

# (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 9 月 10 日 (10.09.2020)



(10) 国际公布号  
**WO 2020/177722 A1**

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

(21) 国际申请号: PCT/CN2020/077809

(22) 国际申请日: 2020 年 3 月 4 日 (04.03.2020)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:  
201910168236.9 2019年3月6日 (06.03.2019) CN

(71) 申请人: 腾讯科技(深圳)有限公司 (TENCENT TECHNOLOGY (SHENZHEN) COMPANY LIMITED) [CN/CN]; 中国广东省深圳市南山区高新区科技中一路腾讯大厦 35 层, Guangdong 518057 (CN)。

(72) 发明人: 唐永毅 (TANG, Yongyi); 中国广东省深圳市南山区高新区科技中一路腾讯大厦 35 层, Guangdong 518057 (CN)。 马林 (MA, Lin); 中国广东省深圳市南山区高新区科技中一路腾讯大厦 35 层, Guangdong 518057 (CN)。 刘威 (LIU, Wei); 中国广东省深圳市南山区高新区科技中一路腾讯大厦 35 层, Guangdong 518057 (CN)。

(74) 代理人: 北京三高永信知识产权代理有限公司 (BEIJING SAN GAO YONG XIN INTELLECTUAL PROPERTY AGENCY CO., LTD.); 中国北京市海淀区学院路蓟门里和景园 A 座 1 单元 102 室, Beijing 100088 (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,

(54) Title: METHOD FOR VIDEO CLASSIFICATION, METHOD AND DEVICE FOR MODEL TRAINING, AND STORAGE MEDIUM

(54) 发明名称: 一种视频分类的方法、模型训练的方法、设备及存储介质

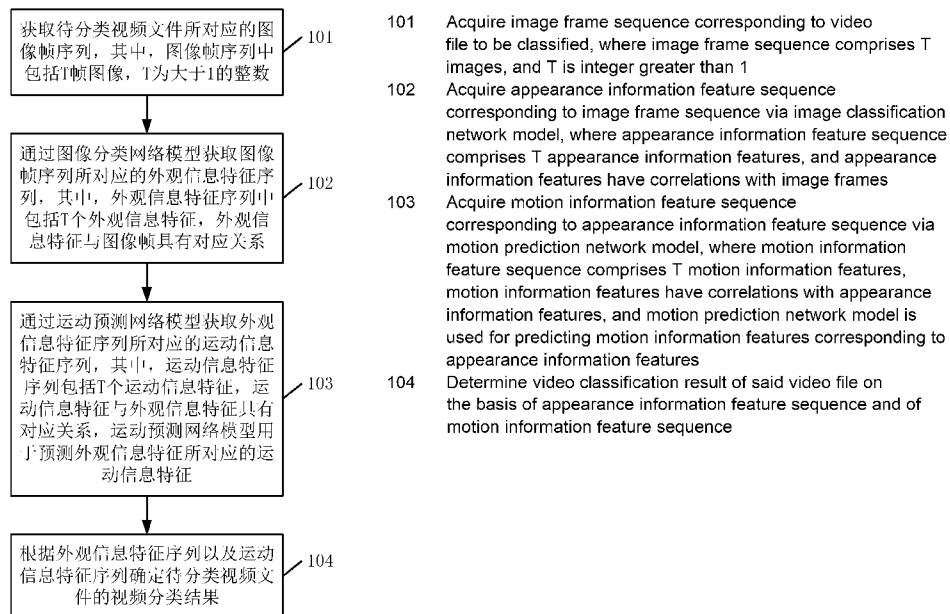


图 4

(57) Abstract: The present application relates to the field of artificial intelligence. Disclosed is a method for video classification, comprising: acquiring an image frame sequence corresponding to a video file to be classified, where the image frame sequence comprises T images; acquiring an appearance information feature sequence corresponding to the image frame sequence via an image classification network model, the appearance information feature sequence comprising T appearance information features, and the appearance information features having correlations with image frames; acquiring a motion information feature sequence corresponding to the

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, 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))。

appearance information feature sequence via a motion prediction network model, the motion information feature sequence comprising T motion information features, and the motion prediction network model being used for predicting motion information features corresponding to the appearance information features; determining a video classification result of said video file on the basis of the appearance information feature sequence and of the motion information feature sequence. Also provided in the present application are a method and device for model training. The present application saves time and resources spent on computation, thus increasing the efficiency of data processing.

