

(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
2020 年 11 月 26 日 (26.11.2020)

(10) 国际公布号
WO 2020/232886 A1

(51) 国际专利分类号:
G06K 9/00 (2006.01) **G06N 3/08** (2006.01)

(21) 国际申请号: PCT/CN2019/103174

(22) 国际申请日: 2019 年 8 月 29 日 (29.08.2019)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201910422559.6 2019 年 5 月 21 日 (21.05.2019) CN

(71) 申请人: 平安科技(深圳)有限公司(PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) [CN/CN]; 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。

(72) 发明人: 周俊琨(ZHOU, Junkun); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。 罗郑楠(LUO, Zhengnan); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。 官民(GUAN, Min); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。 许扬(XU, Yang); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。

(74) 代理人: 深圳中一联合知识产权代理有限公司(SHENZHEN ZHONGYI UNION INTELLECTUAL PROPERTY AGENCY CO., LTD.); 中国广东省深圳市福田区园岭街道深南中路 1014 号报春大厦 9 楼(5 号信箱), Guangdong 518028 (CN)。

(54) Title: VIDEO BEHAVIOR IDENTIFICATION METHOD AND APPARATUS, STORAGE MEDIUM AND SERVER

(54) 发明名称: 一种视频行为识别方法、装置、存储介质和服务服务器

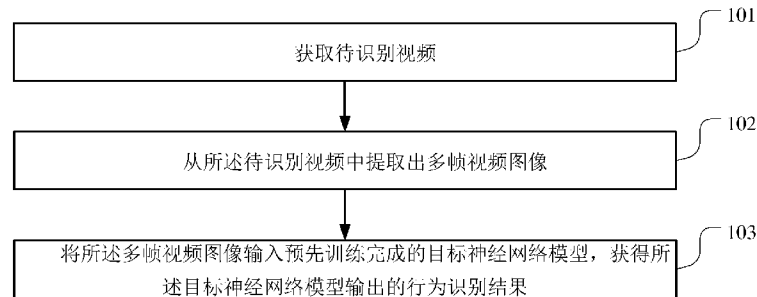


图 1

- 101 Acquire a video to be identified
102 Extract multiple video images from the video to be identified
103 Input the multiple video images into a pre-trained target neural network model, so as to obtain a behavior identification result output by the target neural network model

(57) Abstract: The present application relates to the field of computer technology, and proposes a video behavior recognition method and apparatus, a storage medium and a server. Said method comprises: firstly acquiring a video to be identified, and extracting multiple video images from the video to be identified; and then inputting the multiple video images into a pre-trained neural network model, so as to obtain a behavior identification result output by the neural network model. In the present application, the following data processing logic is added to the structure of the neural network model: acquiring image processing data of the multiple video images, the image processing data being a four-dimensional matrix containing global image feature information; performing convolution processing on

(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, 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, 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, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, 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))。

the four-dimensional matrix to obtain a four-dimensional matrix containing local image feature information; and respectively tiling the four-dimensional matrix containing the global image feature information and the four-dimensional matrix containing the local image feature information, and then multiplying same, so that the local feature and the global feature of the video image are combined. The present invention can improve the accuracy of video behavior identification.

(57) 摘要: 本申请涉及计算机技术领域, 提出一种视频行为识别方法、装置、存储介质和服务器。首先获取待识别视频, 从所述待识别视频中提取出多帧视频图像; 然后将所述多帧视频图像输入预先训练完成的神经网络模型, 以获得该神经网络模型输出的行为识别结果。本申请在该神经网络模型的结构中添加了以下数据处理逻辑: 获取所述多帧视频图像的图像处理数据, 该图像处理数据为包含图像全局特征信息的四维矩阵, 通过该四维矩阵进行卷积处理, 能够得到包含图像局部特征信息的四维矩阵, 然后将该包含图像全局特征信息的四维矩阵和该包含图像局部特征信息的四维矩阵分别平铺后再相乘, 从而将视频图像的局部特征与全局特征相结合, 能够提高视频行为识别的准确率。

一种视频行为识别方法、装置、存储介质和服务器

[0001] 本申请要求于2019年5月21日提交中国专利局、申请号为201910422559.6、申请名称为“一种视频行为识别方法、装置、存储介质和服务器”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

[0002] 本申请涉及计算机技术领域，尤其涉及一种视频行为识别方法、装置、存储介质和服务器。

背景技术

[0003] 目前，市面上的图片分类技术和基于图片的定位技术较为成熟，但是视频的行为识别无法通过一张图片来完成，而是需要通过连续的多帧视频图像来判断行为种类。例如：判断某个视频中的舞蹈类型是伦巴、恰恰还是广场舞。

[0004] 在对视频行为进行识别时，一般是将视频文件输入预先训练好的神经网络模型中，直接得到该视频文件的行为识别结果。然而，现有的神经网络模型的训练过程仅考虑了视频样本的全局特征，没有考虑相应的局部特征，这导致视频行为识别的准确率较低。

发明概述

技术问题

[0005] 本申请实施例提供了一种视频行为识别方法、装置、存储介质和服务器，以解决视频行为识别的准确率低的问题。

问题的解决方案

技术解决方案

[0006] 本申请实施例的第一方面，提供了一种视频行为识别方法，包括：

[0007] 获取待识别视频；

[0008] 从所述待识别视频中提取出多帧视频图像；

[0009] 将所述多帧视频图像输入预先训练完成的目标神经网络模型，获得所述目标神经网络模型输出的行为识别结果；

- [0010] 其中，所述目标神经网络模型包括第一处理层和第二处理层，所述第一处理层和所述第二处理层为所述目标神经网络模型的网络结构中相邻的两个卷积层，所述目标神经网络将所述多帧视频图像转换为四个维度的图像数据进行处理，所述第一处理层和第二处理层之间的处理逻辑如下：
- [0011] 获取所述第一处理层输出的四维矩阵格式的第一图像处理数据，所述第一处理层用于接收上一个处理层输出的图像处理数据，对接收到的图像处理数据执行卷积处理，得到所述第一图像处理数据；
- [0012] 对所述第一图像处理数据进行矩阵的平铺处理，获得二维矩阵格式的第二图像处理数据；
- [0013] 对所述第一图像数据进行卷积处理后，执行矩阵的平铺处理，获得二维矩阵格式的第三图像处理数据；
- [0014] 将所述第三图像处理数据和所述第二图像处理数据相乘，获得二维矩阵格式的第四图像处理数据；
- [0015] 对所述第四图像处理数据进行矩阵的逆平铺处理，得到四维矩阵格式的第五图像处理数据；
- [0016] 将所述第五图像处理数据输入所述第二处理层，所述第二处理层用于对所述第五图像处理数据执行卷积处理，然后将处理后的数据输入下一个处理层。
- [0017] 本申请实施例的第二方面，提供了一种视频行为识别装置，包括：
- [0018] 视频获取模块，用于获取待识别视频；
- [0019] 视频图像提取模块，用于从所述待识别视频中提取出多帧视频图像；
- [0020] 行为识别模块，用于将所述多帧视频图像输入预先训练完成的目标神经网络模型，获得所述目标神经网络模型输出的行为识别结果；
- [0021] 其中，所述目标神经网络模型包括第一处理层和第二处理层，所述第一处理层和所述第二处理层为所述目标神经网络模型的网络结构中相邻的两个卷积层，所述目标神经网络将所述多帧视频图像转换为四个维度的图像数据进行处理，所述第一处理层和第二处理层之间的处理逻辑如下：
- [0022] 获取所述第一处理层输出的四维矩阵格式的第一图像处理数据，所述第一处理层用于接收上一个处理层输出的图像处理数据，对接收到的图像处理数据执行卷

