

(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
2017年10月19日 (19.10.2017)



(10) 国际公布号
WO 2017/177661 A1

- (51) 国际专利分类号:
G06F 17/30 (2006.01)
- (21) 国际申请号: PCT/CN2016/103945
- (22) 国际申请日: 2016年10月31日 (31.10.2016)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201610237628.2 2016年4月15日 (15.04.2016) CN
- (71) 申请人: 乐视控股(北京)有限公司 (LE HOLDINGS (BEIJING) CO., LTD.) [CN/CN]; 中国北京市朝阳区姚家园路105号3号楼10层1102, Beijing 100025 (CN)。 乐视云计算有限公司 (LECLOUD COMPUTING CO., LTD.) [CN/CN]; 中国北京市朝阳区姚家园路105号观湖国际大厦1号楼15层, Beijing 100025 (CN)。
- (72) 发明人: 刘阳 (LIU, Yang); 中国北京市朝阳区姚家园路105号观湖国际大厦1号楼15层, Beijing 100025 (CN)。 白茂生 (BAI, Maosheng); 中国北京市朝阳区姚家园路105号观湖国际大厦1号楼15层, Beijing 100025 (CN)。 魏伟 (WEI, Wei); 中国北京市朝阳区姚家园路105号观湖国际大厦1号楼

15层, Beijing 100025 (CN)。 蔡砚刚 (CAI, Yang-gang); 中国北京市朝阳区姚家园路105号观湖国际大厦1号楼15层, Beijing 100025 (CN)。 祁海 (QI, Hai); 中国北京市朝阳区姚家园路105号观湖国际大厦1号楼15层, Beijing 100025 (CN)。 李兴玉 (LI, Xingyu); 中国北京市朝阳区姚家园路105号观湖国际大厦1号楼15层, Beijing 100025 (CN)。

(74) 代理人: 北京品源专利代理有限公司 (BEYOND ATTORNEYS AT LAW); 中国北京市海淀区莲花池东路39号西金大厦6层, Beijing 100036 (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, JP, KE, KG, 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,

[见续页]

(54) Title: CONVOLUTIONAL NEURAL NETWORK-BASED VIDEO RETRIEVAL METHOD AND SYSTEM

(54) 发明名称: 基于卷积神经网络的视频检索方法及系统

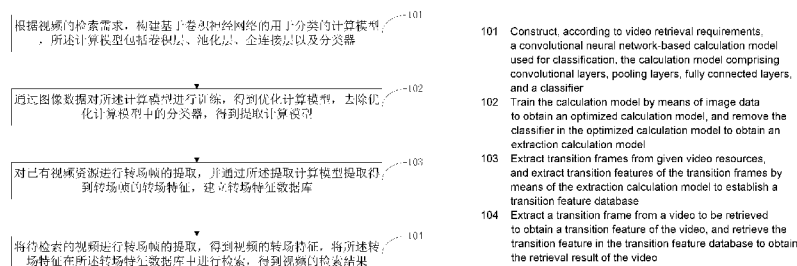


图 1

(57) Abstract: A convolutional neural network-based video retrieval method and system. The method comprises: first constructing a convolutional neural network-based calculation model (101); training the calculation model by means of image data to obtain an optimized calculation model, and removing a classifier from the model to obtain an extraction calculation model (102); extracting transition frames from given video resources, and extracting transition features of the transition frames by means of the extraction calculation model to establish a transition feature database (103); and extracting a transition frame from a video to be retrieved to obtain a transition feature, and retrieving the transition feature in the transition feature database to obtain the retrieval result of the video (104).

(57) 摘要: 一种基于卷积神经网络的视频检索方法及系统, 所述方法包括: 首先构建基于卷积神经网络的计算模型(101); 通过图像数据对计算模型进行训练, 得到优化计算模型; 去除模型中的分类器, 得到提取计算模型(102); 对已有视频资源进行转场帧的提取, 并通过提取计算模型提取得到转场帧的转场特征, 建立转场特征数据库(103); 将待检索的视频进行转场帧的提取得到转场特征, 将转场特征在转场特征数据库中进行检索, 得到视频的检索结果(104)。



WO 2017/177661 A1



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))。

基于卷积神经网络的视频检索方法及系统

本申请要求在2016年4月15日提交中国专利局、申请号为201610237628.2、公开名称为“基于卷积神经网络的视频检索方法及系统”的中国专利申请的优先权，该申请的全部内容通过引用结合在本申请中。

技术领域

本公开涉及卷积神经网络技术领域，特别涉及一种基于卷积神经网络的视频检索方法及系统。

背景技术

随着互联网及多媒体技术的不断发展，人们越来越依赖于通过网络检索获取想要得到的信息，例如通过网络检索获取视频信息。但是由于网络中存在数量巨大的视频文件和相关内容，而用户在检索时通常都是想要获取某一个单一的视频，如何从海量的视频资源中，快速地查找到想要的视频是目前用户进行视频检索的难点。

目前，比较常用的方法一般是根据视频的标题或者人为预先在视频中手动设置的标签进行视频检索，但是这种方法存在一定的局限性，有时并不能检索得到用户满意的结果。例如：检索某一个视频的名字，由于名字的重复、语种的差异用户常常检索得到的结果还是一大堆视频，无法准确得到想要的视频。而且，针对某些特殊情形，用户有时候并不知道想要检索的视频的标题，此时用户将无法实现检索。例如：用户已有部分视频内容或者低版本的视频文件，而想要检索完整的视频或者高版本的视频文件，但是又不知道视频的标题，此

时，相关技术的检索方法均不能较好的实现视频检索的目的。

发明内容

有鉴于此，本公开提出一种基于卷积神经网络的视频检索方法及系统，能够大大提高视频检索的速度和准确性。

本公开提供的一种基于卷积神经网络的视频检索方法，包括：

根据视频的检索需求，构建基于卷积神经网络的用于分类的计算模型，所述计算模型包括卷积层、池化层、全连接层以及分类器；

通过图像数据对所述计算模型进行训练，得到优化计算模型，去除优化计算模型中的分类器，得到提取计算模型；

对已有视频资源进行转场帧的提取，并通过所述提取计算模型提取得到转场帧的转场特征，建立转场特征数据库；

将待检索的视频进行转场帧的提取，通过提取计算模型提取得到视频的转场特征，将所述转场特征在所述转场特征数据库中进行检索，得到视频的检索结果。

可选地，所述基于卷积神经网络的用于分类的计算模型包括依次连接的卷积层C1、卷积层C2、池化层P2、卷积层C3、池化层P3、卷积层C4、卷积层C5、卷积层C6、池化层P6、全连接层fc7、全连接层fc8以及分类器。