(57) 摘要: 本申请公开了一种视频分类的方法, 涉及人工智能领域, 包括: 获取待分类视频文件所对应的图像帧序列, 其中, 图像帧序列中包括T帧图像; 通过图像分类网络模型获取图像帧序列所对应的外观信息特征序列, 外观信息特征序列中包括T个外观信息特征, 外观信息特征与图像帧具有对应关系; 通过运动预测网络模型获取外观信息特征序列所对应的运动信息特征序列, 运动信息特征序列包括T个运动信息特征, 运动预测网络模型用于预测外观信息特征所对应的运动信息特征; 根据外观信息特征序列以及运动信息特征序列确定待分类视频文件的视频分类结果。本申请还提供一种模型训练的方法以及装置。本申请能够节省计算时间和资源, 从而提升了数据处理的效率。

## 一种视频分类的方法、模型训练的方法、设备及存储介质

本申请要求于 2019 年 03 月 06 日提交的申请号为 2019101682369、发明名称为“一种视频分类的方法、模型训练的方法及装置”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

本申请涉及人工智能领域，尤其涉及一种视频分类的方法、模型训练的方法、设备及存储介质。

### 背景技术

视频分类是指给定一个视频片段，对其中包含的内容进行分类。视频类别通常包括动作（比如做蛋糕）、场景（比如海滩）以及物体（比如桌子）等。其中，又以视频动作分类最为热门，毕竟动作本身就包含动态的因素，不是静态的图像所能描述的，因此，基于运动特征对视频进行分类是很常见的。

目前，通常使用的视频分类技术主要是，首先对于从视频红绿蓝（red green blue, RGB）图像中提取的外观信息特征进行分类，然后进一步利用从光流图像中提取的运动信息特征来提升视频分类的性能。在提取运动信息特征的过程中，需要根据连续两帧或多帧的视频 RGB 图片计算光流图像，进而利用卷积神经网络或其他方法获得视频的运动信息特征。

然而，在现有方案中需要根据至少两帧 RGB 图像对光流图像进行计算，对于视频而言，以 24 帧每秒的分辨率计算，短视频通常包含 15 秒以上的内容，而长视频常会长达几十分钟。因此，需要大量的时间和空间对光流图像进行计算，以获得视频的光流图像表示，从而导致数据处理的效率较低，且占用的计算资源和存储资源较多。

### 发明内容

本申请实施例提供了一种视频分类的方法、模型训练的方法、设备及存储介质，能够实现对视频的自动分类，无需根据视频的外观信息特征计算光流图像，以利用光流图像计算运动信息特征。而是直接通过运动预测网络模型即可获取外观信息特征对应的运动信息特征，由此节省了计算时间和资源，从而提升了数据处理的效率。

有鉴于此，本申请第一方面提供一种视频分类的方法，应用于计算机设备，包括：

获取待分类视频文件所对应的图像帧序列，其中，所述图像帧序列中包括 T 帧图像，所述 T 为大于 1 的整数；

获取所述图像帧序列所对应的外观信息特征序列，其中，所述外观信息特征序列中包括 T 个外观信息特征，所述外观信息特征与所述图像帧具有对应关系；

获取所述外观信息特征序列所对应的运动信息特征序列，其中，所述运动信息特征序列包括 T 个运动信息特征，所述运动信息特征与所述外观信息特征具有对应关系；

根据所述外观信息特征序列以及所述运动信息特征序列确定所述待分类视频文件的视频分类结果。

本申请第二方面提供一种模型训练的方法，包括：

获取待训练视频文件所对应的光流图序列，其中，所述光流图序列中包括多个光流图；

获取所述待训练视频文件所对应的待训练外观信息特征序列，其中，所述待训练外观信息特征序列包括多个待训练外观信息特征；

根据所述待训练视频文件所对应的所述光流图序列，获取真实运动信息特征序列，其中，所述真实运动信息特征序列包括 T 个真实运动信息特征，所述 T 为大于 1 的整数；

通过待训练运动预测网络模型获取所述待训练外观信息特征序列所对应的预测运动信息特征序列，其中，所述预测运动信息特征序列包括 T 个预测运动信息特征；

根据所述真实运动信息特征序列与所述预测运动信息特征序列，确定模型参数；

采用所述模型参数对所述待训练运动预测网络模型进行训练，得到运动预测网络模型。

本申请第三方面提供一种视频分类装置，包括：

获取模块，用于获取待分类视频文件所对应的图像帧序列，其中，所述图像帧序列中包括 T 帧图像，所述 T 为大于 1 的整数；

所述获取模块，还用于获取所述图像帧序列所对应的外观信息特征序列，其中，所述外

观信息特征序列中包括 T 个外观信息特征，所述外观信息特征与所述图像帧具有对应关系；  
 所述获取模块，还用于获取所述外观信息特征序列所对应的运动信息特征序列，其中，所述运动信息特征序列包括 T 个运动信息特征，所述运动信息特征与所述外观信息特征具有对应关系，所述运动预测网络模型用于预测所述外观信息特征所对应的所述运动信息特征；