积处理，得到所述第一图像处理数据；

[0023] 对所述第一图像处理数据进行矩阵的平铺处理，获得二维矩阵格式的第二图像处理数据；

[0024] 对所述第一图像数据进行卷积处理后，执行矩阵的平铺处理，获得二维矩阵格式的第三图像处理数据；

[0025] 将所述第三图像处理数据和所述第二图像处理数据相乘，获得二维矩阵格式的第四图像处理数据；

[0026] 对所述第四图像处理数据进行矩阵的逆平铺处理，得到四维矩阵格式的第五图像处理数据；

[0027] 将所述第五图像处理数据输入所述第二处理层，所述第二处理层用于对所述第五图像处理数据执行卷积处理，然后将处理后的数据输入下一个处理层。

[0028] 本申请实施例的第三方面，提供了一种计算机非易失性可读存储介质，所述计算机非易失性可读存储介质存储有计算机可读指令，所述计算机可读指令被处理器执行时实现如本申请实施例的第一方面提出的视频行为识别方法的步骤。

[0029] 本申请实施例的第四方面，提供了一种服务器，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可读指令，所述处理器执行所述计算机可读指令时实现如本申请实施例的第一方面提出的视频行为识别方法的步骤。

发明的有益效果

有益效果

[0030] 本申请实施例提出的视频行为识别方法，首先获取待识别视频，从所述待识别视频中提取出多帧视频图像；然后将所述多帧视频图像输入预先训练完成的神经网络模型，以获得该神经网络模型输出的行为识别结果。本申请实施例在该神经网络模型的结构中添加了以下数据处理逻辑：获取所述多帧视频图像的图像处理数据，该图像处理数据为包含图像全局特征信息的四维矩阵，通过对该四维矩阵进行卷积处理，能够得到包含图像局部特征信息的四维矩阵，然后将该包含图像全局特征信息的四维矩阵和该包含图像局部特征信息的四维矩阵分别平铺后再相乘，从而将视频图像的局部特征与全局特征相结合，能够提高视

频行为识别的准确率。

对附图的简要说明

附图说明

[0031] 图1是本申请实施例提供的一种视频行为识别方法的第一个实施例的流程图；

[0032] 图2是本申请实施例提供的一种视频行为识别方法的第二个实施例的流程图；

[0033] 图3是本申请实施例提供的一种视频行为识别装置的一个实施例的结构图；

[0034] 图4是本申请实施例提供的一种服务器的示意图。

[0035] 本申请的实施方式

[0036] 请参阅图1，本申请实施例中一种视频行为识别方法的第一个实施例包括：

[0037] 101、获取待识别视频；

[0038] 首先，获取待识别视频，待识别视频是需要识别视频人员行为的一定长度的视频文件。

[0039] 102、从所述待识别视频中提取出多帧视频图像；

[0040] 在获取到待识别视频之后，从所述待识别视频中提取出多帧视频图像。具体的，可以采用现有技术中包含的各类视频图像提取方式，比如从视频文件中按照视频长度均匀地提取出多帧视频图像。

[0041] 103、将所述多帧视频图像输入预先训练完成的目标神经网络模型，获得所述目标神经网络模型输出的行为识别结果。

[0042] 然后，将所述多帧视频图像输入预先训练完成的目标神经网络模型，从而获得所述目标神经网络模型输出的行为识别结果。该目标神经网络模型可以采用各类已知行为类别的样本视频作为训练集训练得到，比如可以采用Kinetics行为数据集中的样本视频。其中，为了提高行为识别的准确性，该目标神经网络模型包括第一处理层和第二处理层，所述第一处理层和所述第二处理层为所述目标神经网络模型的网络结构中相邻的两个卷积层，所述目标神经网络将所述多帧视频图像转换为四个维度的图像数据进行处理，这四个维度可以是视频图像的时间、高度、宽度和特征值。所述第一处理层和第二处理层之间的处理逻辑如下：

[0043] （1）获取所述第一处理层输出的四维矩阵格式的第一图像处理数据；

[0044] (2) 对所述第一图像处理数据进行矩阵的平铺处理, 获得二维矩阵格式的第二图像处理数据;

[0045] (3) 对所述第一图像数据进行卷积处理后, 执行矩阵的平铺处理, 获得二维矩阵格式的第三图像处理数据;

[0046] (4) 将所述第三图像处理数据和所述第二图像处理数据相乘, 获得二维矩阵格式的第四图像处理数据;

[0047] (5) 对所述第四图像处理数据进行矩阵的逆平铺处理, 得到四维矩阵格式的第五图像处理数据;

[0048] (6) 将所述第五图像处理数据输入所述第二处理层。

[0049] 对于上述步骤(1), 该第一处理层是目标神经网络模型的结构中包含的某一个卷积层, 其处理的是该模型上一个处理层输出的图像处理数据, 对接收到的图像处理数据执行卷积处理, 得到所述第一图像处理数据。具体的, 该第一处理层输出的是四维矩阵格式的第一图像处理数据, 所述第一图像处理数据为所述多帧视频图像经所述第一处理层处理后得到的数据。矩阵的各个维度分别是视频图像的时间、高度、宽度和特征值, 比如可以为 $T*H*W*1024$ (时间*高度*宽度*特征值), 该第一图像处理数据具有视频图像的全局特征信息。

[0050] 对于上述步骤(2), 对所述第一图像处理数据进行矩阵的平铺处理, 获得二维矩阵格式的第二图像处理数据, 矩阵经过平铺处理, 可以实现降维, 比如一个2行5列的矩阵, 经过平铺可以得到一个1行10列的矩阵。对于 $T*H*W*1024$ 的第一图像处理数据, 经过矩阵的平铺处理, 可以得到 $THW*1024$ 的第二图像处理数据, 即从四维矩阵变为二维矩阵。

[0051] 对于上述步骤(3), 对所述第一图像数据进行卷积处理后, 执行矩阵的平铺处理, 获得二维矩阵格式的第三图像处理数据。经过卷积处理, 可以获得包含视频图像局部特征信息的第三图像处理数据。具体的, 步骤(3)可以包括:

[0052] (3.1) 将所述第一图像处理数据和预先构建的第一卷积矩阵相乘后, 执行矩阵的平铺处理, 得到二维矩阵格式的第六图像处理数据;

[0053] 所述第一图像处理数据和预先构建的第一卷积矩阵相乘, 然后执行矩阵的平铺处理, 得到二维矩阵格式的第六图像处理数据。对于 $T*H*W*1024$ 的第一图像处

理数据，和预先构建的第一卷积矩阵（ $1*1*1$ ）相乘，然后再执行矩阵的平铺处理，可以得到 $THW*1024$ 的第六图像处理数据。

[0054] （3.2）将所述第一图像处理数据和预先构建的第二卷积矩阵相乘后，执行矩阵的平铺处理并转置，得到二维矩阵格式的第七图像处理数据；

[0055] 所述第一卷积矩阵和所述第二卷积矩阵均为 $1*1*1$ 的卷积核，且具有不同的元素值。对于 $T*H*W*1024$ 的第一图像处理数据，和预先构建的第二卷积矩阵（ $1*1*1$ ）相乘，然后再执行矩阵的平铺处理，接着进行矩阵转置；或者在和预先构建的第二卷积矩阵（ $1*1*1$ ）相乘之后先执行矩阵转置，然后再执行矩阵的平铺处理，可以得到 $1024*THW$ 的第七图像处理数据。经过卷积矩阵的处理，获得的第六图像处理数据和第七图像处理数据均包含视频图像的局部特征信息。

[0056] （3.3）将所述第六图像处理数据和所述第七图像处理数据相乘，获得二维矩阵格式的所述第三图像处理数据。

[0057] 将所述第六图像处理数据和所述第七图像处理数据相乘，得到二维矩阵格式的所述第三图像处理数据。对于 $THW*1024$ 的第六图像处理数据和 $1024*THW$ 的第七图像处理数据，它们相乘之后得到（ $THW*THW$ ）的第三图像处理数据。