可选地，所述6个卷积层的核大小均不大于 5×5 ，且所述卷积层C1的核大小为 3×3 ，卷积层C2的核大小为 3×3 。

可选地，所述卷积层C3的核大小为 5×5 ，所述池化层P2、池化层P3、卷积层C4、卷积层C5、卷积层C6、池化层P6的核大小均为 3×3 ；

所述卷积层C1、卷积层C2、卷积层C3、卷积层C4、卷积层C5、卷积层C6

的步数均为1，所述池化层P2、池化层P3、池化层P6的步数均为2；

所述卷积层C1、卷积层C2、卷积层C4、卷积层C5、卷积层C6的pad值均为1，所述卷积层C3的pad值为2，所述池化层P2、池化层P3、池化层P6的pad值为0；

所述卷积层C1、卷积层C2的卷积核的个数均为96个，所述卷积层C3、卷积层C6的卷积核的个数均为256个，所述卷积层C4、卷积层C5的卷积核的个数均为384个。

可选的，所述全连接层fc7、全连接层fc8的节点数目分别为4096、1000；且所述全连接层fc7采用dropout方式进行数据处理；

所述分类器为softmax分类器。

可选的，所述卷积层C1、卷积层C2、卷积层C3、卷积层C4、卷积层C5、卷积层C6以及所述全连接层fc7均采用激活函数LEAKY RELU进行数据的激活处理。

可选的，所述进行转场帧的提取的方法为：

按照预设的算法，判断所述当前视频帧的前后两帧Y平面的灰度直方图之差是否大于预设的视频阈值，若是，则当前视频帧为转场帧，否则，当前视频帧不是转场帧；

算法公式如下：

$$\sum_{i=0}^{255} |H_p[i] - H_n[i]| > T$$

其中，所述Hp[i]、Hn[i]分别为当前视频帧的前一帧和最后一帧的Y平面的灰度直方图对应的数值，T为预设的视频阈值。

可选地，所述预设的视频阈值T的计算公式为：

$T = \text{width} * \text{height} / 8$ ，其中，width和height分别为视频帧的宽度和高度。

本公开还提供了一种基于卷积神经网络的视频检索系统，包括：

构建模块，设置为根据视频的检索需求，构建基于卷积神经网络的用于分类的计算模型，将构建的计算模型发送给训练模块，所述计算模型包括卷积层、池化层、全连接层以及分类器；

训练模块，设置为接收所述构建模块发送的计算模型，通过图像数据对所述计算模型进行训练，得到优化计算模型，去除优化计算模型中的分类器，得到提取计算模型；将提取计算模型发送给数据库模块和检索模块；

数据库模块，设置为接收所述训练模块发送的提取计算模型，对已有视频资源进行转场帧的提取，并通过所述提取计算模型提取得到转场帧的转场特征，建立转场特征数据库；

检索模块，设置为接收所述训练模块发送的提取计算模型，将待检索的视频进行转场帧的提取，通过提取计算模型提取得到视频的转场特征，将所述转场特征在所述数据库模块中的转场特征数据库中进行检索，得到视频的检索结果。

可选的，所述基于卷积神经网络的用于分类的计算模型包括依次连接的卷积层C1、卷积层C2、池化层P2、卷积层C3、池化层P3、卷积层C4、卷积层C5、卷积层C6、池化层P6、全连接层fc7、全连接层fc8以及分类器。

可选地，所述6个卷积层的核大小均不大于 5×5 ，且所述卷积层C1的核大小为 3×3 ，卷积层C2的核大小为 3×3 。

可选地，所述卷积层C3的核大小为 5×5 ，所述池化层P2、池化层P3、卷积层C4、卷积层C5、卷积层C6、池化层P6的核大小均为 3×3 ；

所述卷积层C1、卷积层C2、卷积层C3、卷积层C4、卷积层C5、卷积层C6的步数均为1，所述池化层P2、池化层P3、池化层P6的步数均为2；

所述卷积层C1、卷积层C2、卷积层C4、卷积层C5、卷积层C6的pad值均为1，所述卷积层C3的pad值为2，所述池化层P2、池化层P3、池化层P6的pad值为0；

所述卷积层C1、卷积层C2的卷积核的个数均为96个，所述卷积层C3、卷积层C6的卷积核的个数均为256个，所述卷积层C4、卷积层C5的卷积核的个数均为384个。

可选的，所述全连接层fc7、全连接层fc8的节点数目分别为4096、1000；且所述全连接层fc7采用dropout方式进行数据处理；

所述分类器为softmax分类器。

可选的，所述卷积层C1、卷积层C2、卷积层C3、卷积层C4、卷积层C5、卷积层C6以及所述全连接层fc7均采用激活函数LEAKY RELU进行数据的激活处理。

可选的，所述进行转场帧的提取的方法为：

按照预设的算法，判断所述当前视频帧的前后两帧Y平面的灰度直方图之差是否大于预设的视频阈值，若是，则当前视频帧为转场帧，否则，当前视频帧不是转场帧；

算法公式如下：

$$\sum_{i=0}^{255} |H_p[i] - H_n[i]| > T$$

其中，所述Hp[i]、Hn[i]分别为当前视频帧的前一帧和后一帧的Y平面的灰度直方图对应的数值，T为预设的视频阈值。

可选地，所述预设的视频阈值T的计算公式为：

$T = \text{width} * \text{height} / 8$ ，其中，width和height分别为视频帧的宽度和高度。

本公开还提供了一种非暂态存储介质，存储有计算机可执行指令，所述计算机可执行指令设置为执行上述的基于卷积神经网络的视频检索方法。

本公开还提供了一种计算机程序产品，所述计算机程序产品包括存储在非暂态计算机可读存储介质上的计算机程序，所述计算机程序包括程序指令，当所述程序指令被计算机执行时，使所述计算机执行上述的基于卷积神经网络的视频检索方法。

本公开还提供了一种电子设备，包括至少一个处理器和与所述至少一个处理器通信连接的存储器，所述存储器用于存储可被所述至少一个处理器执行的指令，所述指令被所述至少一个处理器执行时，使所述至少一个处理器执行上述的基于卷积神经网络的视频检索方法。

