

# (12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织  
国际局

(43) 国际公布日  
2022 年 4 月 14 日 (14.04.2022)



(10) 国际公布号  
**WO 2022/073282 A1**

- (51) 国际专利分类号:  
**G06K 9/00** (2006.01)
- (21) 国际申请号: PCT/CN2020/129550
- (22) 国际申请日: 2020 年 11 月 17 日 (17.11.2020)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:  
202011078182.6 2020 年 10 月 10 日 (10.10.2020) CN
- (71) 申请人: 中国科学院深圳先进技术研究院 (SHENZHEN INSTITUTES OF ADVANCED TECHNOLOGY CHINESE ACADEMY OF SCIENCES) [CN/CN]; 中国广东省深圳市南山区深圳大学城学苑大道 1068 号, Guangdong 518055 (CN)。
- (72) 发明人: 任子良 (REN, Ziliang); 中国广东省深圳市南山区深圳大学城学苑大道 1068 号, Guangdong 518055 (CN)。程俊 (CHENG, Jun); 中国广东省深圳市南山区深圳大学城学苑大道 1068

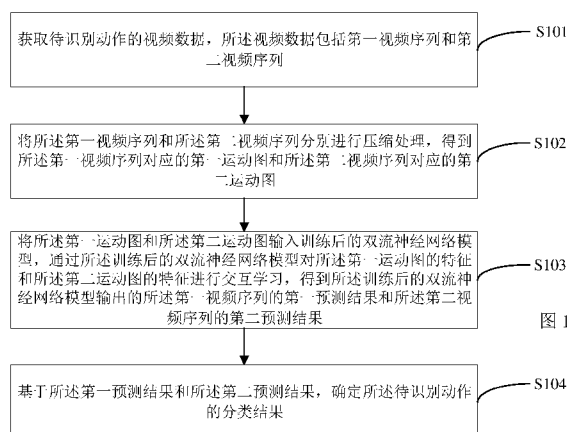
号, Guangdong 518055 (CN)。张锬石 (ZHANG, Qieshi); 中国广东省深圳市南山区深圳大学城学苑大道 1068 号, Guangdong 518055 (CN)。高向阳 (GAO, Xiangyang); 中国广东省深圳市南山区深圳大学城学苑大道 1068 号, Guangdong 518055 (CN)。康宇航 (KANG, Yuhang); 中国广东省深圳市南山区深圳大学城学苑大道 1068 号, Guangdong 518055 (CN)。

(74) 代理人: 北京中巡通大知识产权代理有限公司 (BEIJING ZHONG XUN TONG DA INTELLECTUAL PROPERTY AGENCY CO., LTD.); 中国北京市海淀区杏石口路 50 号中间建筑 1 区 148 号, Beijing 100195 (CN)。

(81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX,

(54) Title: MOTION RECOGNITION METHOD BASED ON FEATURE INTERACTIVE LEARNING, AND TERMINAL DEVICE

(54) 发明名称: 一种基于特征交互学习的动作识别方法及终端设备



- S101 Acquire video data of a motion to be recognized, wherein the video data comprises a first video sequence and a second video sequence
- S102 Respectively perform compression processing on the first video sequence and the second video sequence, so as to obtain a first motion graph corresponding to the first video sequence and a second motion graph corresponding to the second video sequence
- S103 Input the first motion graph and the second motion graph into a trained two-stream neural network model, and perform interactive learning on features of the first motion graph and features of the second motion graph by means of the trained two-stream neural network model, so as to obtain a first prediction result of the first video sequence and a second prediction result of the second video sequence, which are output by the trained two-stream neural network model
- S104 On the basis of the first prediction result and the second prediction result, determine a classification result of the motion to be recognized

(57) Abstract: A motion recognition method based on feature interactive learning, and a terminal device, which are applicable to the technical field of computer vision. The method comprises: acquiring video data of a motion to be recognized, the video data comprising a first video sequence and a second video sequence (S101); respectively performing compression processing on the first video sequence and the second video sequence, so as to obtain a first motion graph and a second motion graph (S102); inputting the first motion graph and the second motion graph into a trained two-stream neural network model, and performing interactive learning on features of the first motion graph and features of the second motion graph by means of the trained two-stream neural network model, so as to obtain a first prediction result of the first video sequence and a second prediction result of the second video sequence (S103); and on the basis of the first prediction result and the second prediction result, determining a classification result of the motion to be recognized (S104). According to the method, the problem of low motion recognition accuracy of sparse sampling is solved. By means of interactive learning of multi-modal input features, the accuracy of motion category recognition is improved.

MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL,  
PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL,  
ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US,  
UZ, VC, VN, WS, ZA, ZM, ZW。

(84) 指定国 (除另有指明, 要求每一种可提供的地区  
保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ,  
NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM,  
AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG,  
CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU,  
IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT,  
RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI,  
CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告 (条约第21条(3))。

(57) 摘要: 一种基于特征交互学习的动作识别方法及终端设备, 适用于计算机视觉技术领域, 所述方法包括: 获取待识别动作的视频数据, 视频数据包括第一视频序列和第二视频序列 (S101); 将第一视频序列和第二视频序列分别进行压缩处理, 得到第一运动图和第二运动图 (S102); 将第一运动图和第二运动图输入训练后的双流神经网络模型, 通过训练后的双流神经网络模型对第一运动图的特征和第二运动图的特征进行交互学习, 得到第一视频序列的第一预测结果和第二视频序列的第二预测结果 (S103); 基于第一预测结果和第二预测结果, 确定待识别动作的分类结果 (S104)。该方法解决了稀疏采样的动作识别准确度低的问题; 通过对多模态输入特征的交互学习, 提高了对动作类别识别的准确度。

# 一种基于特征交互学习的动作识别方法及终端设备

## 技术领域

5           本申请属于计算机视觉技术领域，尤其涉及一种基于特征交互学习的动作识别方法及终端设备。

## 背景技术

近年来，人体动作识别已成为计算机视觉领域的研究热点之一。通过动作识别技术，计算机可以自动理解和描述视频中的人体动作，在诸多领域具有巨大的应用价值，例如：视频监控、人机交互、运动分析、基于内容的视频检索以及自动驾驶等领域。人体动作识别的方法主要包括基于人工设计特征的方法和基于神经网络深度学习特征的方法。

与传统的基于人工设计特征的方法相比，基于神经网络深度学习特征的方法对人体动作的识别取得了一定的成功。然而目前基于神经网络深度学习的人体动作识别方法，在处理长视频序列的动作分类识别时，通过稀疏采样获取一定数量的视频帧作为神经网络的输入，经过神经网络的逐层提取视频帧中的特征，对人体动作进行识别和分类；由于视频拍摄视角、拍摄尺寸、拍摄背景的复杂多变，以及动作的差异性和相似性，针对单一模态的稀疏采样方式，动作识别的准确率较低。

## 20 发明内容

本申请实施例提供了一种基于特征交互学习的动作识别方法及终端设备，可以解决由于单一模态的稀疏采样方式，动作识别的准确率较低的问题。

第一方面，本申请实施例提供了一种基于特征交互学习的动作识别方法，所述方法包括：获取待识别动作的视频数据，所述视频数据包括第一视频序列和第二视频序列；将所述第一视频序列和所述第二视频序列分别进行压缩处理，得到所述第一视频序列对应的第一运动图和所述第二视频序列对应的第二运动图；将所述第一运动图和所述第二运动图输入训练后的双流神经网络模型，通

过所述训练后的双流神经网络模型对所述第一运动图的特征和所述第二运动图的特征进行交互学习，得到所述训练后的双流神经网络模型输出的所述第一视频序列的第一预测结果和所述第二视频序列的第二预测结果；基于所述第一预测结果和所述第二预测结果，确定所述待识别动作的分类结果。

5       在第一方面的一种可能的实现方式中，所述将所述第一视频序列进行压缩处理，得到所述第一视频序列对应的第一运动图，包括：

获取所述第一视频序列中每一视频帧的特征矩阵；根据所述第一视频序列中视频帧的时序，将每一视频帧的所述特征矩阵进行压缩计算，得到用于表示所述第一运动图的特征矩阵。

10       在第一方面的一种可能的实现方式中，所述将所述第二视频序列进行压缩处理，得到所述第二视频序列对应的第二运动图，包括：

将所述第二视频序列进行灰度处理，得到所述第二视频序列对应的灰度序列帧；根据所述第二视频序列中视频帧的时序，将所述灰度序列帧的特征矩阵进行压缩计算，得到用于表示所述第二运动图的特征矩阵。

15       在第一方面的一种可能的实现方式中，所述训练后的双流神经网络模型包括第一神经网络模型、第二神经网络模型和路由模块，所述路由模块设置于所述第一神经网络模型的中间卷积模块和所述第二神经网络模型的中间卷积模块之间；所述第一神经网络模型的输入为所述第一运动图，输出为所述第一视频序列的所述第一预测结果；所述第二神经网络模型的输入为所述第二运动图，  
20       输出为所述第二视频序列的所述第二预测结果；所述路由模块用于在所述第一神经网络模型的中间卷积模块和所述第二神经网络模型的中间卷积模块之间，对所述第一运动图的特征和所述第二运动图的特征进行交互学习。

在第一方面的一种可能的实现方式中，所述第一神经网络模型的中间卷积模块包括预设层数的第一卷积模块，所述第二神经网络模型的中间卷积模块包  
25       括与所述第一卷积模块相对应的第二卷积模块；所述将所述第一运动图和所述第二运动图输入训练后的双流神经网络模型，通过所述训练后的双流神经网络模型对所述第一运动图的特征和所述第二运动图的特征进行交互学习，得到所述第一视频序列的第一预测结果和所述第二视频序列的第二预测结果，包括：