[0058] 对于上述步骤（4），将所述第三图像处理数据和所述第二图像处理数据相乘，获得二维矩阵格式的第四图像处理数据。对于 $THW*THW$ 的第三图像处理数据和 $THW*1024$ 的第二图像处理数据，它们相乘之后得到 $THW*1024$ 的第四图像处理数据。通过将所述第三图像处理数据和所述第二图像处理数据相乘，实现了视频图像全局特征和局部特征的结合。

[0059] 对于上述步骤（5），对所述第四图像处理数据进行矩阵的逆平铺处理，得到四维矩阵格式的第五图像处理数据。矩阵的逆平铺处理是矩阵平铺的逆操作，也可称为矩阵的展开处理，可以实现升维，比如一个1行10列的矩阵，经过逆平铺处理可以得到一个2行5列的矩阵。对 $THW*1024$ 的第四图像处理数据进行矩阵的逆平铺处理，可以得到 $T*H*W*1024$ 的第五图像处理数据。

[0060] 对于上述步骤（6），将所述第五图像处理数据输入所述第二处理层，所述第二处理层用于对所述第五图像处理数据执行卷积处理，然后将处理后的数据输入下一个处理层，以执行后续各个处理层的数据处理操作，最终得到行为识别

结果。通过在神经网络模型的网络结构中添加上述步骤（1）-（6）的处理逻辑，能够将视频图像的全局特征和局部特征相结合，从而提高视频行为识别的准确率。

[0061] 优选的，所述目标神经网络模型为RsesNet101模型，所述第一处理层为RsesNet101模型网络结构第四部分的最后一层处理层，所述第二处理层为RsesNet101模型网络结构第五部分的第一层处理层。

[0062] RsesNet是深度残差网络，包含RsesNet18、RsesNet34、RsesNet50 RsesNet101和RsesNet152等多种不同深度的网络结构，每个网络结构都包含conv1、conv2、conv3、conv4和conv5这5个部分，根据网络结构深度的不同，每个部分包含的处理层的数量也不同。本申请实施例优选采用RsesNet101神经网络模型，在第一处理层和第二处理层之间添加如上述步骤（1）至（8）所示的处理逻辑，第一处理层是RsesNet101网络结构的第四部分（即conv4）的最后一层处理层，第二处理层是RsesNet101网络结构的第五部分（即conv5）的第一层处理层。

[0063] 进一步的，所述目标神经网络模型可以通过以下步骤训练得到：

[0064] （1）采用kinetics数据集训练得到初始神经网络模型；

[0065] （2）从预先选取的样本视频中提取样本视频图像，所述样本视频的行为识别结果是已知的；

[0066] （3）将所述样本视频图像输入所述初始神经网络模型，得到样本行为识别结果；

[0067] （4）将所述样本行为识别结果和所述样本视频的行为识别结果进行比较，并根据比较的结果修正所述初始神经网络模型的参数，重复迭代直至所述样本行为识别结果和所述样本视频的行为识别结果之间的差别小于预设阈值；

[0068] （5）将参数修正完毕的所述初始神经网络模型确定为所述目标神经网络模型。

[0069] Kinetics是行为数据集，主要关注各种人类行为，包含单人行为、多人行为、人物行为等多种不同的类别。该数据集有400个类别，每个类别都有400-1000多个视频片段，可作为训练集训练得到一个初始的神经网络模型。然后，从预先选取的样本视频中提取样本视频图像，这些样本视频的行为识别结果是已知的

，比如可以是交际舞、广场舞的视频。将这些样本视频图像输入所述初始的神经网络模型，得到样本行为识别结果，接着将模型输出的样本行为识别结果和该已知的样本视频的行为识别结果进行比较，得到相应的误差，然后根据该误差修正该初始神经网络模型的参数，如何反复迭代直至一定的迭代次数或者相应的误差小于设定的阈值，最终会得到一个参数修正完毕的神经网络模型，作为所述目标神经网络模型。

- [0070] 具体的，所述从预先选取的样本视频中提取样本视频图像可以包括：
- [0071] (1) 将所述样本视频的分辨率转换为预设的分辨率；
- [0072] (2) 根据所述样本视频的视频长度选取第四数量的样本时间点；
- [0073] (3) 以每个所述样本时间点为起始点，分别从所述样本视频中提取连续的多帧视频图像，得到所述第四数量的样本视频图像组，每个所述样本视频图像组包括第五数量的视频图像；
- [0074] (4) 对所述样本视频图像组包含的每一帧视频图像分别按照左边界对齐、中部对齐和右边界对齐的方式剪切为3个视频图像；
- [0075] (5) 从每个所述样本视频图像组包含的剪切后的视频图像中分别选取第六数量的视频图像，作为提取到的样本视频图像。
- [0076] 首先将样本视频的分辨率转换为预设的分辨率，比如可以转化为统一的分辨率340*256（水平像素×垂直像素）。然后根据所述样本视频的视频长度选取第四数量的样本时间点，比如若样本视频的长度为60min，则可以选取10个样本时间点，分别为0min、6min、12min、18min、24min、30min、36min、42min、48min、54min。接着以每个所述样本时间点为起始点，分别从所述样本视频中提取连续的多帧视频图像，得到所述第四数量的样本视频图像组，每个所述样本视频图像组包括第五数量的视频图像，比如以0min为起始点，从样本视频中提取连续的64帧视频图像，以6min为起始点，从样本视频中提取连续的64帧视频图像...最终得到10个样本视频图像组，每组包含64帧视频图像。对所述样本视频图像组包含的每一帧视频图像分别按照左边界对齐、中部对齐和右边界对齐的方式剪切为3个视频图像，采用左、中、右对齐指的是分别提取图片的左侧、中间和左右侧的图片信息，可以将一张分辨率340*256的图片转化为3张224*224的

图片。经过这样处理，每一个样本视频图像组将包含64*3帧剪切后的视频图像。最后，从每个所述样本视频图像组包含的剪切后的视频图像中分别选取第六数量的视频图像，作为提取到的样本视频图像，比如可以按照每8步移动步长，提取一帧图片的方式进行提取，最终每个所述样本视频图像组得到8*3帧样本视频图像。通过将视频进行均匀分割，并提取预设帧数图片作为新的视频训练数据，既能保证提取训练数据的特征准确，又能降低视频图像的总大小，从而提高模型训练的效率。

[0077] 本申请实施例在该神经网络模型的结构中添加了以下数据处理逻辑：获取所述多帧视频图像的图像处理数据，该图像处理数据为包含图像全局特征信息的四维矩阵，通过对该四维矩阵进行卷积处理，能够得到包含图像局部特征信息的四维矩阵，然后将该包含图像全局特征信息的四维矩阵和该包含图像局部特征信息的四维矩阵分别平铺后再相乘，从而将视频图像的局部特征与全局特征相结合，能够提高视频行为识别的准确率。

[0078] 请参阅图2，本申请实施例中一种视频行为识别方法的第二个实施例包括：

[0079] 201、获取待识别视频；

[0080] 步骤201与步骤101相同，具体可参照步骤101的相关说明。

[0081] 202、将所述待识别视频的分辨率转换为预设的分辨率；

[0082] 在获得待识别视频之后，将所述待识别视频的分辨率转换为预设的分辨率，比如可以转化为统一的分辨率340*256（水平像素×垂直像素）。

[0083] 203、根据所述待识别视频的视频长度选取第一数量的时间点；

[0084] 根据所述待识别视频的视频长度选取第一数量的时间点，比如若待识别视频的长度为60min，则可以选取10个时间点，分别为0min、6min、12min、18min、24min、30min、36min、42min、48min、54min。

[0085] 204、以每个所述时间点为起始点，分别从所述待识别视频中提取连续的多帧视频图像，得到所述第一数量的视频图像组；

[0086] 接着，以每个所述时间点为起始点，分别从所述待识别视频中提取连续的多帧视频图像，得到所述第一数量的视频图像组，每个所述视频图像组包括第二数量的视频图像。比如以0min为起始点，从待识别视频中提取连续的64帧视频图