本公开的基于卷积神经网络的视频检索方法及系统提高了检索过程的鲁棒性，去除了冗余信息，提高了视频检索的速度和准确性。

附图说明

图1为本公开实施例提供的基于卷积神经网络的视频检索方法的流程图；

图2为本公开实施例提供的卷积神经网络计算模型的结构示意图；

图3为本公开实施例提供的基于卷积神经网络的视频检索系统的实施例的结构示意图；

图4是本公开实施例提供的电子设备的硬件结构示意图。

具体实施方式

下面结合附图和实施例对本公开作进一步的详细说明。在不冲突的情况下，以下实施例和实施例中的特征可以相互组合。

需要说明的是，本公开中所有使用“第一”和“第二”的表述均是为了区分两个相同名称非相同的实体或者非相同的参量，可见“第一”“第二”仅为了表述的方便，不应理解为对本公开实施例的限定。

图1为本公开提供的基于卷积神经网络的视频检索方法的流程图。所述基于卷积神经网络的视频检索方法，包括：

步骤101，根据视频的检索需求，构建基于卷积神经网络的用于分类的计算模型，所述计算模型包括卷积层、池化层、全连接层以及分类器；

其中，所述卷积神经网络（Convolutional Neural Network，CNN）是近年发展起来，并引起广泛重视的一种高效识别方法。目前，卷积神经网络已经成为众多科学领域的研究热点之一，特别是在模式分类领域，由于该网络避免了对图像的复杂前期预处理，可以直接输入原始图像，因而得到了更为广泛的应用。同时根据卷积神经网络的分类特性，还可以扩展到其他应用领域中，例如：本公开中用于视频或图片特征的提取。所述基于卷积神经网络的用于分类的计算模型是指用于进行特征提取并最终实现分类的一类计算模型。其计算模型中包含多个卷积层、池化层以及全连接层，所述卷积层用于特征的提取，池化层用于特征数据的降维，也即在保证数据有效性的基础上大大减少数据量，全连接层实现数据映射，分类器用于实现特征的分类。

步骤102，通过图像数据对所述计算模型进行训练，得到优化计算模型，去除优化计算模型中的分类器，得到提取计算模型；

其中，所述图像数据例如为imagenet数据库。通过迭代训练，逐步完成使得计算模型完成自我学习的过程，称为优化的计算模型。将模型最后的分类器去除后，剩余的计算模型只能实现图片或视频特征的提取，也即得到一个训练后的提取计算模型。

步骤103，对已有视频资源进行转场帧的提取，并通过所述提取计算模型提取得到转场帧的转场特征，建立转场特征数据库；

其中，将所有能够获得的视频通过所述提取计算模型提取得到转场特征后，

建立转场特征数据库，当用户后续进行检索时，可以通过转场特征实现视频的检索。所述转场特征数据库通常只需要建立一次，就能够实现后续所有视频的检索，而且，在其他相关领域中所述转场特征数据库也能够重复使用。

步骤104，将待检索的视频进行转场帧的提取，通过提取计算模型提取得到视频的转场特征，将所述转场特征在所述转场特征数据库中进行检索，得到视频的检索结果。

其中，待检索的视频通常为残缺的视频、部分视频、或者较低版本的视频等，而用户想要通过这些视频检索得到更好的、完整的视频。将待检索的视频同样通过所述提取计算模型提取得到视频的转场特征，然后在步骤103中已建立的转场特征数据库中检索，能够得到与待检索的视频相关的视频结果。

由上述实施例可知，所述基于卷积神经网络的视频检索方法通过建立基于卷积神经网络的用于分类的计算模型，将训练后的计算模型中的分类器去除后，能够得到基于卷积神经网络的提取计算模块，将所有的视频资源根据所述提取计算模块提取得到转场帧的转场特征，并建立转场特征数据库，最后通过所述转场特征数据库可以实现视频的快速检索。所述基于卷积神经网络的视频检索方法及系统通过将视频内的转场帧作为数据处理的对象，不仅提高了检索过程的鲁棒性，而且去除了冗余信息；通过采用基于卷积神经网络的计算模型进行特征的提取，大大提高了视频检索的速度和准确性。

需要说明的是：所述视频检索的方法不仅仅用于视频的检索，同样适用于图片、音频等多媒体文件的检索，只需要相应的建立多媒体的特征数据库。

在一些可选地实施例中，所述基于卷积神经网络的用于分类的计算模型包括依次连接的卷积层C1、卷积层C2、池化层P2、卷积层C3、池化层P3、卷积层C4、卷积层C5、卷积层C6、池化层P6、全连接层fc7、全连接层fc8以及分类器。

这样，通过前端两个卷积层的连续特征提取，能够提高计算模型的特征提取的效率和速度。

可选地，所述池化层采用平均值进行池化处理。

可选的，在本公开一个实施例中，所述计算模型采用2个全连接层，最后的全连接层fc8输出特征为1000维特征。并且为了防止过拟合，全连接层中采用了dropout的方式。训练时采用softmax分类器。

完成上述计算模型的设计后，利用ImageNet图像数据库对上述网络进行分类训练。训练迭代次数为30万次。完成训练后，将上述模型中的softmax层去掉，利用模型中其他部分进行特征提取工作，特征输出为模型中的全连接层fc8。

建立视频数据库：对所有现有视频资源，依次进行转场帧提取，之后利用训练好的模型（去除softmax层）对转场帧进行特征提取，并保存，这样一个视频会得到一个特征向量。将所有视频的特征向量进行保存，以便后续检索时应用。

检索：对于待检索的视频，首先进行转场帧提取，之后利用训练好的模型进行特征提取，最后利用kd tree算法，根据已提取的特征以及整个视频库的特征进行快速检索。

在另一些实施例中，所述6个卷积层的核大小均不大于 5×5 ，且所述卷积层C1的核大小为 3×3 ，卷积层C2的核大小为 3×3 。这样，依次连接的卷积层由于具有较小的核，能够更有效地提取视频或图片的特征数据，同时还减少了神经网络计算模型的参数，对于提升特征提取的速度以及防止过度拟合有较大的作用。

可选地，所述卷积层C3的核大小为 5×5 ，所述池化层P2、池化层P3、卷积层C4、卷积层C5、卷积层C6、池化层P6的核大小均为 3×3 ；

所述卷积层C1、卷积层C2、卷积层C3、卷积层C4、卷积层C5、卷积层C6

的步数均为1，所述池化层P2、池化层P3、池化层P6的步数均为2；

所述卷积层C1、卷积层C2、卷积层C4、卷积层C5、卷积层C6的pad值均为1，所述卷积层C3的pad值为2，所述池化层P2、池化层P3、池化层P6的pad值为0；