将第一层的第一卷积模块的输出和第一层的第二卷积模块的输出作为第一

层的路由模块的输入，由所述第一层的路由模块进行特征交互学习，得到第一路由输出；将所述第一层的第一卷积模块的输出与所述第一路由输出的叠加结果作为第二层的第一卷积模块的输入，由所述第二层的第一卷积模块进行特征学习，得到所述第二层的第一卷积模块的输出；将所述第一层的第二卷积模块的输出与所述第一路由输出的叠加结果作为第二层的第二卷积模块的输入，由所述第二层的第二卷积模块进行特征学习，得到所述第二层的第二卷积模块的输出；将所述第二层的第一卷积模块的输出和所述第二层的第二卷积模块的输出作为第二层的路由模块的输入，由所述第二层的路由模块进行特征交互学习，得到第二路由输出；

其中，所述第一层的第一卷积模块与第二层的第一卷积模块为所述第一神经网络模型的中间卷积模块中前后相邻的两层卷积层；所述第一层的第二卷积模块和所述第二层的第二卷积模块为所述第二神经网络模型的中间卷积模块中前后相邻的两层卷积层；所述第一层的路由模块和所述第二层的路由模块为前后相邻的两个计算模块。

在第一方面的一种可能的实现方式中，所述路由模块包括：第一卷积单元、第一归一化单元、第一激活单元、第二卷积单元、第二归一化单元、第二激活单元；通过所述路由模块的所述第一卷积单元、所述第一归一化单元、所述第一激活单元、所述第二卷积单元、所述第二归一化单元、所述第二激活单元，依次对所述第一神经网络模型的卷积计算模块输出的特征矩阵和所述第二神经网络模型的卷积计算模块输出的特征矩阵进行交互学习，得到所述路由模块输出的特征矩阵。

在第一方面的一种可能的实现方式中，所述基于所述第一预测结果和所述第二预测结果，确定所述待识别动作的分类结果，包括：

对所述第一预测结果和所述第二预测结果进行特征融合，得到动作类别的概率分布；将所述概率分布中概率最大的动作类别作为所述待识别动作的所述分类结果。

在第一方面的一种可能的实现方式中，所述第一神经网络模型包括第一损失函数，所述第二神经网络模型包括第二损失函数；通过样本视频数据对所述第一神经网络模型、所述第二神经网络模型和所述路由模块进行训练，依据所

述第一损失函数和所述第二损失函数分别调整所述第一神经网络模型的参数、  
所述第二神经网络模型的参数以及所述路由模块的参数；若所述第一损失函数  
和所述第二损失函数满足预设阈值，则停止对所述第一神经网络模型的参数、  
所述第二神经网络模型的参数以及所述路由模块的训练，得到所述训练后的双  
5 流神经网络模型。

第二方面，本申请实施例提供了一种基于特征交互学习的动作识别装置，  
包括：

获取单元，用于获取待识别动作的视频数据，所述视频数据包括第一视频  
序列和第二视频序列；

10 处理单元，用于将所述第一视频序列和所述第二视频序列分别进行压缩处  
理，得到所述第一视频序列对应的第一运动图和所述第二视频序列对应的第二  
运动图；

计算单元，用于将所述第一运动图和所述第二运动图输入训练后的双流神  
经网络模型，通过所述训练后的双流神经网络模型对所述第一运动图的特征和  
15 所述第二运动图的特征进行交互学习，得到所述第一视频序列的第一预测结果  
和所述第二视频序列的第二预测结果；

输出单元，用于基于所述第一预测结果和所述第二预测结果，确定所述待  
识别动作的分类结果。

第三方面，本申请实施例提供了一种终端设备，包括：存储器、处理器以  
20 及存储在所述存储器中并可在所述处理器上运行的计算机程序，所述处理器执  
行所述计算机程序时实现第一方面及第一方面的可能实现方式所述的方法。

第四方面，本申请实施例提供了一种计算机可读存储介质，所述计算机可  
读存储介质存储有计算机程序，所述计算机程序被处理器执行时实现第一方面  
及第一方面的可能实现方式所述的方法。

25 第五方面，本申请实施例提供了一种计算机程序产品，当计算机程序产品  
在终端设备上运行时，使得终端设备执行上述第一方面中任一项所述的动作识  
别方法。

可以理解的是，上述第二方面至第五方面的有益效果可以参见上述第一方  
面的相关描述，在此不再赘述。

本申请实施例与现有技术相比存在的有益效果是：通过本申请实施例，终端设备可以获取待识别动作的视频数据，视频数据包括第一视频序列和第二视频序列；将第一视频序列和第二视频序列分别进行压缩处理，得到第一视频序列对应的第一运动图和第二视频序列对应的第二运动图；将第一运动图和第二运动图输入训练后的双流神经网络模型，通过训练后的双流神经网络模型对第一运动图的特征和第二运动图的特征进行交互学习，得到第一视频序列的第一预测结果和第二视频序列的第二预测结果；基于第一预测结果和第二预测结果，确定待识别动作的分类结果；通过分别对第一视频序列和第二视频序列的压缩得到第一运动图和第二运动图，对视频数据进行更丰富的时空表示，使得信息表示更全，特征更丰富；从而将第一运动图和第二运动图作为双流神经网络模型的输入，通过神经网络模型对多模态的图像特征进行交互学习，提高了动作识别的准确度；具有较强的易用性与实用性。

## **附图说明**

为了更清楚地说明本申请实施例中的技术方案，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申请的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动性的前提下，还可以根据这些附图获得其他的附图。

图 1 是本申请实施例提供的动作识别方法的流程示意图；

图 2 是本申请实施例提供的视频数据压缩处理的示意图；

图 3 是本申请实施例提供的双流神经网络模型的网络架构示意图；

图 4 是本申请实施例提供的双流神经网络的路由模块的架构示意图；

图 5 是本申请实施例提供的双流神经网络的中层特征交互学习单元的架构示意图；

图 6 是本申请实施例提供的动作识别装置的结构示意图；

图 7 是本申请实施例提供的终端设备的结构示意图。

## **具体实施方式**

以下描述中，为了说明而不是为了限定，提出了诸如特定系统结构、技术之类的具体细节，以便透彻理解本申请实施例。然而，本领域的技术人员应当

清楚,在没有这些具体细节的其它实施例中也可以实现本申请。在其它情况中,省略对众所周知的系统、装置、电路以及方法的详细说明,以免不必要的细节妨碍本申请的描述。

应当理解,当在本申请说明书和所附权利要求书中使用时,术语“包括”指示所描述特征、整体、步骤、操作、元素和/或组件的存在,但并不排除一个或多个其它特征、整体、步骤、操作、元素、组件和/或其集合的存在或添加。

还应当理解,在本申请说明书和所附权利要求书中使用的术语“和/或”是指相关联列出的项中的一个或多个的任何组合以及所有可能组合,并且包括这些组合。

如在本申请说明书和所附权利要求书中所使用的那样,术语“如果”可以依据上下文被解释为“当...时”或“一旦”或“响应于确定”或“响应于检测到”。类似地,短语“如果确定”或“如果检测到[所描述条件或事件]”可以依据上下文被解释为意指“一旦确定”或“响应于确定”或“一旦检测到[所描述条件或事件]”或“响应于检测到[所描述条件或事件]”。

另外,在本申请说明书和所附权利要求书的描述中,术语“第一”、“第二”、“第三”等仅用于区分描述,而不能理解为指示或暗示相对重要性。

在本申请说明书中描述的参考“一个实施例”或“一些实施例”等意味着在本申请的一个或多个实施例中包括结合该实施例描述的特定特征、结构或特点。由此,在本说明书中的不同之处出现的语句“在一个实施例中”、“在一些实施例中”、“在其他一些实施例中”、“在另外一些实施例中”等不是必然都参考相同的实施例,而是意味着“一个或多个但不是所有的实施例”,除非是以其他方式另外特别强调。术语“包括”、“包含”、“具有”及它们的变形都意味着“包括但不限于”,除非是以其他方式另外特别强调。

目前,针对计算机视觉任务,主要采用以卷积神经网络为基础的二维卷积网络 2D-ConvNets、三维卷积网络 3D-ConvNets 以及循环神经网络(RNN)等模型。卷积神经网络的特征学习能力和动作识别效果取得了一定的成果,但是针对长视频序列的动作分类的处理和识别过程中,仍存在诸多挑战。

基于卷积神经网络的人体动作识别方法,给定一定数量的 RGB 或者深度视频序列,通过稀疏采样方法得到一定数量的视频帧作为网络的输入,卷积神

神经网络逐层提取视频帧中的特征，并通过分类器或归一化（Softmax 函数）层对人体动作进行分类和识别。

其中，基于卷积神经网络的人体动作识别方法可分为以下两大类：第一，2D 端到端的网络训练结构：通过大规模标注数据集对深度网络进行有监督训练，之后通过参数微调训练获得实际任务的训练模型。针对视频序列，该类方法主要采用稀疏采样得到整个视频序列的某一帧图像作为网络输入，不能很好的学习到人体动作的时间维度特征。第二，3D 端到端网络训练结构：通过稀疏采样得到某几帧图像作为网络模型的输入，并通过有监督训练和参数微调训练得到分类模型。该方法可获取较好的识别效果，但计算量庞大而制约了其在实际场景中的应用。

另外，基于其他深度网络的人体动作识别方法，应用于人体动作识别的其他深度网络有循环卷积网络（RNN）等。RNN 网络对前面的信息进行记忆并应用于当前输出的计算中，能够对任何长度的序列数据进行处理，通过对输入序列的有序循环学习来实现特征学习和识别分类，其在自然语言处理领域取得了较好的应用，但在人体动作识别上还有待进一步提升。