像，以6min为起始点，从待识别视频中提取连续的64帧视频图像...最终得到10个视频图像组，每组包含64帧视频图像。

[0087] 205、对所述视频图像组包含的每帧视频图像进行剪切，截取预设区域内的图像；

[0088] 然后，对所述视频图像组包含的每帧视频图像进行剪切，截取预设区域内的图像。比如，对于340*256的视频图像来说，可以剪切视频图像中部区域的图像，得到224*224的剪切后的视频图像。

[0089] 206、从每个所述视频图像组包含的剪切后的视频图像中分别选取第三数量的视频图像，作为提取到的视频图像；

[0090] 接着，从每个所述视频图像组包含的剪切后的视频图像中分别选取第三数量的视频图像，作为提取到的视频图像。比如，可以按照每8步移动步长，提取一帧图片的方式进行提取，最终每个所述视频图像组可以提取得到8帧视频图像。

[0091] 207、将所述提取到的视频图像输入预先训练完成的目标神经网络模型，获得所述目标神经网络模型输出的行为识别结果。

[0092] 最后，将所述提取到的视频图像输入预先训练完成的目标神经网络模型，获得所述目标神经网络模型输出的行为识别结果。关于步骤207的具体说明，可以参照步骤103。

[0093] 具体的，步骤207可以包括：

[0094] （1）通过所述目标神经网络模型分别获得每个所述视频图像组的行为识别结果，获取到的行为识别结果包括各个预设行为的匹配度；

[0095] （2）计算各个所述视频图像组的行为识别结果的加权平均值，加权的权重根据各个所述视频图像组所对应的所述时间点确定；

[0096] （3）将所述加权平均值中匹配度最高的行为确定为所述待识别视频的行为识别结果。

[0097] 假设有视频图像组1、2和3，各个预设的行为类别包括A、B和C，通过该目标神经网络模型获得视频图像组1的行为识别结果为A-90%（表示与行为类别A的匹配度为90%），B-30%，C-15%，视频图像组2的行为识别结果为A-50%，B-80%，C-35%，视频图像组3的行为识别结果为A-70%，B-50%，C-75%。在计算加

权平均值时，考虑视频图像组所对应的所述时间点，比如视频图像组对应的是待识别视频的中部时间点，则可以设置较高的权重值，由于大多数视频文件的主要内容处于视频的中部，故这样设置可以在一定的程度上提高视频行为识别的准确率。假设视频图像组1和视频图像组3的权重值为1，视频图像组2的权重值为1.2，则计算的加权平均值为A： $(90\%*1+50\%*1.2+70\%*1)/3=73.3\%$ ，B： $(30\%*1+80\%*1.2+50\%*1)/3=58.7\%$ ，C： $(15\%*1+35\%*1.2+75\%*1)/3=44\%$ 。最后，将所述加权平均值中匹配度最高的行为确定为所述待识别视频的行为识别结果，即行为A为该待识别视频的行为识别结果。

[0098] 与本申请的第一个实施例相比，本实施例提出了一种具体的从所述待识别视频中提取出多帧视频图像的方式，也即将所述待识别视频的分辨率转换为预设的分辨率，根据所述待识别视频的视频长度选取第一数量的时间点，以每个所述时间点为起始点，分别从所述待识别视频中提取连续的多帧视频图像，得到所述第一数量的视频图像组，对所述视频图像组包含的每帧视频图像进行剪切，截取预设区域内的图像，从每个所述视频图像组包含的剪切后的视频图像中分别选取第三数量的视频图像，作为提取到的视频图像。

[0099] 请参阅图3，本申请实施例中一种视频行为识别装置的一个实施例包括：

[0100] 视频获取模块301，用于获取待识别视频；

[0101] 视频图像提取模块302，用于从所述待识别视频中提取出多帧视频图像；

[0102] 行为识别模块303，用于将所述多帧视频图像输入预先训练完成的目标神经网络模型，获得所述目标神经网络模型输出的行为识别结果；

[0103] 其中，所述目标神经网络模型包括第一处理层和第二处理层，所述第一处理层和所述第二处理层为所述目标神经网络模型的网络结构中相邻的两个卷积层，所述目标神经网络将所述多帧视频图像转换为四个维度的图像数据进行处理，所述第一处理层和第二处理层之间的处理逻辑如下：

[0104] 获取所述第一处理层输出的四维矩阵格式的第一图像处理数据，所述第一处理层用于接收上一个处理层输出的图像处理数据，对接收到的图像处理数据执行卷积处理，得到所述第一图像处理数据；

[0105] 对所述第一图像处理数据进行矩阵的平铺处理，获得二维矩阵格式的第二图像

处理数据；

[0106] 对所述第一图像数据进行卷积处理后，执行矩阵的平铺处理，获得二维矩阵格式的第三图像处理数据；

[0107] 将所述第三图像处理数据和所述第二图像处理数据相乘，获得二维矩阵格式的第四图像处理数据；

[0108] 对所述第四图像处理数据进行矩阵的逆平铺处理，得到四维矩阵格式的第五图像处理数据；

[0109] 将所述第五图像处理数据输入所述第二处理层，所述第二处理层用于对所述第五图像处理数据执行卷积处理，然后将处理后的数据输入下一个处理层。

[0110] 优选的，所述目标神经网络模型为RsesNet101模型，所述第一处理层为RsesNet101模型网络结构第四部分的最后一层处理层，所述第二处理层为RsesNet101模型网络结构第五部分的第一层处理层。

[0111] 进一步的，所述视频图像提取模块可以包括：

[0112] 第一分辨率转换单元，用于将所述待识别视频的分辨率转换为预设的分辨率；

[0113] 时间点选取单元，用于根据所述待识别视频的视频长度选取第一数量的时间点；

[0114] 视频图像提取单元，用于以每个所述时间点为起始点，分别从所述待识别视频中提取连续的多帧视频图像，得到所述第一数量的视频图像组，每个所述视频图像组包括第二数量的视频图像；

[0115] 视频图像剪切单元，用于对所述视频图像组包含的每帧视频图像进行剪切，截取预设区域内的图像；

[0116] 视频图像选取单元，用于从每个所述视频图像组包含的剪切后的视频图像中分别选取第三数量的视频图像，作为提取到的视频图像。

[0117] 更进一步的，所述行为识别模块可以包括：

[0118] 行为识别结果获取单元，用于通过所述目标神经网络模型分别获得每个所述视频图像组的行为识别结果，获取到的行为识别结果包括各个预设行为的匹配度；

[0119] 加权计算单元，用于计算各个所述视频图像组的行为识别结果的加权平均值，

加权的权重根据各个所述视频图像组所对应的所述时间点确定；

[0120] 行为识别结果确定单元，用于将所述加权平均值中匹配度最高的行为确定为所述待识别视频的行为识别结果。

[0121] 进一步的，所述视频行为识别装置还可以包括：

[0122] 初始模型训练模块，用于采用kinetics数据集训练得到初始神经网络模型；

[0123] 样本视频图像提取模块，用于从预先选取的样本视频中提取样本视频图像，所述样本视频的行为识别结果是已知的；

[0124] 样本行为识别模块，用于将所述样本视频图像输入所述初始神经网络模型，得到样本行为识别结果；

[0125] 模型参数修正模块，用于将所述样本行为识别结果和所述样本视频的行为识别结果进行比较，并根据比较的结果修正所述初始神经网络模型的参数，重复迭代直至所述样本行为识别结果和所述样本视频的行为识别结果之间的差别小于预设阈值；

[0126] 目标模型确定模块，用于将参数修正完毕的所述初始神经网络模型确定为所述目标神经网络模型。

[0127] 更进一步的，所述样本视频图像提取模块可以包括：

[0128] 第二分辨率转换单元，用于将所述样本视频的分辨率转换为预设的分辨率；

[0129] 样本时间点选取单元，用于根据所述样本视频的视频长度选取第四数量的样本时间点；

[0130] 样本视频图像提取单元，用于以每个所述样本时间点为起始点，分别从所述样本视频中提取连续的多帧视频图像，得到所述第四数量的样本视频图像组，每个所述样本视频图像组包括第五数量的视频图像；