所述卷积层C1、卷积层C2的卷积核的个数均为96个，所述卷积层C3、卷积层C6的卷积核的个数均为256个，所述卷积层C4、卷积层C5的卷积核的个数均为384个。

其中，所述卷积层的步数指卷积层的核每次移动的步长，所述pad值是指在输入数据的周围是否添加一圈数据参与运算，pad值的大小也即添加数据的圈数。这样，能够提高计算模型的处理效率和速度，进而提高视频特征提取的效率。

在本公开一些实施例中，所述全连接层fc7、全连接层fc8的节点数目分别为4096、1000；且所述全连接层fc7采用dropout方式进行数据处理；所述分类器为softmax分类器。这里，所述的节点数目也可以理解为特征数目。所述dropout方式是通过随机开启一定数目的数据，而将剩下的数据丢弃，这样能够有效地防止数据的过拟合，进而提高特征提取的速度和效率。

可选地，所述卷积层C1、卷积层C2、卷积层C3、卷积层C4、卷积层C5、卷积层C6以及所述全连接层fc7均采用激活函数LEAKY RELU进行数据的激活处理。所述激活函数是将上一次的输出数据通过激活函数中的算法计算得到新的输出结果，将所述新的输出结果作为下一层的输入数据。同时，本公开所采用的激活函数LEAKY RELU相比于传统的激活函数RELU，在函数值小于零时，也具有一定的输出值，也即使得函数值小于零这一部分的数据也可以参与训练过程。这里，当函数值小于0时，输出值为输入值乘以一个系数a，所述系数a优选采用固定值。

在本公开可选的实施例中，所述进行转场帧的提取的方法为：

按照预设的算法，判断所述当前视频帧的前后两帧Y平面的灰度直方图之差是否大于预设的视频阈值，若当前视频帧的前后两帧Y平面的灰度直方图之差大于预设的视频阈值，则当前视频帧为转场帧，当前视频帧的前后两帧Y平面的灰度直方图之差小于或等于预设的视频阈值，当前视频帧不是转场帧；

算法公式如下：

$$\sum_{i=0}^{255} |H_p[i] - H_n[i]| > T$$

其中，所述Hp[i]、Hn[i]分别为当前视频帧的前一帧和后一帧的Y平面的灰度直方图对应的数值，T为预设的视频阈值。

对于输入的视频，考虑到视频中含有大量的帧信息，如果对视频的每一个帧都进行检测，则会非常耗时。故本公开首先对视频的重要帧进行提取。这里的重要帧是指能够很好的代表视频内容的帧。在视频中，一般关键帧可以表示视频的大致内容，但考虑到关键帧在多次编码后会发生变化，为了提高系统的鲁棒性，本公开将视频内的转场帧作为处理对象，保证了鲁棒性的同时，去除了冗余信息，最终提高了视频检索的准确率和速度。

可选地，所述预设的视频阈值T的计算公式为：

$T = \text{width} * \text{height} / 8$ ，其中，width和height分别为视频帧的宽度和高度。

可选的，本公开准备训练样本约130万张图像，对卷积神经网络进行训练，获得优化计算模型。然后利用模型对网络中现有的视频库进行特征提取并建库。其中训练模型中卷积层采用标准偏差为0.01的高斯分布进行初始化。LEAKY RELU函数的a参数为0.01。全连接层中参数采用标准偏差为0.002的高斯分布进行初始化。dropout模块的参数为0.5。训练过程采用反向传播算法（BP算法）进行参数的训练及更新。本公开中一共训练30万次迭代。

参照图2所示，为本公开提供的基于卷积神经网络的计算模型的结构示意图。所述基于卷积神经网络的计算模型包括：依次连接的卷积层C1、卷积层C2、池化层P2、卷积层C3、池化层P3、卷积层C4、卷积层C5、卷积层C6、池化层P6、全连接层fc7、全连接层fc8以及分类器SVM。所有卷积层以及全连接层fc7均通过激活函数LEAKY RELU进行数据的处理，使得前一层的数据能够向下一层传递。且所述全连接层fc7采用dropout方式进行数据处理。

参照图3所示，为本公开提供的基于卷积神经网络的视频检索系统的实施例的结构示意图。所述基于卷积神经网络的视频检索系统包括：构建模块301，训练模块302，数据库模块303以及检索模块304。

构建模块301，设置为根据视频的检索需求，构建基于卷积神经网络的用于分类的计算模型，将构建的计算模型发送给训练模块302，所述计算模型包括卷积层、池化层、全连接层以及分类器；

训练模块302，设置为接收所述构建模块301发送的计算模型，通过图像数据对所述计算模型进行训练，得到优化计算模型，去除优化计算模型中的分类器，得到提取计算模型；将提取计算模型发送给数据库模块303和检索模块304；

数据库模块303，设置为接收所述训练模块302发送的提取计算模型，对已有视频资源进行转场帧的提取，并通过所述提取计算模型提取得到转场帧的转场特征，建立转场特征数据库；

检索模块304，设置为接收所述训练模块302发送的提取计算模型，将待检索的视频进行转场帧的提取，通过提取计算模型提取得到视频的转场特征，将所述转场特征在所述数据库模块303中的转场特征数据库中进行检索，得到视频的检索结果。

由上述实施例可知，所述基于卷积神经网络的视频检索系统通过所述构建

模块301建立基于卷积神经网络的用于分类的计算模型，通过所述训练模块302训练得到优化计算模型，去除优化计算模型中的分类器，得到提取计算模型；通过所述数据库模块303建立了基于转场帧的转场特征数据库，最后通过所述检索模块304实现视频的准确检索。所述基于卷积神经网络的视频检索系统通过将视频内的转场帧作为数据处理的对象，不仅提高了检索过程的鲁棒性，而且去除了冗余信息；通过采用基于卷积神经网络的计算模型进行特征的提取，大大提高了视频检索的速度和准确性。

可选地，本公开采用卷积神经网络进行视频整体特征的提取与组合，利用kd tree算法完成检索，检索结果具有准确、快速等优点。

在一些可选的实施例中，所述基于卷积神经网络的用于分类的计算模型包括依次连接的卷积层C1、卷积层C2、池化层P2、卷积层C3、池化层P3、卷积层C4、卷积层C5、卷积层C6、池化层P6、全连接层fc7、全连接层fc8以及分类器。