综上，现有的计算机视觉中，基于卷积神经网络或其它深度网络的动作识别方法，缺乏对长视频序列的多模态时空信息的表示以及对多模态特征的相互学习，从而动作识别的准确度仍有待提高。

本申请将基于对长视频序列的多模态时空信息的表示以及对多模态特征的相互学习，实现对用户动作分类的处理和识别，进一步提高了用户动作识别的准确度。下面结合图示及具体实施例来详细说明本申请的技术方案。

请参见图 1，是本申请实施例提供的动作识别方法的流程示意图，该方法的执行主体可以是独立的终端，例如可以是手机、电脑、多媒体设备、流媒体设备、监控装置等终端设备；还可以为终端设备中的集成模块，作为终端设备中的某一项功能实现。下面以该方法由终端设备执动作例进行说明，但本申请实施例不限于此。如图 1 所示，该方法包括：

步骤 S101，获取待识别动作的视频数据，视频数据包括第一视频序列和第二视频序列。

在一些实施例中，视频数据为按时序组合的连续的多帧图像序列，为待识

别动作的整段视频的所有图像帧的序列。终端设备可以通过 RGB-D 摄像装置实时获取待识别动作的视频数据；待识别动作的视频数据还可以是预存在终端设备中的视频数据。第一视频序列和第二视频序列为两种不同模态的视频帧序列，即同一段视频数据的不同的特征表示形式，例如分别可以为 RGB 格式的彩色视频序列、以深度信息表示的视频序列、光流图格式的视频序列或骨架序列等。

其中，以第一视频序列为 RGB 格式的彩色视频序列和第二视频序列为深度信息表示的视频序列为例进行详细说明。彩色视频序列为 RGB 格式的多帧图像序列，即通过红绿蓝 RGB 三种颜色表示每一帧图像中的像素点信息的彩色图像序列；以深度信息表示的视频序列为通过深度值表示每一帧图像中的像素点信息的深度图像序列，每一帧图像的图像深度决定了图像在每个像素点可能的颜色数或者可能的灰度等级。

需要说明的是，通过设置拍摄装置的拍摄参数，使得所获取的第一视频序列和第二视频序列的序列帧可以按时序分别一一对应，即相同时刻的第一视频序列的每一视频帧和第二视频序列的每一视频帧相对应，例如设置每秒拍摄 20 帧，同一段视频的第一视频序列和第二视频序列可以包括相同数量的视频帧；从而可以通过不同的特征量表示相同时刻的视频帧的时空信息。

其中，时空信息包括视频帧序列的时间维度的信息和空间维度的信息；时间维度的信息表现在不同时间点对应一个视频帧，由时间的连续性使得连续的视频帧序列构成动态效果；空间维度的信息可以表现为每个视频帧的纹理信息或色彩信息。例如，RGB 格式的彩色视频序列的视频帧通过 3 通道\*宽 W\*高 H 矩阵形式进行表示，通过三个通道中的元素表示视频帧中每个像素点的色彩信息；以深度信息表示的视频序列的视频帧，深度信息通过长度单位（如毫米等）进行衡量，为便于计算机处理，将表示深度的距离信息与灰度信息进行对应的转换，得到以 1 通道\*宽 W\*高 H 矩阵形式表示的灰度图。

另外，第一视频序列的序列帧和第二视频序列的序列帧按时序一一对应，且为相同的一段视频数据，例如，通过设置拍摄装置的拍摄参数，在相同的拍摄时间内，第一视频序列包括 50 帧图像，第二视频帧也包括 50 帧图像。

在一些实施例中，若第一视频序列为 RGB 格式的彩色视频数据，则可以

通过 RGB 格式的相机获取；若第二视频序列为以深度信息表示的视频序列，则可以通过深度相机获取；两种相机设置的拍摄参数可以相同，且为同一时间段对相同目标的拍摄，在此不做具体限制。

其中，终端设备可以为与摄像装置一体的设备，终端设备可以通过摄像装置直接获取待识别动作的视频数据；终端设备还可以是与摄像装置独立分开的设备，终端设备通过有线或无线的方式与摄像装置进行通信连接，以获取待识别动作的视频数据。待识别动作可以为人的行为或活动动作，也可以是动物的行为或活动动作，不做具体限制。

上述实施例，终端设备获取了待识别动作的整段视频数据的图像帧序列，记录了视频数据的多模态的底层特征，并且很好的利用了第一视频序列和第二视频序列两种不同模态的时空信息的特征，为后续神经网络模型对特征学习的多种可能性提供基础，增强了神经网络模型对图像特征的表达及识别能力。

步骤 S102，将第一视频序列和第二视频序列分别进行压缩处理，得到第一视频序列对应的第一运动图和第二视频序列对应的第二运动图。

在一些实施例中，为了可以对待识别动作的整段视频数据进行特征学习，终端设备将第一视频序列的多帧图像和第二视频序列的多帧图像分别进行压缩处理，得到包含丰富的时空信息的第一运动图和第二运动图。其中第一运动图的特征表示方式不同于第二运动图的特征表示方式，为同一段视频的不同底层特征的表示，即通过不同的图像信息分别表示第一视频序列和第二视频序列中视频帧的图像特征。

第一运动图的时空信息包含第一视频序列的所有视频帧的时空信息，第二运动图的时空信息包含第二视频序列所有视频帧的时空信息；例如，将 RGB 视频序列和深度 depth 视频序列的时间维度信息和空间维度信息分别压缩表示为单张三通道图像和单通道图像，呈现出动态的效果以及色彩、纹理等信息。

在实际计算过程中，第一视频序列的每一视频帧对应一个特征矩阵，第二视频序列的每一视频帧对应一个特征矩阵；例如第一视频序列或第二视频序列分别可以包括 T 帧图像，每一帧图像对应的特征矩阵为  $I_t$ ，则第一视频序列的特征矩阵集合或第二视频序列的特征矩阵集合可以表示为  $\{I_1, I_2, I_3, \dots, I_T\}$ ，其中， $I_1$  为视频序列中按照时序排列的第一帧图像的特征矩阵，以此类推， $I_T$  为视频

序列中按照时序排列的第  $T$  帧图像的特征矩阵。

在一些实施例中，将第一视频序列和第二视频序列分别进行压缩处理，将视频序列的多帧图像压缩合成为一张图像，该图像包含通过时间和空间表示动作的特征信息，可以称为运动图，从而得到包含整段视频序列时空信息的成对的第一运动图和第二运动图；即将多帧图像的特征矩阵合并在一张图像中进行表示，从而可以获取视频序列中所有视频帧的特征。

示例性的，第一运动图可以是 RGB 格式的视频序列帧压缩合成的图像，第二运动图可以为以深度信息表示的视频序列压缩合成的图像；第一运动图和第二运动图还可以是其它模态的一一对应的视频序列分别压缩合成的图像。

在一些实施例中，将第一视频序列进行压缩处理，得到第一视频序列对应的第一运动图，包括：

A1、获取所述第一视频序列中每一视频帧的特征矩阵；

A2、根据所述第一视频序列中视频帧的时序，将所述特征矩阵进行压缩计算，得到用于表示所述第一运动图的特征矩阵。

在一些实施例中，第一视频序列包括多帧图像，每一帧图像对应一个特征矩阵；若第一视频序列为 RGB 格式的彩色视频数据，则第一视频序列的每一帧图像的特征矩阵为 3 通道\*宽度  $W$ \*高度  $H$  的矩阵，其中宽度  $W$  和高度  $H$  以像素为单位，特征矩阵中的元素与像素相对应。特征矩阵中的每个元素的值表示对应位置的像素点的特征，例如 RGB 格式的彩色图像，每个元素代表每个像素点分别在红 R、绿 G、蓝 B 三个通道的特征值。

在一些实施例中，第一视频序列的每一帧图像对应一个特征矩阵，将所有视频帧的特征矩阵中的相同位置处的元素相加，再除以第一视频序列的视频帧的总帧数，得到特征矩阵中的每个位置处的元素值，对每个元素值取整，比如 2.6 向下取整则得到 2，得到第一视频序列对应的第一运动图的特征矩阵。

如图 2 所示，本申请实施例提供的视频数据压缩处理的示意图，当视频序列为 RGB 格式的彩色视频数据时，将 RGB 视频序列进行压缩处理，得到对应的 RGB 运动图，将多帧图像的时空信息合成为一张运动图的时空信息。其中，RGB 视频序列对应的运动图的特征矩阵可以为  $3*W*H$  的矩阵。可以通过下面公式进行计算：

$$MI = \frac{1}{T} \sum_{\tau=1}^T I_{\tau}$$

其中，MI 为第一视频序列对应运动图的特征矩阵，T 为第一视频序列的总帧数， $I_{\tau}$  为第一视频序列中第  $\tau$  帧图像的特征矩阵， $\tau$  的取值范围为[1, T]的整数。

5 另外，第一视频序列的每一帧图像的特征矩阵中元素的取值范围可以为[0,255]的整数，对第一视频序列压缩处理后的运动图 MI 的特征矩阵中每个元素的取值范围也为[0,255]的整数。

在一些实施例中，将第二视频序列进行压缩处理，得到第二视频序列对应的第二运动图，包括：

10 B1、将第二视频序列进行灰度处理，得到第二视频序列对应的灰度序列帧；  
B2、根据灰度序列帧的时序，将灰度序列帧的特征矩阵进行压缩计算，得到用于表示第二运动图的特征矩阵。

15 在一些实施例中，第二视频序列包括多帧图像，每一帧图像对应一个特征矩阵；若第二视频序列为以深度信息表示每一视频帧的图像序列，则第二视频序列的每一帧图像的特征矩阵为 1 通道\*宽度 W\*高度 H 的矩阵，其中宽度 W 和高度 H 以像素为单位，特征矩阵中的元素与像素相对应。特征矩阵中的每个元素的值表示对应位置的像素点的特征。由于第二视频序列为深度信息表示的图像序列，可以将第二视频序列中的每一帧深度图进行灰度处理，将深度图中每个像素点的深度信息进行[0,255]的映射转化，得到视频帧的灰度图像，灰度  
20 图像的特征矩阵中每个元素的取值范围为[0,255]的整数。