[0131] 样本视频图像剪切单元，用于对所述样本视频图像组包含的每一帧视频图像分别按照左边界对齐、中部对齐和右边界对齐的方式剪切为3个视频图像；

[0132] 样本视频图像选取单元，用于从每个所述样本视频图像组包含的剪切后的视频图像中分别选取第六数量的视频图像，作为提取到的样本视频图像。

[0133] 本申请实施例还提供一种计算机非易失性可读存储介质，所述计算机非易失性可读存储介质存储有计算机可读指令，所述计算机可读指令被处理器执行时实

现如图1或图2表示的任意一种视频行为识别方法的步骤。

[0134] 本申请实施例还提供一种服务器，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可读指令，所述处理器执行所述计算机可读指令时实现如图1或图2表示的任意一种视频行为识别方法的步骤。

[0135] 图4是本申请一实施例提供的服务器的示意图。如图4所示，该实施例的服务器4包括：处理器40、存储器41以及存储在所述存储器41中并可在所述处理器40上运行的计算机可读指令42。所述处理器40执行所述计算机可读指令42时实现上述各个视频行为识别方法实施例中的步骤，例如图1所示的步骤101至103。或者，所述处理器40执行所述计算机可读指令42时实现上述各装置实施例中各模块/单元的功能，例如图3所示模块301至303的功能。

[0136] 本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程，是可以通过计算机可读指令来指令相关的硬件来完成，所述的计算机可读指令可存储于一非易失性计算机可读存储介质中，该计算机可读指令在执行时，可包括如上述各方法的实施例的流程。其中，本申请所提供的各实施例中所使用的对存储器、存储、数据库或其它介质的任何引用，均可包括非易失性和/或易失性存储器。非易失性存储器可包括只读存储器（ROM）、可编程ROM（PROM）、电可编程ROM（EPROM）、电可擦除可编程ROM（EEPROM）或闪存。易失性存储器可包括随机存取存储器（RAM）或者外部高速缓冲存储器。作为说明而非局限，RAM以多种形式可得，诸如静态RAM（SRAM）、动态RAM（DRAM）、同步DRAM（SDRAM）、双数据率SDRAM（DDRSDRAM）、增强型SDRAM（ESDRAM）、同步链路（Synchlink）DRAM（SLDRAM）、存储器总线（Rambus）直接RAM（RDRAM）、直接存储器总线动态RAM（DRDRAM）、以及存储器总线动态RAM（RDRAM）等。

权利要求书

[权利要求 1]

一种视频行为识别方法，其特征在于，包括：

获取待识别视频；

从所述待识别视频中提取出多帧视频图像；

将所述多帧视频图像输入预先训练完成的目标神经网络模型，获得所述目标神经网络模型输出的行为识别结果；

其中，所述目标神经网络模型包括第一处理层和第二处理层，所述第一处理层和所述第二处理层为所述目标神经网络模型的网络结构中相邻的两个卷积层，所述目标神经网络将所述多帧视频图像转换为四个维度的图像数据进行处理，所述第一处理层和第二处理层之间的处理逻辑如下：

获取所述第一处理层输出的四维矩阵格式的第一图像处理数据，所述第一处理层用于接收上一个处理层输出的图像处理数据，对接收到的图像处理数据执行卷积处理，得到所述第一图像处理数据；

对所述第一图像处理数据进行矩阵的平铺处理，获得二维矩阵格式的第二图像处理数据；

对所述第一图像数据进行卷积处理后，执行矩阵的平铺处理，获得二维矩阵格式的第三图像处理数据；

将所述第三图像处理数据和所述第二图像处理数据相乘，获得二维矩阵格式的第四图像处理数据；

对所述第四图像处理数据进行矩阵的逆平铺处理，得到四维矩阵格式的第五图像处理数据；

将所述第五图像处理数据输入所述第二处理层，所述第二处理层用于对所述第五图像处理数据执行卷积处理，然后将处理后的数据输入下一个处理层。

[权利要求 2]

根据权利要求1所述的视频行为识别方法，其特征在于，所述对所述第一图像数据进行卷积处理后，执行矩阵的平铺处理，获得二维矩阵格式的第三图像处理数据包括：

将所述第一图像处理数据和预先构建的第一卷积矩阵相乘后，执行矩阵的平铺处理，得到二维矩阵格式的第六图像处理数据；

将所述第一图像处理数据和预先构建的第二卷积矩阵相乘后，执行矩阵的平铺处理并转置，得到二维矩阵格式的第七图像处理数据，所述第一卷积矩阵和所述第二卷积矩阵均为 $1*1*1$ 的卷积核，且具有不同的元素值；

将所述第六图像处理数据和所述第七图像处理数据相乘，获得二维矩阵格式的所述第三图像处理数据。

[权利要求 3] 根据权利要求1所述的视频行为识别方法，其特征在于，所述目标神经网络模型为RsesNet101模型，所述第一处理层为RsesNet101模型网络结构第四部分的最后一层处理层，所述第二处理层为RsesNet101模型网络结构第五部分的第一层处理层。

[权利要求 4] 根据权利要求1所述的视频行为识别方法，其特征在于，所述从所述待识别视频中提取出多帧视频图像包括：

将所述待识别视频的分辨率转换为预设的分辨率；

根据所述待识别视频的视频长度选取第一数量的时间点；

以每个所述时间点为起始点，分别从所述待识别视频中提取连续的多帧视频图像，得到所述第一数量的视频图像组，每个所述视频图像组包括第二数量的视频图像；

对所述视频图像组包含的每帧视频图像进行剪切，截取预设区域内的图像；

从每个所述视频图像组包含的剪切后的视频图像中分别选取第三数量的视频图像，作为提取到的视频图像。

[权利要求 5] 根据权利要求4所述的视频行为识别方法，其特征在于，所述获得所述目标神经网络模型输出的行为识别结果包括：

通过所述目标神经网络模型分别获得每个所述视频图像组的行为识别结果，获取到的行为识别结果包括各个预设行为的匹配度；

计算各个所述视频图像组的行为识别结果的加权平均值，加权的权重

根据各个所述视频图像组所对应的所述时间点确定；

将所述加权平均值中匹配度最高的行为确定为所述待识别视频的行为识别结果。

[权利要求 6]

根据权利要求1至5中任一项所述的视频行为识别方法，其特征在于，所述目标神经网络模型通过以下步骤训练得到：

采用kinetics数据集训练得到初始神经网络模型；

从预先选取的样本视频中提取样本视频图像，所述样本视频的行为识别结果是已知的；

将所述样本视频图像输入所述初始神经网络模型，得到样本行为识别结果；

将所述样本行为识别结果和所述样本视频的行为识别结果进行比较，并根据比较的结果修正所述初始神经网络模型的参数，重复迭代直至所述样本行为识别结果和所述样本视频的行为识别结果之间的差别小于预设阈值；

将参数修正完毕的所述初始神经网络模型确定为所述目标神经网络模型。

[权利要求 7]

根据权利要求6所述的视频行为识别方法，其特征在于，所述从预先选取的样本视频中提取样本视频图像包括：

将所述样本视频的分辨率转换为预设的分辨率；

根据所述样本视频的视频长度选取第四数量的样本时间点；

以每个所述样本时间点为起始点，分别从所述样本视频中提取连续的多帧视频图像，得到所述第四数量的样本视频图像组，每个所述样本视频图像组包括第五数量的视频图像；

对所述样本视频图像组包含的每一帧视频图像分别按照左边界对齐、中部对齐和右边界对齐的方式剪切为3个视频图像；

从每个所述样本视频图像组包含的剪切后的视频图像中分别选取第六数量的视频图像，作为提取到的样本视频图像。

[权利要求 8]