在另一些可选的实施例中，所述6个卷积层的核大小均不大于 5×5 ，且所述卷积层C1的核大小为 3×3 ，卷积层C2的核大小为 3×3 。

在一些可选地实施例中，所述卷积层C3的核大小为 5×5 ，所述池化层P2、池化层P3、卷积层C4、卷积层C5、卷积层C6、池化层P6的核大小均为 3×3 ；

所述卷积层C1、卷积层C2、卷积层C3、卷积层C4、卷积层C5、卷积层C6的步数均为1，所述池化层P2、池化层P3、池化层P6的步数均为2；

所述卷积层C1、卷积层C2、卷积层C4、卷积层C5、卷积层C6的pad值均为1，所述卷积层C3的pad值为2，所述池化层P2、池化层P3、池化层P6的pad值为0；

所述卷积层C1、卷积层C2的卷积核的个数均为96个，所述卷积层C3、卷积层C6的卷积核的个数均为256个，所述卷积层C4、卷积层C5的卷积核的个数均为384个。

在另一些可选地实施例中，所述全连接层fc7、全连接层fc8的节点数目分别为4096、1000；且所述全连接层fc7采用dropout方式进行数据处理；所述分类器为softmax分类器。

可选的，所述卷积层C1、卷积层C2、卷积层C3、卷积层C4、卷积层C5、卷积层C6以及所述全连接层fc7均采用激活函数LEAKY RELU进行数据的激活处理。

可选的，所述进行转场帧的提取的方法为：

按照预设的算法，判断所述当前视频帧的前后两帧Y平面的灰度直方图之差是否大于预设的视频阈值，若当前视频帧的前后两帧Y平面的灰度直方图之差大于预设的视频阈值，则当前视频帧为转场帧，若当前视频帧的前后两帧Y平面的灰度直方图之差小于或等于预设的视频阈值，当前视频帧不是转场帧；

算法公式如下：

$$\sum_{i=0}^{255} |H_p[i] - H_n[i]| > T$$

其中，所述Hp[i]、Hn[i]分别为当前视频帧的前一帧和最后一帧的Y平面的灰度直方图对应的数值，T为预设的视频阈值。

可选地，所述预设的视频阈值T的计算公式为：

$T = \text{width} * \text{height} / 8$ ，其中，width和height分别为视频帧的宽度和高度。

在一些可选的实施例中，本公开所述的基于卷积神经网络的视频检索系统应用于手机、电脑、平板、智能电视等各类智能终端设备中，或者，还可以用于各类服务器中，例如网页搜索的服务器中。同时，将本公开所述基于卷积神经网络的视频检索系统中的部分模块分别作为终端、服务器的功能模块也属于本公开保护的范围。

本公开的实施例还提供一种包含计算机可执行指令的非暂态存储介质，所

述计算机可执行指令在由计算机处理器执行时用于执行上述实施例的基于卷积神经网络的视频检索方法。

本公开的实施例还提供了一种计算机程序产品，所述计算机程序产品包括存储在非暂态计算机可读存储介质上的计算机程序，所述计算机程序包括程序指令，当所述程序指令被计算机执行时，使所述计算机执行上述实施例的基于卷积神经网络的视频检索方法。

图4为本申请实施例提供的一种电子设备的硬件结构示意图，如图5所示，该电子设备包括：一个或多个处理器501以及存储器502，图4中以一个处理器501为例。

电子设备还可以包括：输入装置503和输出装置504。

电子设备中的处理器501、存储器502、输入装置503和输出装置504可以通过总线或者其他方式连接，图4中以通过总线连接为例。

存储器502作为一种非易失性计算机可读存储介质，可用于存储非易失性软件程序、非易失性计算机可执行程序以及模块，如本申请实施例中的基于卷积神经网络的视频检索方法对应的程序指令/模块。处理器501通过运行存储在存储器502中的非易失性软件程序、指令以及模块，从而执行服务器的各种功能应用以及数据处理，即实现上述实施例的基于卷积神经网络的视频检索方法。

存储器502可以包括存储程序区和存储数据区，其中，存储程序区可存储操作系统、至少一个功能所需要的应用程序；存储数据区可存储根据拍摄照片的方法使用所创建的数据等。此外，存储器502可以包括高速随机存取存储器，还可以包括非易失性存储器，例如至少一个磁盘存储器件、闪存器件、或其他非易失性固态存储器件。在一些实施例中，存储器502可选包括相对于处理器501远程设置的存储器。

输入装置503可用于接收输入的数字或字符信息，以及用户设置以及功能控制有关的键信号输入。输出装置504可包括显示屏等显示设备。

所述一个或者多个模块存储在所述存储器502中，当被所述一个或者多个处理器501执行时，执行上述任意方法实施例中的基于卷积神经网络的视频检索方法。

上述产品可执行本公开实施例所提供的方法，具备执行方法相应的功能模块和有益效果。未在本实施例中详尽描述的技术细节，可参见本公开实施例所提供的方法。

本公开的实施例的电子设备以多种形式存在，包括但不限于：

(1)移动通信设备：这类设备的特点是具备移动通信功能，并且以提供话音、数据通信为主要目标。这类终端包括：智能手机(例如iPhone)、多媒体手机、功能性手机，以及低端手机等。

(2)超移动个人计算机设备：这类设备属于个人计算机的范畴，有计算和处理功能，一般也具备移动上网特性。这类终端包括：PDA、MID和UMPC设备等，例如iPad。

(3)便携式娱乐设备：这类设备可以显示和播放多媒体内容。该类设备包括：音频、视频播放器(例如iPod)，掌上游戏机，电子书，以及智能玩具和便携式车载导航设备。

(4)服务器：提供计算服务的设备，服务器的构成包括处理器、硬盘、内存、系统总线等，服务器和通用的计算机架构类似，但是由于需要提供高可靠的服务，因此在处理能力、稳定性、可靠性、安全性、可扩展性、可管理性等方面要求较高。

(5)其他具有数据交互功能的电子设备。

以上所描述的装置实施例仅仅是示意性的，其中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模块来实现本实施例方案的目的。

通过以上的实施方式的描述，本领域的技术人员可以清楚地了解到实施方式可借助软件加通用硬件平台的方式来实现，当然也可以通过硬件。基于这样的理解，上述技术方案本质上或者说对相关技术做出贡献的部分可以以软件产品的形式体现出来，该计算机软件产品可以存储在计算机可读存储介质中，如ROM/RAM、磁碟、光盘等，包括若干指令用以使得一台计算机设备（可以是个人计算机，服务器，或者网络设备等）执行实施例或者实施例的某些部分所述的方法。

注意，上述仅为本公开的较佳实施例及所运用技术原理。本领域技术人员会理解，本公开不限于这里所述的特定实施例，对本领域技术人员来说能够进行各种明显的变化、重新调整和替代而不会脱离本公开的保护范围。