示例性的，以深度信息表示的视频序列的取值可能是 0 到 10000mm，而计算机视觉中图像的表示范围是[0,255]，所以需要将深度信息表示的视频序列缩放为与视觉表示匹配的取值范围，即将深度信息表示的视频序列向灰度图进行映射转化。其中，缩放方式有多种，假设深度信息表示的视频序列为 1\*W\*H  
25 矩阵，设所有元素的最大值和最小值之差为 max-min，对视频序列中每一个深度图像的矩阵的元素进行缩放和取整操作。例如：假设最大深度值 max-最小深度值 min=10000，某一个元素值为 7580，则操作后此元素的对应值为  $(7580/10000) * 255 = 193.29$ ，然后再取整得到 193，即所对应的元素值为 193，从而实现向灰度图像的转化。

在一些实施例中，第二视频序列与第一视频序列的压缩处理过程类似，将第二视频序列的每一帧图像进行灰度处理后，得到灰度图像的特征矩阵，将第二视频序列中所有视频帧对应的灰度图像的特征矩阵的相同位置处的元素相加，再除以第二视频序列的视频帧的总帧数，得到特征矩阵中的每个位置处的元素值，对每个元素值取整，得到第二视频序列对应的运动图的特征矩阵。

如图 2 所示，本申请实施例提供的视频数据压缩处理的示意图，当视频序列为以深度信息表示的视频序列时，将深度视频序列进行灰度处理，得到深度视频序列对应的灰度图像，将灰度图像进行压缩处理，得到对应的深度运动图，将多帧图像的时空信息合成为一张运动图的时空信息。其中，深度视频序列对应的运动图的特征矩阵可以为  $1*W*H$  的矩阵。可以通过下面公式进行计算：

$$MJ = \frac{1}{N} \sum_{n=1}^N J_n$$

其中， $MJ$  为第二视频序列对应运动图的特征矩阵， $N$  为第二视频序列的总帧数， $J_n$  为第二视频序列中第  $n$  帧图像的特征矩阵， $n$  的取值范围为  $[1, N]$  的整数。 $N$  和  $T$  可以相等， $n$  和  $\tau$  取值可以相等，即第一视频序列的视频帧和第二视频序列的视频帧按时序一一对应。

另外，第二视频序列对应的每一帧灰度图像的特征矩阵中元素的取值范围可以为  $[0, 255]$  的整数，第二视频序列对应的运动图  $MJ$  的特征矩阵中每个元素的取值范围可以为  $[0, 255]$  的整数。

需要说明的是，**RGB** 格式的第一视频序列中的视频帧和以深度信息表示的第二视频序列中视频帧可以一一对应。将以深度信息表示的第二视频序列中视频帧进行灰度处理后得到的灰度图序列与 **RGB** 格式的第一视频序列中的视频帧也一一对应。

步骤 S103，将第一运动图和第二运动图输入训练后的双流神经网络模型，通过训练后的双流神经网络模型对第一运动图的特征和第二运动图的特征进行交互学习，得到第一视频序列的第一预测结果和第二视频序列的第二预测结果。

在一些实施例中，双流神经网络模型为包括两路独立的卷积神经网络模型以及路由模块的整体模型。双流神经网络模型包括两路输入和两路输出。其中，两路输入分别对应视频数据两种模态的特征信息，两路输出分别对应两种模态输入信息的预测结果。

如图 3 所示，本申请实施例提供的双流神经网络模型的网络架构示意图，双流神经网络模型包括两路独立的卷积神经网络模型和路由模块，两路卷积神经网络模型的输入分别为第一运动图和第二运动图；每一路的卷积神经网络模型包括多个卷积层，例如卷积模块 Conv1、卷积模块 Conv2\_x、卷积模块 Conv5\_x 以及全连接层，其中卷积模块 Conv2\_x、卷积模块 Conv5\_x 分别表示一个总的卷积模块，一个总的卷积模块可以包括若干数量的卷积层或卷积计算单元。两路卷积神经网络模型的每个卷积模块之后通过路由模块对上一模块的输出结果进行交互学习，路由模块的输出作为和上一卷积模块的输出叠加，作为下一卷积模块的输入，通过路由模块学习双流神经网络模型中不同模态的中层交互特征。

其中，两路卷积神经网络模型的基础网络可以为残差网络（ResNet），由于残差网络的高度模块化，可以将残差网络中的各个模块作为基础模块对第一运动图和第二运动图不同模态的特征信息进行模型训练和特征的交互学习。双流神经网络模型通过双损失函数进行模型的优化和训练。

示例性的，双流神经网络模型的基础网络模型可以为 Inception、ImageNet、TSN 以及双流网络等深度网络模型；通过微调对基础网络模型的参数进行训练及调整；还可以根据需要设计网络模型进行参数的训练集调整。通过双流神经网络模型对不同模态的运动图像的特征学习后，通过双损失函数进行联合优化训练，得到与输入的不同模态的图像特征对应模态的双流高层特征；如输入的模态为 RGB 格式的运动图像和深度信息表示的运动图像，则可以得 RGB 格式及深度信息两种模态的双流高层特征。

在一些实施例中，两路输入可以包括多个通道输入；例如若其中一路输入为 RGB 运动图，则该路输入可以包括三个通道的输入，分别对应输入 RGB 运动图的红色 R 通道的特征矩阵、绿色 G 通道的特征矩阵以及蓝色 B 通道的特征矩阵。

在一些实施例中，训练后的双流神经网络模型包括第一神经网络模型、第二神经网络模型和路由模块，路由模块设置于第一神经网络模型的中间卷积模块和第二神经网络模型的中间卷积模块之间；第一神经网络模型的输入为第一运动图，输出为第一视频序列的第一预测结果；第二神经网络模型的输入为所

述第二运动图，输出为所述第二视频序列的所述第二预测结果；路由模块用于在第一神经网络模型的中间卷积模块和第二神经网络模型的中间卷积模块之间，对双流神经网络模型的每一层卷积模块的输出特征进行交互学习。

如图 4 所示，本申请实施例提供的双流神经网络模型的架构图，第一神经网络模型对应一路输入、输出，第二神经网络模型对应另一路输入、输出。输入第一神经网络模型的第一运动图可以为 RGB 运动图；第一神经网络模型输出的第一预测结果为第一视频序列对应的识别结果，第一视频序列可以为 RGB 格式的 RGB 视频序列；RGB 运动图为由 RGB 格式的 RGB 视频序列压缩得到。输入第二神经网络模型的第二运动图可以为深度运动图；第二神经网络模型输出的第二预测结果为第二视频序列对应的识别结果，第二视频序列可以为以深度信息表示的深度视频序列；深度运动图为由深度信息表示的深度视频序列压缩得到。

在双流神经网络的中层，包括多个卷积模块和多个路由模块，如图 4 中所示的卷积模块 Conv1、卷积模块 Conv2\_x 以及卷积模块 Conv5\_x；路由模块设置在两路卷积神经网络模型的每个卷积模块之后，通过路由模块对上一模块的输出结果进行交互学习，路由模块的输出作为和上一卷积模块的输出叠加，作为下一卷积模块的输入，通过路由模块学习双流神经网络模型中不同模态的中层交互特征。

在一些实施例中，第一神经网络模型的中间卷积模块包括预设层数的第一卷积模块，第二神经网络模型的中间卷积模块包括与第一卷积模块相对应的第二卷积模块。

如图 4 所示，将第一运动图和第二运动图输入训练后的双流神经网络模型，通过训练后的双流神经网络模型对第一运动图的特征和第二运动图的特征进行交互学习，得到第一视频序列的第一预测结果和第二视频序列的第二预测结果，包括：

C1、将第一层的第一卷积模块的输出和第二卷积模块的输出作为第一层的路由模块的输入，由所述第一层的路由模块进行特征交互学习，得到第一路由输出；

C2、将所述第一层的第一卷积模块的输出与所述第一路由输出的叠加结果

作为第二层的第一卷积模块的输入，由所述第二层的第一卷积模块进行特征学习，得到所述第二层的第一卷积模块的输出；

C3、将所述第一层的第二卷积模块的输出与所述第一路由输出的叠加结果作为第二层的第二卷积模块的输入，由所述第二层的第二卷积模块进行特征学习，得到所述第二层的第二卷积模块的输出；

C4、将所述第二层的第一卷积模块的输出和所述第二层的第二卷积模块的输出作为第二层的路由模块的输入，由所述第二层的路由模块进行特征交互学习，得到第二路由输出。

其中，所述第一层的第一卷积模块与第二层的第一卷积模块为所述第一神经网络模型的中间卷积模块中前后相邻的两层卷积层；所述第一层的第二卷积模块和所述第二层的第二卷积模块为所述第二神经网络模型的中间卷积模块中前后相邻的两层卷积层；所述第一层的路由模块和所述第二层的路由模块为前后相邻的两个计算模块。

在一些实施例中，一个卷积模块包括若干数量的卷积层或卷积计算单元；一个卷积层可以为一组平行的特征图，通过在输入图像上滑动不同的卷积并执行一定的运算而组成；在每一个滑动的位置上，卷积核与输入图像之间会执行一个元素对应乘积并求和的运算，以将感受野内的信息投影到特征图中的一个元素。其中，卷积核的尺寸小于输入图像的尺寸，且重叠或平行的作用于输入图像中，双流神经网络模型中间的每一层卷积模块输出的特征图中的所有元素都是通过一个卷积核计算得出。