确定模块，用于根据所述获取模块获取的所述外观信息特征序列以及所述运动信息特征序列确定所述待分类视频文件的视频分类结果。

在一种可能的设计中，在本申请实施例的第三方面的第一种实现方式中，

所述获取模块，具体用于通过图像分类网络模型获取所述图像帧序列所对应的外观信息特征序列；通过运动预测网络模型获取所述外观信息特征序列所对应的运动信息特征序列，所述运动预测网络模型用于预测所述外观信息特征所对应的所述运动信息特征。

在一种可能的设计中，在本申请实施例的第三方面的第一种实现方式中，

所述获取模块，具体用于获取所述待分类视频文件；

对所述待分类视频文件进行解码处理，得到所述图像帧序列，其中，每个帧图像具有相同的采样频率。

在一种可能的设计中，在本申请实施例的第三方面的第二种实现方式中，

所述获取模块，具体用于通过所述图像分类网络模型获取第 t 帧图像所对应的外观信息特征，其中，所述 t 大于或等于 1，且小于或等于所述 T 的整数；

当获取到所述 T 帧图像中每帧图像所对应的外观信息特征时，根据所述 T 个外观信息特征生成所述外观信息特征序列。

在一种可能的设计中，在本申请实施例的第三方面的第三种实现方式中，

所述获取模块，具体用于获取第 i 个步骤所对应的第 t 个外观信息特征，其中，所述 i 为大于或等于 1 的整数，所述 t 大于或等于 1，且小于或等于所述 T 的整数；

获取第(i-1)个步骤所对应的第(t+1)个运动信息特征以及第(t-1)个运动信息特征；

通过所述运动预测网络模型，根据所述第 t 个外观信息特征、所述第(t+1)个运动信息特征以及第(t-1)个运动信息特征，确定所述第 i 个步骤所对应的第 t 个运动信息特征；

当获取到所述 T 个外观信息特征中每个外观信息特征所对应的运动信息特征时，根据所述 T 个运动信息特征生成所述运动信息特征序列。

在一种可能的设计中，在本申请实施例的第三方面的第四种实现方式中，

所述获取模块，具体用于通过运动预测网络模型的更新门，根据所述第 t 个外观信息特征、所述第(t+1)个运动信息特征以及第(t-1)个运动信息特征，计算得到所述第 i 个步骤所对应的第(t+1)个更新门向量以及第(t-1)个更新门向量；

通过所述运动预测网络模型的输出层，根据所述第 t 个外观信息特征、所述第(t+1)个运动信息特征以及第(t-1)个运动信息特征，计算得到所述第 i 个步骤所对应的第(t+1)个输出层向量以及第(t-1)个输出层向量；

通过所述运动预测网络模型的隐藏层，根据所述第 t 个外观信息特征、所述第(t+1)个运动信息特征、第(t-1)个运动信息特征、所述第(t+1)个更新门向量以及所述第(t-1)个更新门向量，计算得到所述第 i 个步骤所对应的第 t 个隐藏层特征向量；

通过所述运动预测网络模型的最终输出层，根据所述第 t 个隐藏层特征向量、所述第 i 个步骤所对应的第(t+1)个输出层向量以及第(t-1)个输出层向量，计算得到所述第 i 个步骤所对应的第 t 个运动信息特征。

在一种可能的设计中，在本申请实施例的第三方面的第五种实现方式中，

所述获取模块，具体用于采用如下方式计算所述第(t+1)个更新门向量以及所述第(t-1)个更新门向量：

$$r_{t-1} = \sigma(W_r x_t + U_{r,t-1} s_{t-1}^{i-1});$$

$$r_{t+1} = \sigma(W_r x_t + U_{r,t+1} s_{t+1}^{i-1});$$

其中，所述  $r_{t-1}$  表示所述第 i 个步骤所对应的所述第(t-1)个更新门向量，所述 t 表示第 t 个时刻，所述  $r_{t+1}$  表示所述第 i 个步骤所对应的所述第(t+1)个更新门向量，所述  $\sigma(\cdot)$  表示线