一种视频行为识别装置，其特征在于，包括：

视频获取模块，用于获取待识别视频；

视频图像提取模块，用于从所述待识别视频中提取出多帧视频图像；

行为识别模块，用于将所述多帧视频图像输入预先训练完成的目标神经网络模型，获得所述目标神经网络模型输出的行为识别结果；

其中，所述目标神经网络模型包括第一处理层和第二处理层，所述第一处理层和所述第二处理层为所述目标神经网络模型的网络结构中相邻的两个卷积层，所述目标神经网络将所述多帧视频图像转换为四个维度的图像数据进行处理，所述第一处理层和第二处理层之间的处理逻辑如下：

获取所述第一处理层输出的四维矩阵格式的第一图像处理数据，所述第一处理层用于接收上一个处理层输出的图像处理数据，对接收到的图像处理数据执行卷积处理，得到所述第一图像处理数据；

对所述第一图像处理数据进行矩阵的平铺处理，获得二维矩阵格式的第二图像处理数据；

对所述第一图像数据进行卷积处理后，执行矩阵的平铺处理，获得二维矩阵格式的第三图像处理数据；

将所述第三图像处理数据和所述第二图像处理数据相乘，获得二维矩阵格式的第四图像处理数据；

对所述第四图像处理数据进行矩阵的逆平铺处理，得到四维矩阵格式的第五图像处理数据；

将所述第五图像处理数据输入所述第二处理层，所述第二处理层用于对所述第五图像处理数据执行卷积处理，然后将处理后的数据输入下一个处理层。

[权利要求 9]

根据权利要求8所述的视频行为识别装置，其特征在于，所述视频图像提取模块包括：

第一分辨率转换单元，用于将所述待识别视频的分辨率转换为预设的分辨率；

时间点选取单元，用于根据所述待识别视频的视频长度选取第一数量

的时间点；

视频图像提取单元，用于以每个所述时间点为起始点，分别从所述待识别视频中提取连续的多帧视频图像，得到所述第一数量的视频图像组，每个所述视频图像组包括第二数量的视频图像；

视频图像剪切单元，用于对所述视频图像组包含的每帧视频图像进行剪切，截取预设区域内的图像；

视频图像选取单元，用于从每个所述视频图像组包含的剪切后的视频图像中分别选取第三数量的视频图像，作为提取到的视频图像。

[权利要求 10] 根据权利要求9所述的视频行为识别装置，其特征在于，所述行为识别模块包括：

行为识别结果获取单元，用于通过所述目标神经网络模型分别获得每个所述视频图像组的行为识别结果，获取到的行为识别结果包括各个预设行为的匹配度；

加权计算单元，用于计算各个所述视频图像组的行为识别结果的加权平均值，加权的权重根据各个所述视频图像组所对应的所述时间点确定；

行为识别结果确定单元，用于将所述加权平均值中匹配度最高的行为确定为所述待识别视频的行为识别结果。

[权利要求 11] 根据权利要求8至10中任一项所述的视频行为识别装置，其特征在于，所述视频行为识别装置还包括：

初始模型训练模块，用于采用kinetics数据集训练得到初始神经网络模型；

样本视频图像提取模块，用于从预先选取的样本视频中提取样本视频图像，所述样本视频的行为识别结果是已知的；

样本行为识别模块，用于将所述样本视频图像输入所述初始神经网络模型，得到样本行为识别结果；

模型参数修正模块，用于将所述样本行为识别结果和所述样本视频的行为识别结果进行比较，并根据比较的结果修正所述初始神经网络模

型的参数，重复迭代直至所述样本行为识别结果和所述样本视频的行为识别结果之间的差别小于预设阈值；

目标模型确定模块，用于将参数修正完毕的所述初始神经网络模型确定为所述目标神经网络模型。

[权利要求 12] 根据权利要求11所述的视频行为识别装置，其特征在于，所述样本视频图像提取模块包括：

第二分辨率转换单元，用于将所述样本视频的分辨率转换为预设的分辨率；

样本时间点选取单元，用于根据所述样本视频的视频长度选取第四数量的样本时间点；

样本视频图像提取单元，用于以每个所述样本时间点为起始点，分别从所述样本视频中提取连续的多帧视频图像，得到所述第四数量的样本视频图像组，每个所述样本视频图像组包括第五数量的视频图像；

样本视频图像剪切单元，用于对所述样本视频图像组包含的每一帧视频图像分别按照左边界对齐、中部对齐和右边界对齐的方式剪切为3个视频图像；

样本视频图像选取单元，用于从每个所述样本视频图像组包含的剪切后的视频图像中分别选取第六数量的视频图像，作为提取到的样本视频图像。

[权利要求 13] 一种计算机非易失性可读存储介质，所述计算机非易失性可读存储介质存储有计算机可读指令，其特征在于，所述计算机可读指令被处理器执行时实现如下步骤：

获取待识别视频；

从所述待识别视频中提取出多帧视频图像；

将所述多帧视频图像输入预先训练完成的目标神经网络模型，获得所述目标神经网络模型输出的行为识别结果；

其中，所述目标神经网络模型包括第一处理层和第二处理层，所述第一处理层和所述第二处理层为所述目标神经网络模型的网络结构中相

邻的两个卷积层，所述目标神经网络将所述多帧视频图像转换为四个维度的图像数据进行处理，所述第一处理层和第二处理层之间的处理逻辑如下：

获取所述第一处理层输出的四维矩阵格式的第一图像处理数据，所述第一处理层用于接收上一个处理层输出的图像处理数据，对接收到的图像处理数据执行卷积处理，得到所述第一图像处理数据；

对所述第一图像处理数据进行矩阵的平铺处理，获得二维矩阵格式的第二图像处理数据；

对所述第一图像数据进行卷积处理后，执行矩阵的平铺处理，获得二维矩阵格式的第三图像处理数据；

将所述第三图像处理数据和所述第二图像处理数据相乘，获得二维矩阵格式的第四图像处理数据；

对所述第四图像处理数据进行矩阵的逆平铺处理，得到四维矩阵格式的第五图像处理数据；

将所述第五图像处理数据输入所述第二处理层，所述第二处理层用于对所述第五图像处理数据执行卷积处理，然后将处理后的数据输入下一个处理层。

[权利要求 14]

根据权利要求13所述的计算机非易失性可读存储介质，其特征在于，所述对所述第一图像数据进行卷积处理后，执行矩阵的平铺处理，获得二维矩阵格式的第三图像处理数据包括：

将所述第一图像处理数据和预先构建的第一卷积矩阵相乘后，执行矩阵的平铺处理，得到二维矩阵格式的第六图像处理数据；

将所述第一图像处理数据和预先构建的第二卷积矩阵相乘后，执行矩阵的平铺处理并转置，得到二维矩阵格式的第七图像处理数据，所述第一卷积矩阵和所述第二卷积矩阵均为 $1*1*1$ 的卷积核，且具有不同的元素值；

将所述第六图像处理数据和所述第七图像处理数据相乘，获得二维矩阵格式的所述第三图像处理数据。

- [权利要求 15] 根据权利要求13所述的计算机非易失性可读存储介质，其特征在于，所述从所述待识别视频中提取出多帧视频图像包括：
将所述待识别视频的分辨率转换为预设的分辨率；
根据所述待识别视频的视频长度选取第一数量的时间点；
以每个所述时间点为起始点，分别从所述待识别视频中提取连续的多帧视频图像，得到所述第一数量的视频图像组，每个所述视频图像组包括第二数量的视频图像；
对所述视频图像组包含的每帧视频图像进行剪切，截取预设区域内的图像；
从每个所述视频图像组包含的剪切后的视频图像中分别选取第三数量的视频图像，作为提取到的视频图像。
- [权利要求 16] 根据权利要求13至15中任一项所述的计算机非易失性可读存储介质，其特征在于，所述计算机可读指令被处理器执行时还实现如下步骤：
采用kinetics数据集训练得到初始神经网络模型；
从预先选取的样本视频中提取样本视频图像，所述样本视频的行为识别结果是已知的；
将所述样本视频图像输入所述初始神经网络模型，得到样本行为识别结果；
将所述样本行为识别结果和所述样本视频的行为识别结果进行比较，并根据比较的结果修正所述初始神经网络模型的参数，重复迭代直至所述样本行为识别结果和所述样本视频的行为识别结果之间的差别小于预设阈值；
将参数修正完毕的所述初始神经网络模型确定为所述目标神经网络模型。
- [权利要求 17] 一种服务器，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可读指令，其特征在于，所述处理器执行所述计算机可读指令时实现如下步骤：
获取待识别视频；