工业实用性

本公开的基于卷积神经网络的视频检索方法及系统提高了检索过程的鲁棒性，去除了冗余信息，提高了视频检索的速度和准确性。

权利要求书

1. 一种基于卷积神经网络的视频检索方法，应用于电子设备，包括：

根据视频的检索需求，构建基于卷积神经网络的用于分类的计算模型，所述计算模型包括卷积层、池化层、全连接层以及分类器；

通过图像数据对所述计算模型进行训练，得到优化计算模型，去除优化计算模型中的分类器，得到提取计算模型；

对已有视频资源进行转场帧的提取，并通过所述提取计算模型提取得到转场帧的转场特征，建立转场特征数据库；以及

将待检索的视频进行转场帧的提取，通过提取计算模型提取得到视频的转场特征，将所述转场特征在所述转场特征数据库中进行检索，得到视频的检索结果。

2. 根据权利要求1所述的方法，其中，所述基于卷积神经网络的用于分类的计算模型包括依次连接的卷积层C1、卷积层C2、池化层P2、卷积层C3、池化层P3、卷积层C4、卷积层C5、卷积层C6、池化层P6、全连接层fc7、全连接层fc8以及分类器。

3. 根据权利要求2所述的方法，其中，所述6个卷积层的核大小均不大于 5×5 ，且所述卷积层C1的核大小为 3×3 ，卷积层C2的核大小为 3×3 。

4. 根据权利要求3所述的方法，其中，所述卷积层C3的核大小为 5×5 ，所述池化层P2、池化层P3、卷积层C4、卷积层C5、卷积层C6、池化层P6的核大小均为 3×3 ；

所述卷积层C1、卷积层C2、卷积层C3、卷积层C4、卷积层C5、卷积层C6的步数均为1，所述池化层P2、池化层P3、池化层P6的步数均为2；

所述卷积层C1、卷积层C2、卷积层C4、卷积层C5、卷积层C6的pad值均为1，所述卷积层C3的pad值为2，所述池化层P2、池化层P3、池化层P6的pad值为0；

所述卷积层C1、卷积层C2的卷积核的个数均为96个，所述卷积层C3、卷积层C6的卷积核的个数均为256个，所述卷积层C4、卷积层C5的卷积核的个数均为384个。

5. 根据权利要求2所述的方法，其中，所述全连接层fc7、全连接层fc8的节点数目分别为4096、1000；且所述全连接层fc7采用dropout方式进行数据处理；

所述分类器为softmax分类器。

6. 根据权利要求2所述的方法，其中，所述卷积层C1、卷积层C2、卷积层C3、卷积层C4、卷积层C5、卷积层C6以及所述全连接层fc7均采用激活函数LEAKY RELU进行数据的激活处理。

7. 根据权利要求1所述的方法，其中，所述进行转场帧的提取的方法为：

按照预设的算法，判断所述当前视频帧的前后两帧Y平面的灰度直方图之差是否大于预设的视频阈值，若当前视频帧的前后两帧Y平面的灰度直方图之差大于预设的视频阈值，则当前视频帧为转场帧，若当前视频帧的前后两帧Y平面的灰度直方图之差小于等于预设的视频阈值，当前视频帧不是转场帧；

算法公式如下：

$$\sum_{i=0}^{255} |H_p[i] - H_n[i]| > T$$

其中，所述Hp[i]、Hn[i]分别为当前视频帧的前一帧和后一帧的Y平面的灰度直方图对应的数值，T为预设的视频阈值。

8. 根据权利要求7所述的方法，其中，所述预设的视频阈值T的计算公式为：

$T = \text{width} * \text{height} / 8$ ，其中，width和height分别为视频帧的宽度和高度。

9. 一种基于卷积神经网络的视频检索系统，包括：

构建模块，设置为根据视频的检索需求，构建基于卷积神经网络的用于分

类的计算模型，将构建的计算模型发送给训练模块，所述计算模型包括卷积层、池化层、全连接层以及分类器；

训练模块，设置为接收所述构建模块发送的计算模型，通过图像数据对所述计算模型进行训练，得到优化计算模型，去除优化计算模型中的分类器，得到提取计算模型；将提取计算模型发送给数据库模块和检索模块；

数据库模块，设置为接收所述训练模块发送的提取计算模型，对已有视频资源进行转场帧的提取，并通过所述提取计算模型提取得到转场帧的转场特征，建立转场特征数据库；以及

检索模块，设置为接收所述训练模块发送的提取计算模型，将待检索的视频进行转场帧的提取，通过提取计算模型提取得到视频的转场特征，将所述转场特征在所述数据库模块中的转场特征数据库中进行检索，得到视频的检索结果。

10. 根据权利要求9所述的系统，其中，所述基于卷积神经网络的用于分类的计算模型包括依次连接的卷积层C1、卷积层C2、池化层P2、卷积层C3、池化层P3、卷积层C4、卷积层C5、卷积层C6、池化层P6、全连接层fc7、全连接层fc8以及分类器。

11. 根据权利要求10所述的系统，其中，所述6个卷积层的核大小均不大于 5×5 ，且所述卷积层C1的核大小为 3×3 ，卷积层C2的核大小为 3×3 。

12. 根据权利要求11所述的系统，其中，所述卷积层C3的核大小为 5×5 ，所述池化层P2、池化层P3、卷积层C4、卷积层C5、卷积层C6、池化层P6的核大小均为 3×3 ；

所述卷积层C1、卷积层C2、卷积层C3、卷积层C4、卷积层C5、卷积层C6的步数均为1，所述池化层P2、池化层P3、池化层P6的步数均为2；

所述卷积层C1、卷积层C2、卷积层C4、卷积层C5、卷积层C6的pad值均为1，所述卷积层C3的pad值为2，所述池化层P2、池化层P3、池化层P6的pad值为0；

所述卷积层C1、卷积层C2的卷积核的个数均为96个，所述卷积层C3、卷积层C6的卷积核的个数均为256个，所述卷积层C4、卷积层C5的卷积核的个数均为384个。

13. 根据权利要求10所述的系统，其中，所述全连接层fc7、全连接层fc8的节点数目分别为4096、1000；且所述全连接层fc7采用dropout方式进行数据处理；所述分类器为softmax分类器。

14. 根据权利要求10所述的系统，其中，所述卷积层C1、卷积层C2、卷积层C3、卷积层C4、卷积层C5、卷积层C6以及所述全连接层fc7均采用激活函数LEAKY RELU进行数据的激活处理。

