6中任一项所述的方法，其特征在于，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所述待处理视频中的各目标视频帧对应；

45 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。
- 5 8、根据权利要求4-7中任意一项所述的方法, 其特征在于, 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。
- 10 9、根据权利要求1-8中任意一项所述的方法, 其特征在于, 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, 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。
- 10、根据权利要求1-9中任意一项所述的方法, 其特征在于, 所述方法还包括:
- 15 从待处理视频中确定出多个目标视频帧。
- 11、根据权利要求10所述的方法, 其特征在于, 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。
- 20 12、根据权利要求1-11中任一项所述的方法, 其特征在于, 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。
- 13、根据权利要求12所述的方法, 其特征在于, 通过样本视频及所述样本视频的类别标注, 对所
- 25 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。
- 30 14、一种视频处理装置, 其特征在于, 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
- 35 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。
- 15、根据权利要求14所述的装置, 其特征在于, 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
- 40 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。
- 45 16、根据权利要求15所述的装置, 其特征在于, 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；
- 5 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。
- 17、根据权利要求16所述的装置，其特征在于，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。
- 18、根据权利要求16所述的装置，其特征在于，所述动作识别模块被进一步配置为：
- 10 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；
- 15 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。
- 19、根据权利要求18所述的装置，其特征在于，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
- 20 第二特征信息包括一个行向量或列向量。
- 20、根据权利要求16-19中任一项所述的装置，其特征在于，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
- 25 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第k+1个目标视频帧对应的第五特征信息及与第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。
- 21、根据权利要求17-20中任一项所述的装置，其特征在于，所述动作识别模块被进一步配置为：
- 30 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。
- 22、根据权利要求14-21中任一项所述的装置，其特征在于，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；
- 35 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。
- 23、根据权利要求14-22中任一项所述的装置，其特征在于，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。
- 24、根据权利要求23所述的装置，其特征在于，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；
- 40 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。
- 25、根据权利要求14-24中任一项所述的装置，其特征在于，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类别标注，对所述神经网络进行训练。
- 45 26、根据权利要求25所述的装置，其特征在于，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。

5 27、一种电子设备，其特征在于，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调用所述存储器存储的指令，以执行权利要求1至13中任意一项所述的方法。

10 28、一种计算机可读存储介质，其上存储有计算机程序指令，其特征在于，所述计算机程序指令被处理器执行时实现权利要求1至13中任意一项所述的方法。

29、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实现权利要求1-13中的任一权利要求所述的方法。