另外，双流神经网络模型还包括全连接层、第一损失函数以及第二损失函数。如图4所示，卷积模块 Conv5\_x 输出的特征作为一个全连接层的输入，最后一层的路由模块的输出特征作为一个全连接层的输入，将两个全连接层的结果相加，作为总的全连接层的输出，得到第一预测结果和第二预测结果。

在一些实施例中，路由模块包括：第一卷积单元、第一归一化单元、第一激活单元、第二卷积单元、第二归一化单元、第二激活单元；通过路由模块的第一卷积单元、第一归一化单元、第一激活单元、第二卷积单元、第二归一化单元、第二激活单元，依次对第一神经网络模型的卷积计算模块输出的特征矩阵和第二神经网络模型的卷积计算模块输出的特征矩阵进行交互学习，得到路

由模块输出的特征矩阵。

如图 5 所示，本申请实施例提供的路由模块的架构示意图。路由模块包括两层卷积单元、两层归一化单元以及两层激活单元；分别可以为第一卷积单元 Conv1D、第一归一化单元 Batch Normlization、第一激活单元 ReLU、第二卷积单元 Conv1D、第二归一化单元 Batch Normlization、第二激活单元 ReLU。将双流神经网络模型中间卷积模块的每一层的双路的卷积模块的输出作为对应的路由模块的输入，将每一层路由模块的输出作为下一层卷积模块的输入或全连接层的输入。其中，路由模块可以为 1\*1 的卷积为基础的计算单元；上一层的双路的卷积模块的输出通过 1\*1 卷积的学习和重定向之后，输出给后续层的卷积模块。双路的卷积模块的输出可以为多模态的图像特征的信息流，例如 RGB 格式的信息流和深度图像特征的信息流等。

在一些实施例中，第一神经网络模型包括第一损失函数，第二神经网络模型包括第二损失函数；通过视频样本数据对第一神经网络模型、第二神经网络模型和路由模块进行训练，依据第一损失函数和第二损失函数分别调整第一神经网络模型的参数、第二神经网络模型的参数以及路由模块的参数；若第一损失函数和第二损失函数满足预设阈值，则停止对第一神经网络模型的参数、第二神经网络模型的参数以及路由模块的训练，得到训练后的双流神经网络模型。

在一些实施例中，双流神经网络模型通过双损失函数进行模型的优化和训练。根据第一路的卷积神经网络的全连接层输出结果，通过第一损失函数对第一路的卷积神经网络的参数进行训练及调整；根据第二路的卷积神经网络的全连接层输出结果，通过第二损失函数对第二路的卷积神经网络的参数进行训练及调整；同时通过第一损失函数和第二损失函数训练并调整路由模块的参数。

步骤 S104，基于第一预测结果和第二预测结果，确定所述待识别动作的分类结果。

在一些实施例中，第一预测结果和第二预测结果为训练后的神经网络模型输出的多模态的双流高层特征。对双流高层特征进行特征融合，得到双流神经网络模型的网络架构中最后的输出结果。最后的输出结果为一个一维的得分向量（概率），最终分类结果根据得分向量中概率最大的进行确定；即得分最大的对应的类别为待识别动作的分类结果。

在一些实施例中，基于第一预测结果和第二预测结果，确定待识别动作的分类结果，包括：

D1、对所述第一预测结果和所述第二预测结果进行特征融合，得到动作类别的概率分布；

5 D2、将所述概率分布中概率最大的动作类别作为所述待识别动作的所述分类结果。

在一些实施例中，特征融合是双流神经网络模型的网络架构中的一个计算过程，即双流神经网络模型得到了 RGB 格式的信息流和深度信息流的特征信息后，会进行融合，融合后进行概率映射，最后进行类别判断。例如，最后的输出结果为一个一维的得分向量(概率)，得分向量为包含 10 个元素的一维向量，  
10 每个元素都是 0 到 1 的概率，10 个元素的和为 1，假设第二个元素最大为 0.3，则判定待识别动作的分类结果为第二类。

其中，特征融合的过程可以通过将网络架构最后输出的两个矩阵进行点乘、加权相加或最大值等方式进行融合计算，得到最终的概率分布，根据概率分布  
15 中的最大值对应的类别确定待识别动作的种类。

通过本申请实施例，终端设备可以获取待识别动作的视频数据，视频数据包括第一视频序列和第二视频序列，分别对第一视频序列和第二视频序列的压缩得到第一运动图和第二运动图，对视频数据进行更丰富的时空表示，使得信息表示更全，特征更丰富；从而将第一运动图和第二运动图作为双流神经网络  
20 模型的输入，通过神经网络模型对多模态的图像特征进行交互学习，提高了动作识别的准确度。

应理解，以上所述实施例仅用以说明本申请的技术方案，而非对其限制；对前述各实施例所记载的技术方案的修改，或者对其中部分技术特征的等同替换；例如增加模型的维度，增加多个模态的视频序列的特征作为模型的输入，  
25 将双流神经网络模型修改为多路独立的卷积神经网络模型和路由模块，对多个模态的视频序列的特征进行交互学习等，属于类似的发明构思，并不使相应技术方案的本质脱离本申请各实施例技术方案的精神和范围，均应包含在本申请的保护范围之内。

还应理解，上述实施例中各步骤的序号的大小并不意味着执行顺序的先后，

各过程的执行顺序应以其功能和内在逻辑确定，而不应对本申请实施例的实施过程构成任何限定。

对应于上文实施例所述的动作识别方法，图 6 示出了本申请实施例提供的动作识别装置的结构框图，为了便于说明，仅示出了与本申请实施例相关的部分。

参照图 6，该装置包括：

获取单元 61，用于获取待识别动作的视频数据，所述视频数据包括第一视频序列和第二视频序列；

处理单元 62，用于将所述第一视频序列和所述第二视频序列分别进行压缩处理，得到所述第一视频序列对应的第一运动图和所述第二视频序列对应的第二运动图；

计算单元 63，用于将所述第一运动图和所述第二运动图输入训练后的双流神经网络模型，通过所述训练后的双流神经网络模型对所述第一运动图的特征和所述第二运动图的特征进行交互学习，得到所述第一视频序列的第一预测结果和所述第二视频序列的第二预测结果；

输出单元 64，用于基于所述第一预测结果和所述第二预测结果，确定所述待识别动作的分类结果。

通过本申请实施例，终端设备可以获取待识别动作的视频数据，视频数据包括第一视频序列和第二视频序列，分别对第一视频序列和第二视频序列的压缩得到第一运动图和第二运动图，对视频数据进行更丰富的时空表示，使得信息表示更全，特征更丰富；从而将第一运动图和第二运动图作为双流神经网络模型的输入，通过神经网络模型对多模态的图像特征进行交互学习，提高了动作识别的准确度。

需要说明的是，上述装置/单元之间的信息交互、执行过程等内容，由于与本申请方法实施例基于同一构思，其具体功能及带来的技术效果，具体可参见方法实施例部分，此处不再赘述。

所属领域的技术人员可以清楚地了解到，为了描述的方便和简洁，仅以上述各功能单元、模块的划分进行举例说明，实际应用中，可以根据需要而将上述功能分配由不同的功能单元、模块完成，即将所述装置的内部结构划分成不

同的功能单元或模块，以完成以上描述的全部或者部分功能。实施例中的各功能单元、模块可以集成在一个处理单元中，也可以是各个单元单独物理存在，也可以两个或两个以上单元集成在一个单元中，上述集成的单元既可以采用硬件的形式实现，也可以采用软件功能单元的形式实现。另外，各功能单元、模块的具体名称也只是为了便于相互区分，并不用于限制本申请的保护范围。上述系统中单元、模块的具体工作过程，可以参考前述方法实施例中的对应过程，在此不再赘述。

图 7 为本申请一实施例提供的终端设备的结构示意图。如图 7 所示，该实施例的终端设备 7 包括：至少一个处理器 70（图 7 中仅示出一个）处理器、存储器 71 以及存储在所述存储器 71 中并可在所述至少一个处理器 70 上运行的计算机程序 72，所述处理器 70 执行所述计算机程序 72 时实现上述任意各个训练板的识别方法实施例中的步骤。

该终端设备 7 可包括，但不仅限于，处理器 70、存储器 71。本领域技术人员可以理解，图 7 仅仅是终端设备 7 的举例，并不构成对终端设备 7 的限定，可以包括比图示更多或更少的部件，或者组合某些部件，或者不同的部件，例如还可以包括输入输出设备、网络接入设备等。

所称处理器 70 可以是中央处理单元 (Central Processing Unit, CPU)，该处理器 70 还可以是其他通用处理器、数字信号处理器 (Digital Signal Processor, DSP)、专用集成电路 (Application Specific Integrated Circuit, ASIC)、现场可编程门阵列 (Field-Programmable Gate Array, FPGA) 或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件等。通用处理器可以是微处理器或者该处理器也可以是任何常规的处理器等。

所述存储器 71 在一些实施例中可以是所述终端设备 7 的内部存储单元，例如终端设备 7 的硬盘或内存。所述存储器 71 在另一些实施例中也可以是所述终端设备 7 的外部存储设备，例如所述终端设备 7 上配备的插接式硬盘，智能存储卡 (Smart Media Card, SMC)，安全数字 (Secure Digital, SD) 卡，闪存卡 (Flash Card) 等。进一步地，所述存储器 71 还可以既包括所述终端设备 7 的内部存储单元也包括外部存储设备。所述存储器 71 用于存储操作系统、应用程序、引导装载程序 (BootLoader)、数据以及其他程序等，例如所述计算机程序的程序代码