15. 根据权利要求9所述的系统，其中，所述进行转场帧的提取的方法为：

按照预设的算法，判断所述当前视频帧的前后两帧Y平面的灰度直方图之差是否大于预设的视频阈值，若当前视频帧的前后两帧Y平面的灰度直方图之差大于预设的视频阈值，则当前视频帧为转场帧，若当前视频帧的前后两帧Y平面的灰度直方图之差小于等于预设的视频阈值，当前视频帧不是转场帧；

算法公式如下：

$$\sum_{i=0}^{255} |H_p[i] - H_n[i]| > T$$

其中，所述Hp[i]、Hn[i]分别为当前视频帧的前一帧和后一帧的Y平面的灰度直方图对应的数值，T为预设的视频阈值。

16. 根据权利要求15所述的系统，其中，所述预设的视频阈值T的计算公式为：

$T = \text{width} * \text{height} / 8$ ，其中，width和height分别为视频帧的宽度和高度。

17、一种非暂态存储介质，存储有计算机可执行指令，所述计算机可执行指令设置为执行权利要求1至8任一项所述的基于卷积神经网络的视频检索方法。

18、一种计算机程序产品，所述计算机程序产品包括存储在非暂态计算机可读存储介质上的计算机程序，所述计算机程序包括程序指令，当所述程序指令被计算机执行时，使所述计算机执行权利要求1至8任一项所述的基于卷积神经网络的视频检索方法。

19、一种电子设备，包括至少一个处理器和与所述至少一个处理器通信连接的存储器，所述存储器用于存储可被所述至少一个处理器执行的指令，所述指令被所述至少一个处理器执行时，使所述至少一个处理器执行权利要求1至8任一项所述的基于卷积神经网络的视频检索方法。