从所述待识别视频中提取出多帧视频图像；

将所述多帧视频图像输入预先训练完成的目标神经网络模型，获得所述目标神经网络模型输出的行为识别结果；

其中，所述目标神经网络模型包括第一处理层和第二处理层，所述第一处理层和所述第二处理层为所述目标神经网络模型的网络结构中相邻的两个卷积层，所述目标神经网络将所述多帧视频图像转换为四个维度的图像数据进行处理，所述第一处理层和第二处理层之间的处理逻辑如下：

获取所述第一处理层输出的四维矩阵格式的第一图像处理数据，所述第一处理层用于接收上一个处理层输出的图像处理数据，对接收到的图像处理数据执行卷积处理，得到所述第一图像处理数据；

对所述第一图像处理数据进行矩阵的平铺处理，获得二维矩阵格式的第二图像处理数据；

对所述第一图像数据进行卷积处理后，执行矩阵的平铺处理，获得二维矩阵格式的第三图像处理数据；

将所述第三图像处理数据和所述第二图像处理数据相乘，获得二维矩阵格式的第四图像处理数据；

对所述第四图像处理数据进行矩阵的逆平铺处理，得到四维矩阵格式的第五图像处理数据；

将所述第五图像处理数据输入所述第二处理层，所述第二处理层用于对所述第五图像处理数据执行卷积处理，然后将处理后的数据输入下一个处理层。

[权利要求 18]

根据权利要求17所述的服务器，其特征在于，所述对所述第一图像数据进行卷积处理后，执行矩阵的平铺处理，获得二维矩阵格式的第三图像处理数据包括：

将所述第一图像处理数据和预先构建的第一卷积矩阵相乘后，执行矩阵的平铺处理，得到二维矩阵格式的第六图像处理数据；

将所述第一图像处理数据和预先构建的第二卷积矩阵相乘后，执行矩

阵的平铺处理并转置，得到二维矩阵格式的第七图像处理数据，所述第一卷积矩阵和所述第二卷积矩阵均为 $1*1*1$ 的卷积核，且具有不同的元素值；

将所述第六图像处理数据和所述第七图像处理数据相乘，获得二维矩阵格式的所述第三图像处理数据。

[权利要求 19] 根据权利要求17所述的服务器，其特征在于，所述从所述待识别视频中提取出多帧视频图像包括：

将所述待识别视频的分辨率转换为预设的分辨率；

根据所述待识别视频的视频长度选取第一数量的时间点；

以每个所述时间点为起始点，分别从所述待识别视频中提取连续的多帧视频图像，得到所述第一数量的视频图像组，每个所述视频图像组包括第二数量的视频图像；

对所述视频图像组包含的每帧视频图像进行剪切，截取预设区域内的图像；

从每个所述视频图像组包含的剪切后的视频图像中分别选取第三数量的视频图像，作为提取到的视频图像。

[权利要求 20] 根据权利要求17至19中任一项所述的服务器，其特征在于，所述处理器执行所述计算机可读指令时还实现如下步骤：

采用kinetics数据集训练得到初始神经网络模型；

从预先选取的样本视频中提取样本视频图像，所述样本视频的行为识别结果是已知的；

将所述样本视频图像输入所述初始神经网络模型，得到样本行为识别结果；

将所述样本行为识别结果和所述样本视频的行为识别结果进行比较，并根据比较的结果修正所述初始神经网络模型的参数，重复迭代直至所述样本行为识别结果和所述样本视频的行为识别结果之间的差别小于预设阈值；