等。所述存储器 71 还可以用于暂时地存储已经输出或者将要输出的数据。

本申请实施例还提供了一种计算机可读存储介质，所述计算机可读存储介质存储有计算机程序，所述计算机程序被处理器执行时实现可实现上述各个方法实施例中的步骤。

5 本申请实施例提供了一种计算机程序产品，当计算机程序产品在移动终端上运行时，使得移动终端执行时实现可实现上述各个方法实施例中的步骤。

所述集成的单元如果以软件功能单元的形式实现并作为独立的产品销售或使用时，可以存储在一个计算机可读取存储介质中。基于这样的理解，本申请实现上述实施例方法中的全部或部分流程，可以通过计算机程序来指令相关的  
10 硬件来完成，所述的计算机程序可存储于一计算机可读存储介质中，该计算机程序在被处理器执行时，可实现上述各个方法实施例的步骤。其中，所述计算机程序包括计算机程序代码，所述计算机程序代码可以为源代码形式、对象代码形式、可执行文件或某些中间形式等。所述计算机可读介质至少可以包括：  
15 能够将计算机程序代码携带到拍照装置/终端设备的任何实体或装置、记录介质、计算机存储器、只读存储器 (ROM, Read-Only Memory)、随机存取存储器 (RAM, Random Access Memory)、电载波信号、电信信号以及软件分发介质。例如 U 盘、移动硬盘、磁碟或者光盘等。在某些司法管辖区，根据立法和专利实践，计算机可读介质不可以是电载波信号和电信信号。

在上述实施例中，对各个实施例的描述都各有侧重，某个实施例中  
20 没有详述或记载的部分，可以参见其它实施例的相关描述。

本领域普通技术人员可以意识到，结合本文中所公开的实施例描述的各示例的单元及算法步骤，能够以电子硬件、或者计算机软件和电子硬件的结合来实现。这些功能究竟以硬件还是软件方式来执行，取决于技术方案的特定应用和设计约束条件。专业技术人员可以对每个特定的应用来使用不同方法来实现  
25 所描述的功能，但是这种实现不应认为超出本申请的范围。

在本申请所提供的实施例中，应该理解到，所揭露的装置/网络设备和方法，可以通过其它的方式实现。例如，以上所描述的装置/网络设备实施例仅仅是示意性的，例如，所述模块或单元的划分，仅仅为一种逻辑功能划分，实际实现时可以有另外的划分方式，例如多个单元或组件可以结合或者可以集成到另一

个系统，或一些特征可以忽略，或不执行。另一点，所显示或讨论的相互之间的耦合或直接耦合或通讯连接可以是通过一些接口，装置或单元的间接耦合或通讯连接，可以是电性，机械或其它的形式。

5 所述作为分离部件说明的单元可以是或者也可以不是物理上分开的，作为单元显示的部件可以是或者也可以不是物理单元，即可以位于一个地方，或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部单元来实现本实施例方案的目的。

10 以上所述实施例仅用以说明本申请的技术方案，而非对其限制；尽管参照前述实施例对本申请进行了详细的说明，本领域的普通技术人员应当理解：其依然可以对前述各实施例所记载的技术方案进行修改，或者对其中部分技术特征进行等同替换；而这些修改或者替换，并不使相应技术方案的本质脱离本申请各实施例技术方案的精神和范围，均应包含在本申请的保护范围之内。

## 权利要求书

1. 一种基于特征交互学习的动作识别方法，其特征在于，包括：

获取待识别动作的视频数据，所述视频数据包括第一视频序列和第二视频序列；

5 将所述第一视频序列和所述第二视频序列分别进行压缩处理，得到所述第一视频序列对应的第一运动图和所述第二视频序列对应的第二运动图；

将所述第一运动图和所述第二运动图输入训练后的双流神经网络模型，通过所述训练后的双流神经网络模型对所述第一运动图的特征和所述第二运动图的特征进行交互学习，得到所述训练后的双流神经网络模型输出的所述第一视频序列的第一预测结果和所述第二视频序列的第二预测结果；

10 基于所述第一预测结果和所述第二预测结果，确定所述待识别动作的分类结果。

2. 如权利要求 1 所述的方法，其特征在于，所述将所述第一视频序列进行压缩处理，得到所述第一视频序列对应的第一运动图，包括：

15 获取所述第一视频序列中每一视频帧的特征矩阵；

根据所述第一视频序列中视频帧的时序，将每一视频帧的所述特征矩阵进行压缩计算，得到用于表示所述第一运动图的特征矩阵。

3. 如权利要求 1 所述的方法，其特征在于，所述将所述第二视频序列进行压缩处理，得到所述第二视频序列对应的第二运动图，包括：

20 将所述第二视频序列进行灰度处理，得到所述第二视频序列对应的灰度序列帧；

根据所述第二视频序列中视频帧的时序，将所述灰度序列帧的特征矩阵进行压缩计算，得到用于表示所述第二运动图的特征矩阵。

4. 如权利要求 1 所述的方法，其特征在于，所述训练后的双流神经网络模型包括第一神经网络模型、第二神经网络模型和路由模块，所述路由模块设置于所述第一神经网络模型的中间卷积模块和所述第二神经网络模型的中间卷积模块之间；

所述第一神经网络模型的输入为所述第一运动图，输出为所述第一视频序列的所述第一预测结果；

所述第二神经网络模型的输入为所述第二运动图，输出为所述第二视频序列的所述第二预测结果；

所述路由模块用于在所述第一神经网络模型的中间卷积模块和所述第二神经网络模型的中间卷积模块之间，对所述第一运动图的特征和所述第二运动图的特征进行交互学习。

5. 如权利要求4所述的方法，其特征在于，所述第一神经网络模型的中间卷积模块包括预设层数的第一卷积模块，所述第二神经网络模型的中间卷积模块包括与所述第一卷积模块相对应的第二卷积模块；

所述将所述第一运动图和所述第二运动图输入训练后的双流神经网络模型，通过所述训练后的双流神经网络模型对所述第一运动图的特征和所述第二运动图的特征进行交互学习，得到所述第一视频序列的第一预测结果和所述第二视频序列的第二预测结果，包括：

将第一层的第一卷积模块的输出和第一层的第二卷积模块的输出作为第一层的路由模块的输入，由所述第一层的路由模块进行特征交互学习，得到第一路由输出；

将所述第一层的第一卷积模块的输出与所述第一路由输出的叠加结果作为第二层的第一卷积模块的输入，由所述第二层的第一卷积模块进行特征学习，得到所述第二层的第一卷积模块的输出；

将所述第一层的第二卷积模块的输出与所述第一路由输出的叠加结果作为第二层的第二卷积模块的输入，由所述第二层的第二卷积模块进行特征学习，得到所述第二层的第二卷积模块的输出；

将所述第二层的第一卷积模块的输出和所述第二层的第二卷积模块的输出作为第二层的路由模块的输入，由所述第二层的路由模块进行特征交互学习，得到第二路由输出；

其中，所述第一层的第一卷积模块与第二层的第一卷积模块为所述第一神经网络模型的中间卷积模块中前后相邻的两层卷积层；所述第一层的第二卷积模块和所述第二层的第二卷积模块为所述第二神经网络模型的中间卷积模块中前后相邻的两层卷积层；所述第一层的路由模块和所述第二层的路由模块为前后相邻的两个计算模块。

6. 如权利要求 4 所述的方法，其特征在于，所述路由模块包括：第一卷积单元、第一归一化单元、第一激活单元、第二卷积单元、第二归一化单元、第二激活单元；

5 通过所述路由模块的所述第一卷积单元、所述第一归一化单元、所述第一激活单元、所述第二卷积单元、所述第二归一化单元、所述第二激活单元，依次对所述第一神经网络模型的卷积计算模块输出的特征矩阵和所述第二神经网络模型的卷积计算模块输出的特征矩阵进行交互学习，得到所述路由模块输出的特征矩阵。

10 7. 如权利要求 1 至 6 任一项所述的方法，其特征在于，所述基于所述第一预测结果和所述第二预测结果，确定所述待识别动作的分类结果，包括：

对所述第一预测结果和所述第二预测结果进行特征融合，得到动作类别的概率分布；

将所述概率分布中概率最大的动作类别作为所述待识别动作的所述分类结果。

15 8. 如权利要求 4 所述的方法，其特征在于，所述第一神经网络模型包括第一损失函数，所述第二神经网络模型包括第二损失函数；

20 通过样本视频数据对所述第一神经网络模型、所述第二神经网络模型和所述路由模块进行训练，依据所述第一损失函数和所述第二损失函数分别调整所述第一神经网络模型的参数、所述第二神经网络模型的参数以及所述路由模块的参数；

若所述第一损失函数和所述第二损失函数满足预设阈值，则停止对所述第一神经网络模型的参数、所述第二神经网络模型的参数以及所述路由模块的训练，得到所述训练后的双流神经网络模型。

25 9. 一种终端设备，其特征在于，包括存储器、处理器以及存储在所述存储器中并可在所述处理器上运行的计算机程序，所述处理器执行所述计算机程序时实现如权利要求 1 至 8 任一项所述的方法。