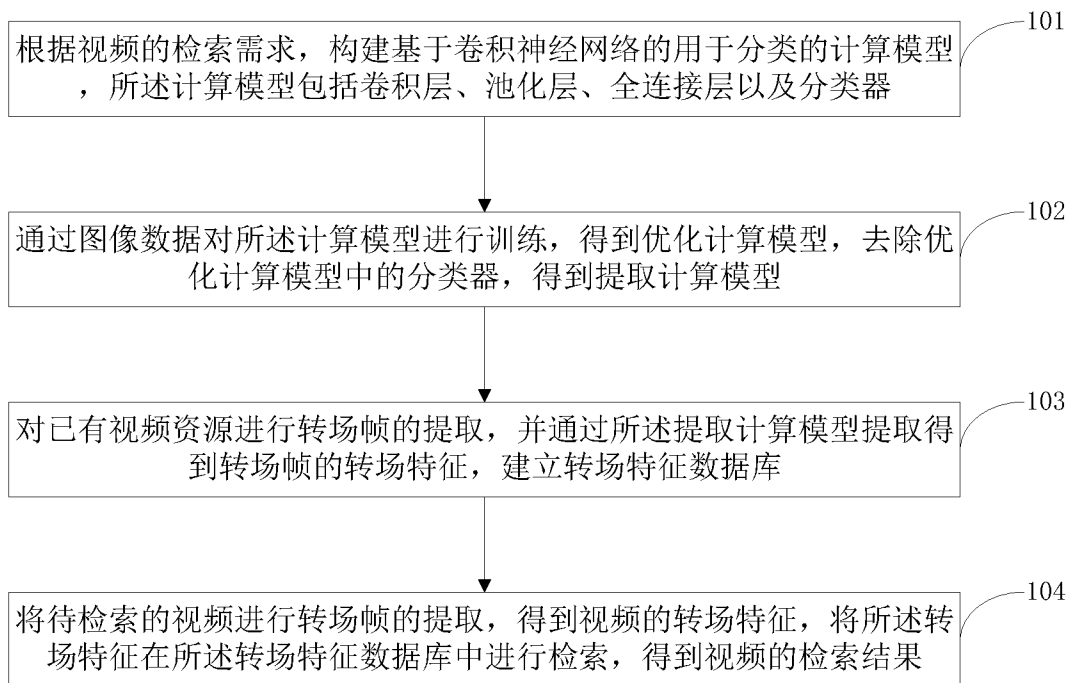


图 1

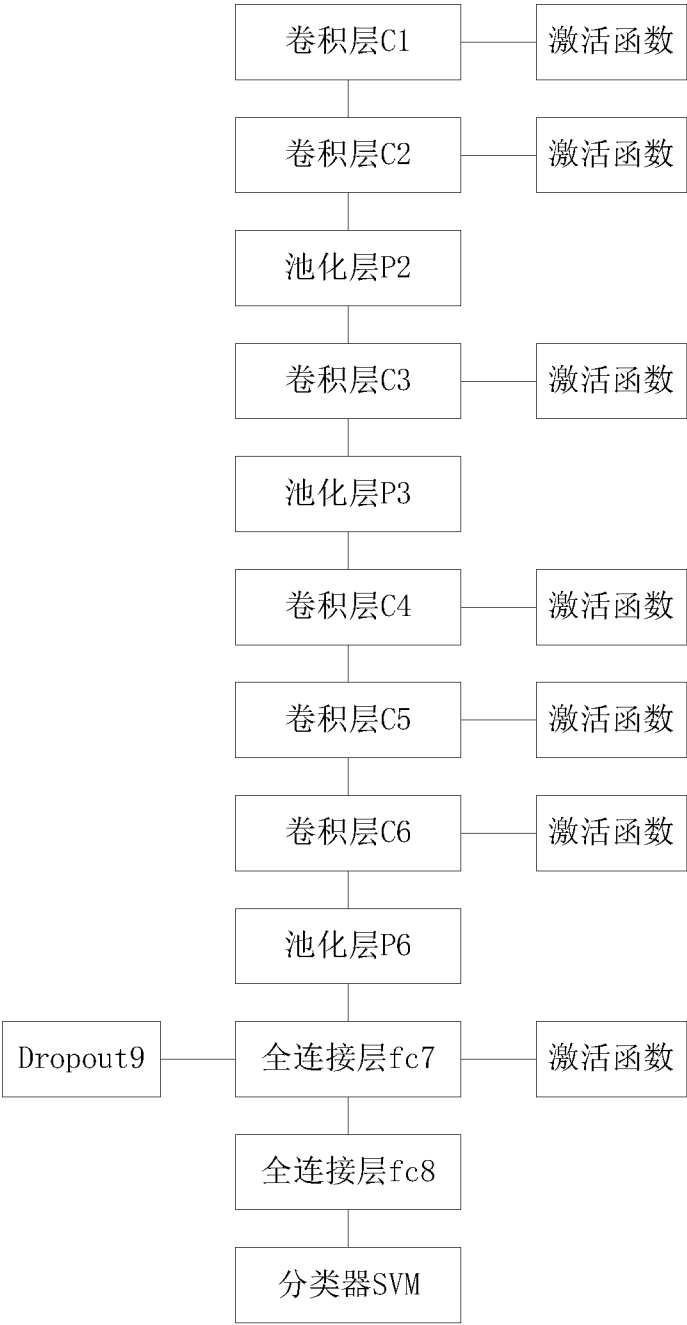


图 2

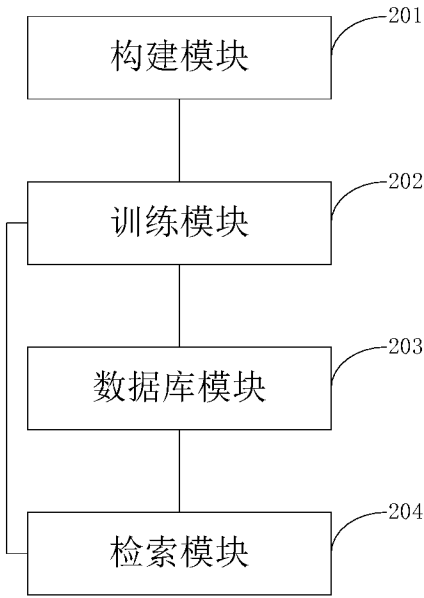


图 3

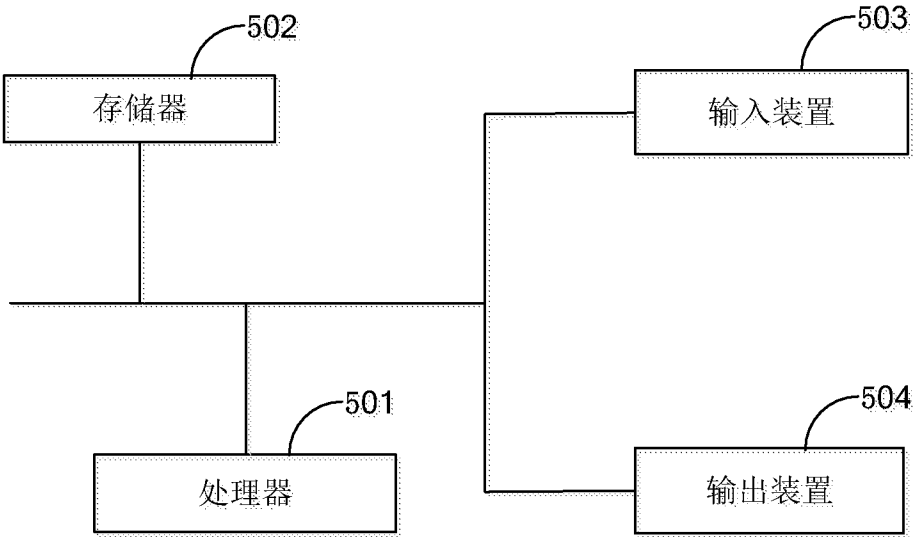


图 4

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2016/103945

A. CLASSIFICATION OF SUBJECT MATTER

G06F 17/30 (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)

G06F

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)

CNPAT, WPI, EPODOC, CNKI, GOOGLE: roll, product, classifier, module, video, search, frame, scene, change, train, neural, neural network

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 105930402 A (LETV HOLDING (BEIJING) CO., LTD. et al.) 07 September 2016 (07.09.2016) claims 1-16, and description, paragraphs [0110]-[0113]	1-19
A	CN 104156464 A (CHONGQING INSTITUTE OF GREEN AND INTELLIGENT TECHNOLOGY, CHINESE ACADEMY OF SCIENCES) 19 November 2014 (19.11.2014) description, paragraphs [0040]-[0057]	1-19
A	CN 104980625 A (CHINA DIGITAL VIDEO (BEIJING) LIMITED) 14 October 2015 (14.10.2015) the whole document	1-19
A	US 2010302269 A1 (KABUSHIKI KAISHA TOSHIBA) 02 December 2010 (02.12.2010) the whole document	1-19

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents:	“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
“A” document defining the general state of the art which is not considered to be of particular relevance	“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
“E” earlier application or patent but published on or after the international filing date	“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
“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)	“&” document member of the same patent family
“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	

Date of the actual completion of the international search 19 January 2017	Date of mailing of the international search report 04 February 2017
Name and mailing address of the ISA State Intellectual Property Office of the P. R. China No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088, China Facsimile No. (86-10) 62019451	Authorized officer ZHAO, Tianqi Telephone No. (86-10) 61648112

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.
PCT/CN2016/103945

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 105930402 A	07 September 2016	None	
CN 104156464 A	19 November 2014	None	
CN 104980625 A	14 October 2015	None	
US 2010302269 A1	02 December 2010	JP 2010276691 A	09 December 2010

国际检索报告

国际申请号

PCT/CN2016/103945

A. 主题的分类

G06F 17/30(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)

G06F

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))

CNPAT, WPI, EPODOC, CNKI, GOOGLE: 卷积, 神经网络, 视频, 检索, 模型, 训练, 分类器, 转场帧, roll, product, classifier, module, video, search, frame, scene, change, train, neural

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
PX	CN 105930402 A (乐视控股北京有限公司 等) 2016年 9月 7日 (2016 - 09 - 07) 权利要求1-16, 说明书第110-113段	1-19
A	CN 104156464 A (中国科学院重庆绿色智能技术研究院) 2014年 11月 19日 (2014 - 11 - 19) 说明书第40-57段	1-19
A	CN 104980625 A (新奥特北京视频技术有限公司) 2015年 10月 14日 (2015 - 10 - 14) 全文	1-19
A	US 2010302269 A1 (KABUSHIKI KAISHA TOSHIBA) 2010年 12月 2日 (2010 - 12 - 02) 全文	1-19

☐ 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

2017年 1月 19日

国际检索报告邮寄日期

2017年 2月 4日

ISA/CN的名称和邮寄地址

中华人民共和国国家知识产权局(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

赵天奇

电话号码 (86-10)61648112

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2016/103945

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	105930402	A	2016年 9月 7日	无			
CN	104156464	A	2014年 11月 19日	无			
CN	104980625	A	2015年 10月 14日	无			
US	2010302269	A1	2010年 12月 2日	JP	2010276691	A	2010年 12月 9日