性整流函数, 所述  $W_r$  表示第一参数矩阵, 所述  $U_{r,t-1}$  表示所述第(t-1)个更新门向量所对应的第二参数矩阵, 所述  $U_{r,t+1}$  表示所述第(t+1)个更新门向量所对应的第三参数矩阵, 所述  $x_t$  表示所述第t个外观信息特征, 所述  $s_{t-1}^{i-1}$  表示所述第(i-1)个步骤所对应的所述第(t-1)个运动信息特征, 所述  $s_{t+1}^{i-1}$  表示所述第(i-1)个步骤所对应的所述第(t+1)个运动信息特征;

采用如下方式计算所述第(t+1)个输出门向量以及所述第(t-1)个输出门向量:

$$[z_{t+1}, z, z_{t-1}] = \text{soft max}([\tilde{z}_{t+1}, 1, \tilde{z}_{t-1}]);$$

$$\tilde{z}_{t-1} = \sigma(W_z x_t + U_{z,t-1} s_{t-1}^{i-1});$$

$$\tilde{z}_{t+1} = \sigma(W_z x_t + U_{z,t+1} s_{t+1}^{i-1});$$

其中, 所述  $z_{t+1}$  表示所述第i个步骤所对应的所述第(t+1)个输出门向量, 所述  $z_{t-1}$  表示所述第i个步骤所对应的所述第(t-1)个输出门向量, 所述  $z$  表示预设输出门向量, 所述  $\tilde{z}_{t-1}$  表示所述第i个步骤所对应的第(t-1)个待处理输出门向量,  $\tilde{z}_{t+1}$  表示所述第i个步骤所对应的第(t+1)个待处理输出门向量, 所述  $\text{soft max}()$  表示归一化处理函数, 所述  $W_z$  表示第四参数矩阵, 所述  $U_{z,t-1}$  表示所述第(t-1)个输出门向量所对应的第五参数矩阵, 所述  $U_{z,t+1}$  表示所述第(t+1)个输出门向量所对应的第六参数矩阵;

采用如下方式计算所述第t个隐藏层特征向量:

$$h_t = \phi\left(W_h x_t + U_h \left[s_{t+1}^{i-1} \circ r_{t+1}, s_{t-1}^{i-1} \circ r_{t-1}\right]^T\right);$$

其中, 所述  $h_t$  表示所述第i个步骤所对应的所述第t个隐藏层特征向量, 所述  $\phi()$  表示sigmoid函数, 所述  $W_h$  表示第七参数矩阵, 所述  $U_h$  表示第八参数矩阵, 所述  $[\ ]^T$  表示矩阵转置, 所述  $\circ$  表示点乘;

采用如下方式计算所述第t个运动信息特征:

$$s_t^i = z \circ h_t + z_{t-1} \circ s_{t-1}^{i-1} + z_{t+1} \circ s_{t+1}^{i-1};$$

其中, 所述  $s_t^i$  表示所述第i个步骤所对应的所述第t个运动信息特征。

在一种可能的设计中, 在本申请实施例的第三方面的第六种实现方式中,

所述确定模块, 具体用于获取所述外观信息特征序列以及所述运动信息特征序列所对应的视频类别向量, 其中, 所述视频类别向量包括C个视频类别元素, 所述C为大于1的整数;

从所述视频类别向量中确定至少一个目标视频类别元素;

根据所述至少一个目标视频类别元素生成所述待分类视频文件的所述视频分类结果。

本申请第四方面提供一种模型训练装置, 包括:

获取模块, 用于获取待训练视频文件所对应的光流图序列, 其中, 所述光流图序列中包括多个光流图;

所述获取模块, 还用于获取所述待训练视频文件所对应的待训练外观信息特征序列, 其中, 所述待训练外观信息特征序列包括多个待训练外观信息特征;

所述获取模块, 还用于根据所述待训练视频文件所对应的所述光流图序列, 获取真实运动信息特征序列, 其中, 所述真实运动信息特征序列包括T个真实运动信息特征, 所述T为大于1的整数;

所述获取模块, 还用于通过待训练运动预测网络模型获取所述待训练外观信息特征序列所对应的预测运动信息特征序列, 其中, 所述预测运动信息特征序列包括T个预测运动信息