10. 一种计算机可读存储介质，所述计算机可读存储介质存储有计算机程序，其特征在于，所述计算机程序被处理器执行时实现如权利要求 1 至 8 任一项所述的方法。

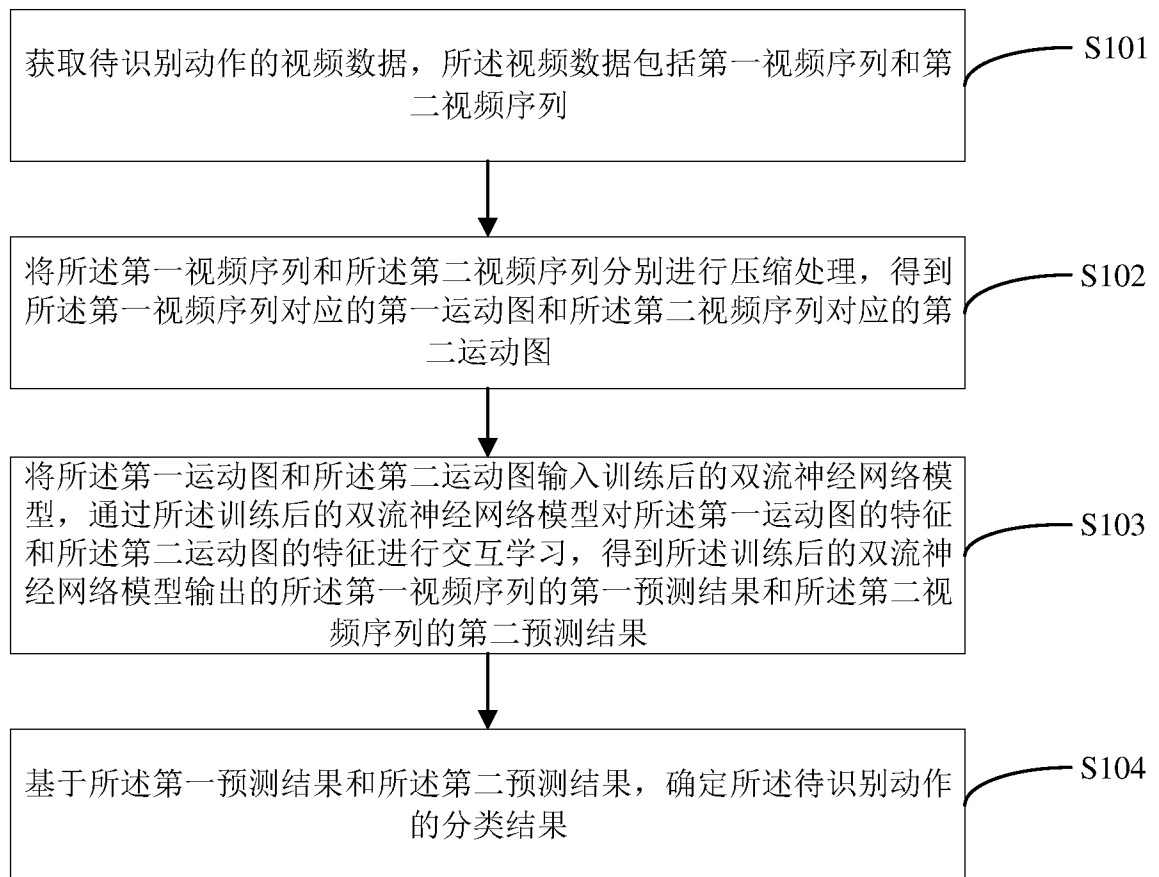


图 1

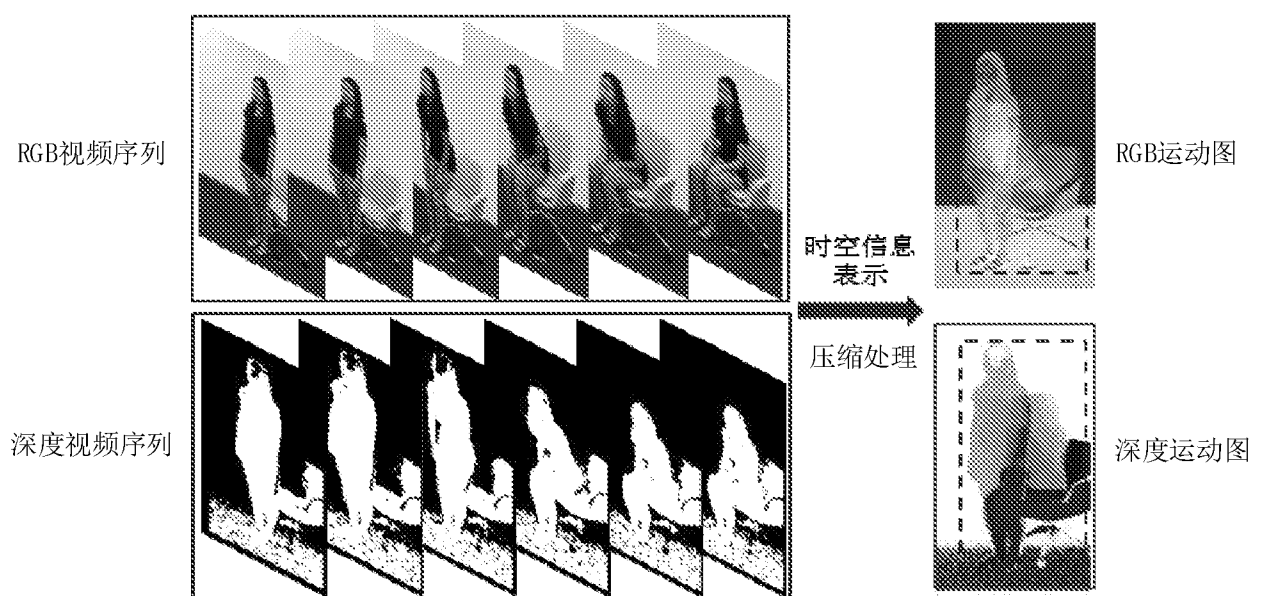


图 2

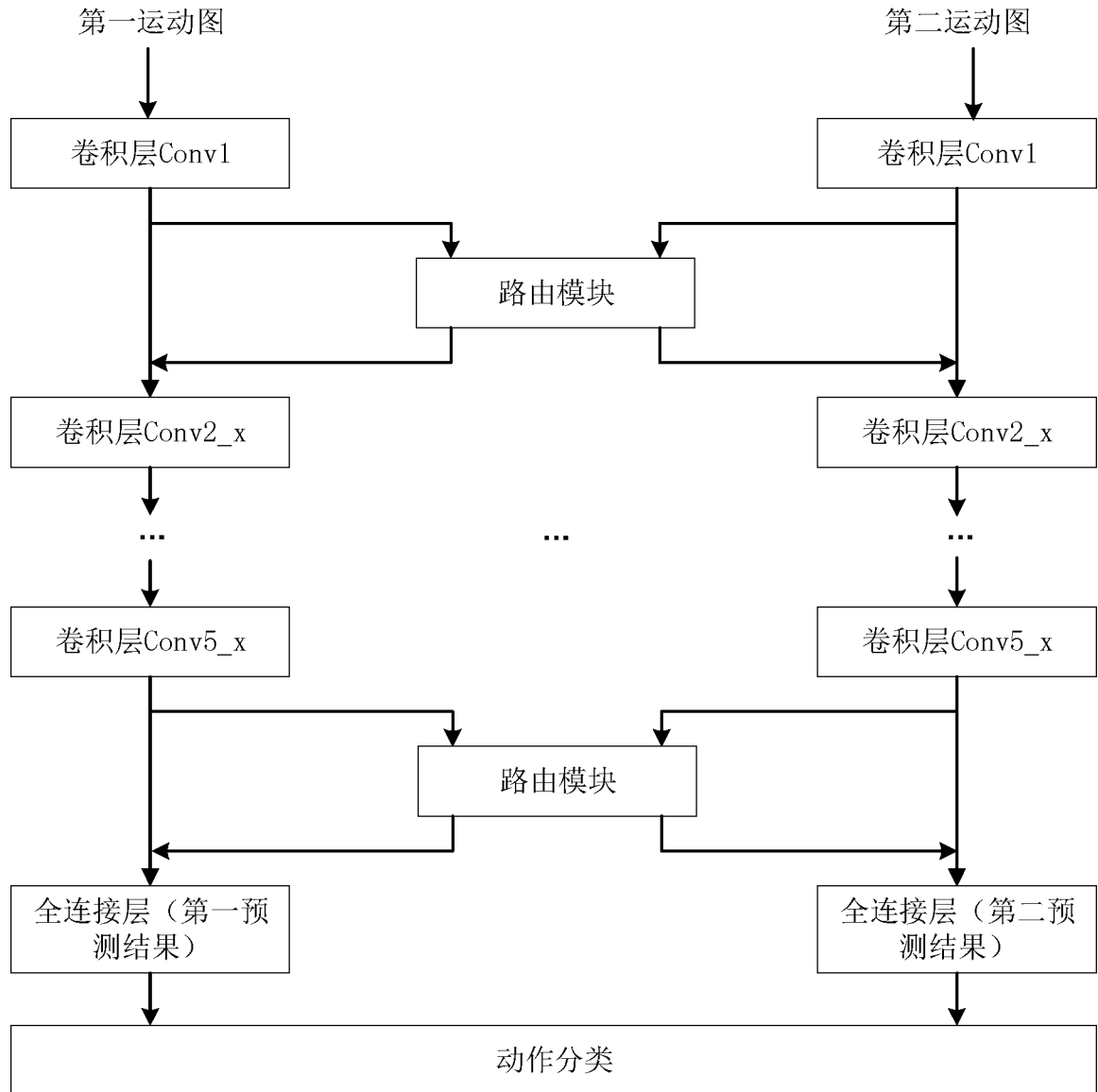


图 3

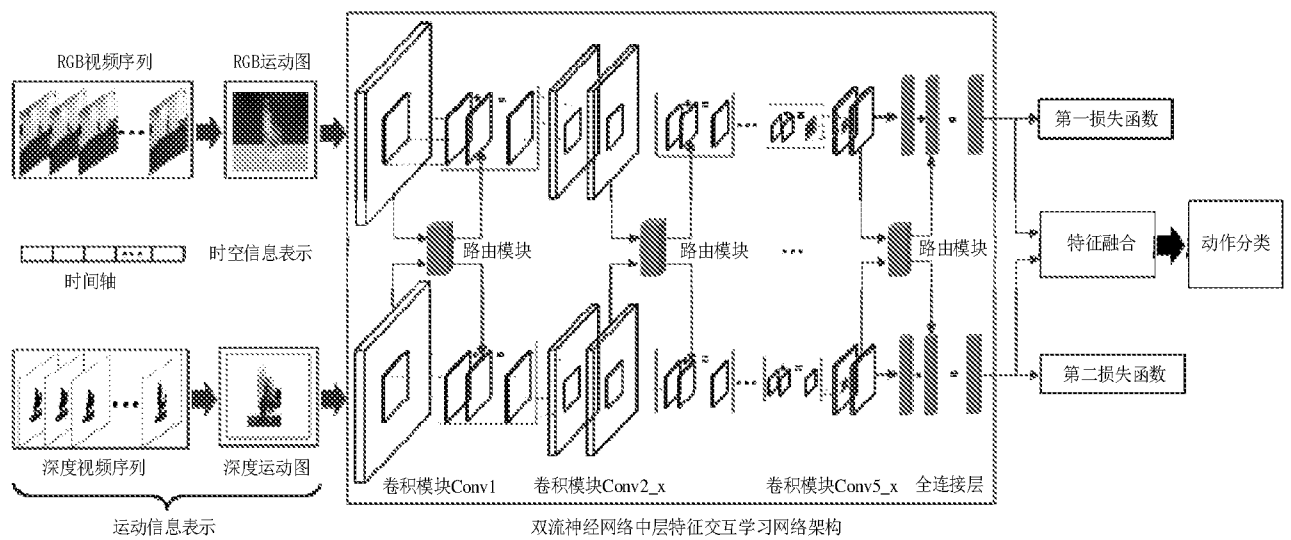


图 4

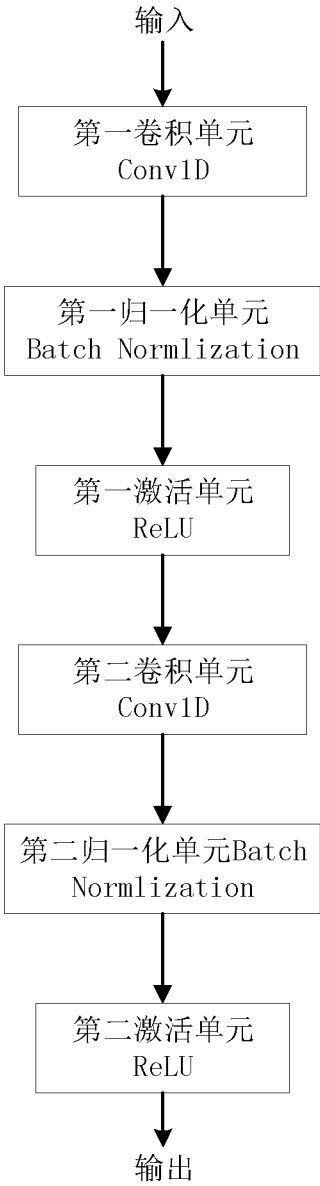


图 5

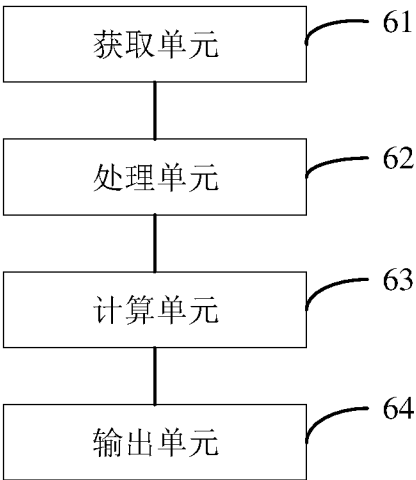


图 6

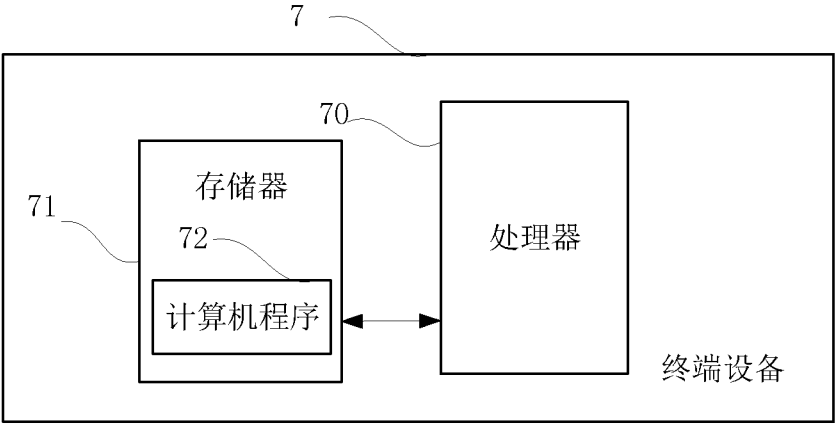


图 7

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2020/129550

**A. CLASSIFICATION OF SUBJECT MATTER**

G06K 9/00(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

G06K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

WPI, EPODOC, CNPAT, CNKI: 动作识别, 交互学习, 视频, 压缩, 运动图, 预测, 分类, action w recogni+, interactive w learn+, video, compress+, motion w graph, predict+, classif+

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

| Category* | Citation of document, with indication, where appropriate, of the relevant passages                                                                         | Relevant to claim No. |
|-----------|------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------|
| Y         | CN 110633630 A (SHENZHEN INSTITUTES OF ADVANCED TECHNOLOGY, CHINESE ACADEMY OF SCIENCES) 31 December 2019 (2019-12-31)<br>description paragraphs 0044-0181 | 1-3, 7, 9-10          |
| Y         | CN 110020658 A (DALIAN UNIVERSITY OF TECHNOLOGY) 16 July 2019 (2019-07-16)<br>description, paragraphs 0004-0012                                            | 1-3, 7, 9-10          |
| A         | CN 111199238 A (SF TECHNOLOGY CO., LTD.) 26 May 2020 (2020-05-26)<br>entire document                                                                       | 1-10                  |
| A         | CN 110175596 A (CHONGQING UNIVERSITY OF POSTS AND TELECOMMUNICATIONS) 27 August 2019 (2019-08-27)<br>entire document                                       | 1-10                  |



Further documents are listed in the continuation of Box C.



See patent family annex.

\* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

08 June 2021

Date of mailing of the international search report

24 June 2021

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/  
CN)  
No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing  
100088  
China

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

**INTERNATIONAL SEARCH REPORT**  
**Information on patent family members**

International application No.

**PCT/CN2020/129550**

| Patent document<br>cited in search report |           |   | Publication date<br>(day/month/year) | Patent family member(s) | Publication date<br>(day/month/year) |