将参数修正完毕的所述初始神经网络模型确定为所述目标神经网络模

型。

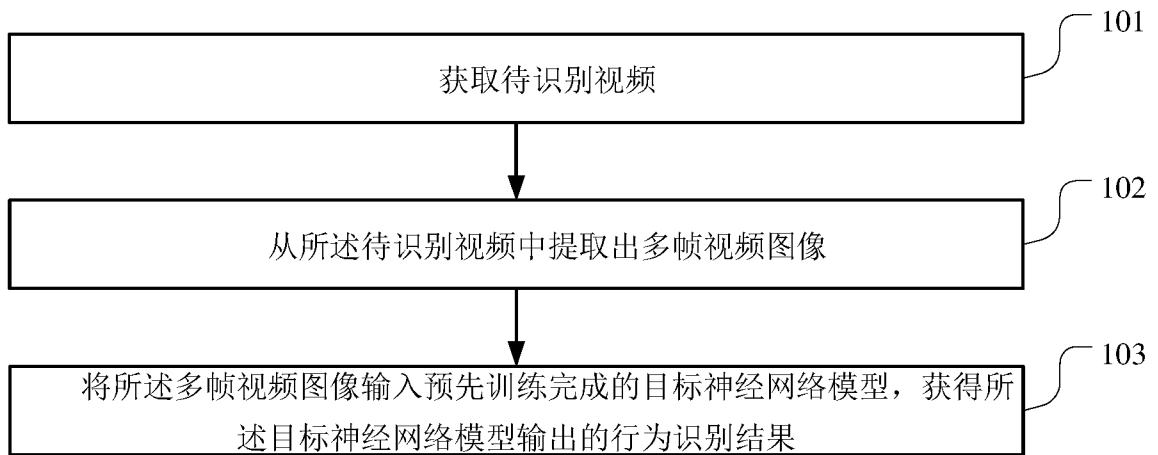


图 1

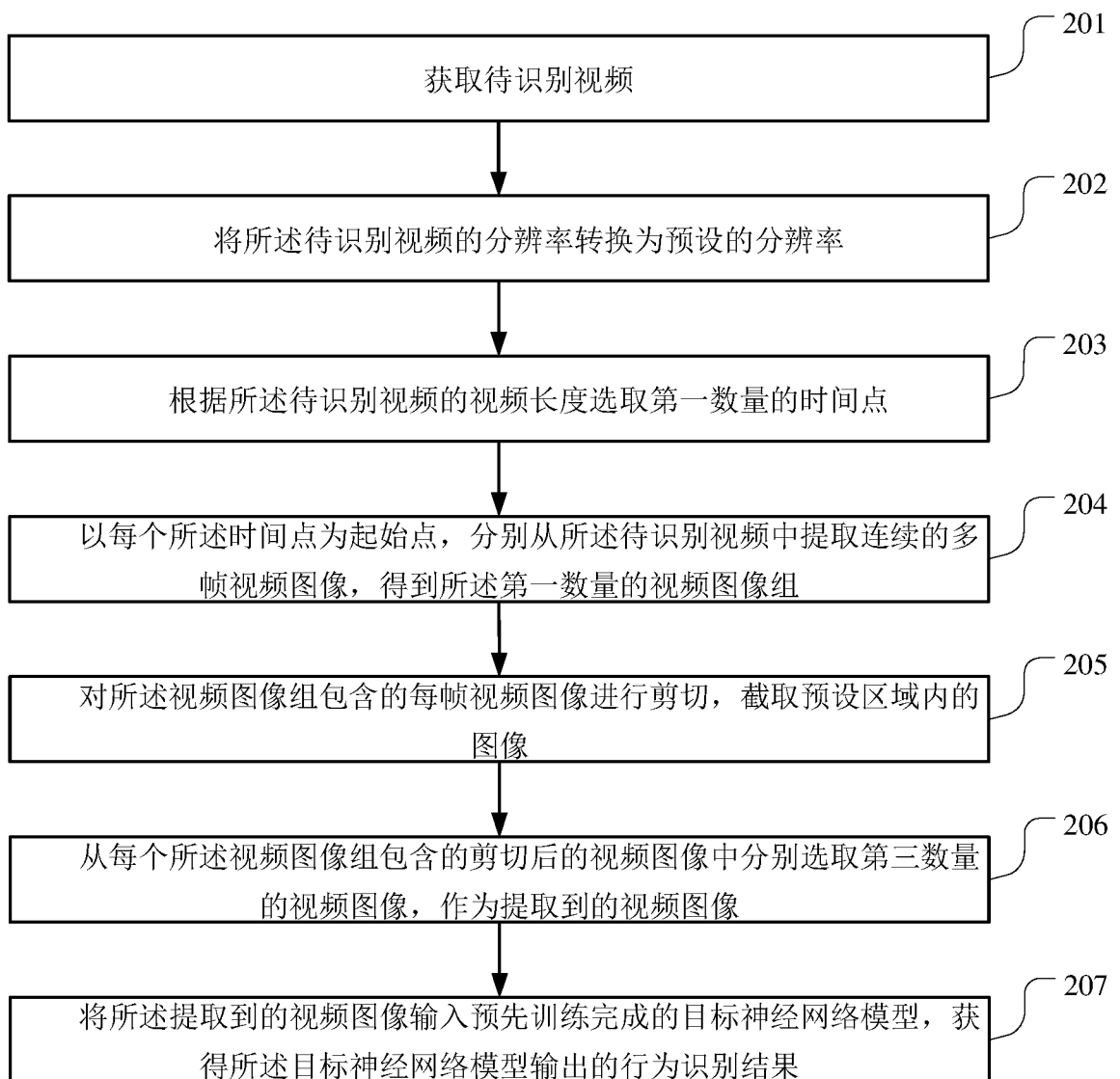


图 2

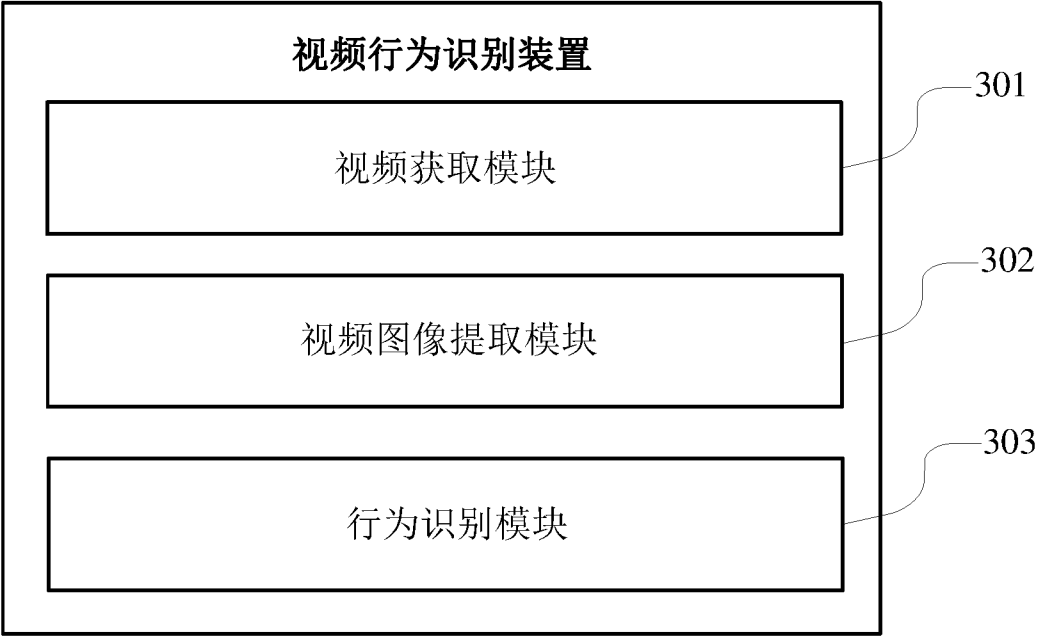


图 3

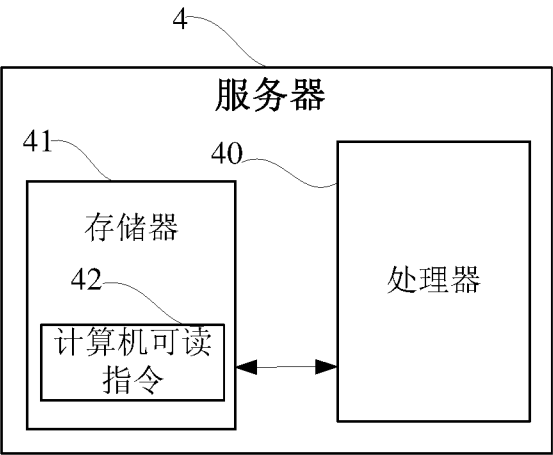


图 4

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2019/103174

A. CLASSIFICATION OF SUBJECT MATTER

G06K 9/00(2006.01)i; G06N 3/08(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; G06N

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)

CNABS; CNTXT; DWPI; SIPOABS; EPTXT; WOTXT; USTXT; CNKI: 视频, 帧, 图像, 训练, 神经网络, 矩阵, 全局, 局部, 特征, 行为, 识别, video, frame, image, train, NN, matrix, global, local, feature, behavior, recognize

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 105740773 A (CHONGQING UNIVERSITY OF TECHNOLOGY) 06 July 2016 (2016-07-06) description, paragraphs [0035]-[0074], figure 1	1-20
A	CN 107220611 A (SHANGHAI JIAO TONG UNIVERSITY) 29 September 2017 (2017-09-29) entire document	1-20
A	CN 107886061 A (SOUTH CHINA UNIVERSITY OF TECHNOLOGY) 06 April 2018 (2018-04-06) entire document	1-20
A	CN 108241849 A (NORTH CHINA UNIVERSITY OF TECHNOLOGY) 03 July 2018 (2018-07-03) entire document	1-20



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

27 December 2019

Date of mailing of the international search report

06 February 2020

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/CN2019/103174

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)		Publication date (day/month/year)	
CN	105740773	A	06 July 2016	CN	105740773	B	01 February 2019
CN	107220611	A	29 September 2017	None			
CN	107886061	A	06 April 2018	None			
CN	108241849	A	03 July 2018	None			

国际检索报告

国际申请号

PCT/CN2019/103174

A. 主题的分类

G06K 9/00(2006.01)i; G06N 3/08(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; G06N

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))

CNABS; CNTXT; DWPI; SIPOABS; EPTXT; WOTXT; USTXT; CNKI: 视频, 帧, 图像, 训练, 神经网络, 矩阵, 全局, 局部, 特征, 行为, 识别, video, frame, image, train, NN, matrix, global, local, feature, behavior, recognize

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
X	CN 105740773 A (重庆理工大学) 2016年 7月 6日 (2016 - 07 - 06) 说明书第[0035]-[0074]段, 图1	1-20
A	CN 107220611 A (上海交通大学) 2017年 9月 29日 (2017 - 09 - 29) 全文	1-20
A	CN 107886061 A (华南理工大学) 2018年 4月 6日 (2018 - 04 - 06) 全文	1-20
A	CN 108241849 A (北方工业大学) 2018年 7月 3日 (2018 - 07 - 03) 全文	1-20

☐ 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

“&” 同族专利的文件

国际检索实际完成的日期

2019年 12月 27日

国际检索报告邮寄日期

2020年 2月 6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

张剑峰

电话号码 (86-512) 88996211

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2019/103174

检索报告引用的专利文件			公布日 (年/月/日)	同族专利	公布日 (年/月/日)
CN	105740773	A	2016年 7月 6日	CN 105740773 B	2019年 2月 1日
CN	107220611	A	2017年 9月 29日	无	
CN	107886061	A	2018年 4月 6日	无	
CN	108241849	A	2018年 7月 3日	无	