|-------------------------------------------|-----------|---|--------------------------------------|-------------------------|--------------------------------------|
| CN                                        | 110633630 | A | 31 December 2019                     | None                    |                                      |
| CN                                        | 110020658 | A | 16 July 2019                         | None                    |                                      |
| CN                                        | 111199238 | A | 26 May 2020                          | None                    |                                      |
| CN                                        | 110175596 | A | 27 August 2019                       | None                    |                                      |

## 国际检索报告

国际申请号

PCT/CN2020/129550

## A. 主题的分类

G06K 9/00 (2006.01) i

按照国际专利分类 (IPC) 或者同时按照国家分类和 IPC 两种分类

## B. 检索领域

检索的最低限度文献 (标明分类系统和分类号)

G06K

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库 (数据库的名称, 和使用的检索词 (如使用))

WPI, EPDOC, CNPAT, CNKI: 动作识别, 交互学习, 视频, 压缩, 运动图, 预测, 分类, action w recogni+, interactive w learn+, video, compress+, motion w graph, predict+, classif+

## C. 相关文件

| 类 型* | 引用文件, 必要时, 指明相关段落                                                                | 相关的权利要求      |
|------|----------------------------------------------------------------------------------|--------------|
| Y    | CN 110633630 A (中国科学院深圳先进技术研究院) 2019年 12月 31日 (2019 - 12 - 31)<br>说明书第0044-0181段 | 1-3, 7, 9-10 |
| Y    | CN 110020658 A (大连理工大学) 2019年 7月 16日 (2019 - 07 - 16)<br>说明书第0004-0012段          | 1-3, 7, 9-10 |
| A    | CN 111199238 A (顺丰科技有限公司) 2020年 5月 26日 (2020 - 05 - 26)<br>全文                    | 1-10         |
| A    | CN 110175596 A (重庆邮电大学) 2019年 8月 27日 (2019 - 08 - 27)<br>全文                      | 1-10         |

☐ 其余文件在C栏的续页中列出。☒ 见同族专利附件。

\* 引用文件的具体类型:

“A” 认为不特别相关的表示了现有技术一般状态的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的)

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性

“&amp;” 同族专利的文件

国际检索实际完成的日期

2021年 6月 8日

国际检索报告邮寄日期

2021年 6月 24日

ISA/CN的名称和邮寄地址

中国国家知识产权局 (ISA/CN)

中国北京市海淀区蓟门桥西土城路6号 100088

传真号 (86-10)62019451

受权官员

焦永涵

电话号码 86-(10)-53961463

国际检索报告  
关于同族专利的信息

国际申请号

PCT/CN2020/129550

| 检索报告引用的专利文件 |           |   | 公布日<br>(年/月/日) | 同族专利 | 公布日<br>(年/月/日) |
|-------------|-----------|---|----------------|------|----------------|
| CN          | 110633630 | A | 2019年 12月 31日  | 无    |                |
| CN          | 110020658 | A | 2019年 7月 16日   | 无    |                |
| CN          | 111199238 | A | 2020年 5月 26日   | 无    |                |
| CN          | 110175596 | A | 2019年 8月 27日   | 无    |                |