特征；

确定模块，用于根据所述获取模块获取的所述真实运动信息特征序列与所述获取模块获取的所述预测运动信息特征序列，确定模型参数；

训练模块，用于采用所述确定模块确定的所述模型参数对所述待训练运动预测网络模型进行训练，得到运动预测网络模型。

在一种可能的设计中，在本申请实施例的第四方面的第一种实现方式中，

所述确定模块，具体用于采用损失函数对所述真实运动信息特征序列与所述预测运动信息特征序列进行计算，得到目标损失结果；

根据所述目标损失结果确定所述模型参数。

在一种可能的设计中，在本申请实施例的第四方面的第二种实现方式中，

所述确定模块，具体用于采用如下方式计算所述目标损失结果：

$$L_{MSE} = \frac{1}{T} \sum_t^T \left\| \tilde{S}^t - S^t \right\|_2^2;$$

其中，所述  $L_{MSE}$  表示所述目标损失结果，所述  $T$  表示所述真实运动信息特征序列中的真实运动信息特征个数，所述  $\tilde{S}^t$  表示所述真实运动信息特征序列中第  $t$  个真实运动信息特征，所述  $S^t$  表示所述预测运动信息特征序列中第  $t$  个预测运动信息特征，所述  $t$  为大于或等于 1，且小于或等于所述  $T$  的整数。

本申请第五方面提供一种计算机设备，包括：存储器、收发器、处理器以及总线系统；

其中，所述存储器用于存储程序；

所述处理器用于执行所述存储器中的程序，包括如下步骤：

获取待分类视频文件所对应的图像帧序列，其中，所述图像帧序列中包括  $T$  帧图像，所述  $T$  为大于 1 的整数；

获取所述图像帧序列所对应的外观信息特征序列，其中，所述外观信息特征序列中包括  $T$  个外观信息特征，所述外观信息特征与所述图像帧具有对应关系；

获取所述外观信息特征序列所对应的运动信息特征序列，其中，所述运动信息特征序列包括  $T$  个运动信息特征，所述运动信息特征与所述外观信息特征具有对应关系；

根据所述外观信息特征序列以及所述运动信息特征序列确定所述待分类视频文件的视频分类结果；

所述总线系统用于连接所述存储器以及所述处理器，以使所述存储器以及所述处理器进行通信。

本申请第六方面提供一种服务器，包括：存储器、收发器、处理器以及总线系统；

其中，所述存储器用于存储程序；

所述处理器用于执行所述存储器中的程序，包括如下步骤：

获取待训练视频文件所对应的光流图序列，其中，所述光流图序列中包括多个光流图；

获取所述待训练视频文件所对应的待训练外观信息特征序列，其中，所述待训练外观信息特征序列包括多个待训练外观信息特征；

根据所述待训练视频文件所对应的所述光流图序列，获取真实运动信息特征序列，其中，所述真实运动信息特征序列包括  $T$  个真实运动信息特征，所述  $T$  为大于 1 的整数；

通过待训练运动预测网络模型获取所述待训练外观信息特征序列所对应的预测运动信息特征序列，其中，所述预测运动信息特征序列包括  $T$  个预测运动信息特征；

根据所述真实运动信息特征序列与所述预测运动信息特征序列，确定模型参数；

采用所述模型参数对所述待训练运动预测网络模型进行训练，得到运动预测网络模型；

所述总线系统用于连接所述存储器以及所述处理器，以使所述存储器以及所述处理器进行通信。

本申请的第七方面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中存储有指令，当其在计算机上运行时，使得计算机执行上述各方面所述的方法。

附图说明

图 1 为本申请实施例中视频分类系统的一个架构示意图；

图 2 为本申请实施例中结合外观信息特征和运动信息特征的一个视频分类流程图；

图 3 为本申请实施例中视频分类的一个应用场景示意图；

图 4 为本申请实施例中视频分类的方法一个实施例示意图；

图 5 为本申请实施例中运动预测网络模型的一个结构单元示意图；

图 6 为本申请实施例中运动预测网络模型的一个递归展开示意图；

图 7 为本申请实施例中模型训练的方法一个实施例示意图；

图 8 为本申请实施例中视频分类装置一个实施例示意图；

图 9 为本申请实施例中模型训练装置一个实施例示意图；

图 10 为本申请实施例中终端设备一个结构示意图；

图 11 为本申请实施例中服务器一个结构示意图。

### 具体实施方式

本申请实施例提供了一种视频分类的方法、模型训练的方法及装置，能够实现对视频的自动分类，无需根据视频的外观信息特征计算光流图像，以利用光流图像计算运动信息特征。而是直接通过运动预测网络模型即可获取外观信息特征对应的运动信息特征，由此节省了计算时间和资源，从而提升了数据处理效率。

本申请的说明书和权利要求书及上述附图中的术语“第一”、“第二”、“第三”、“第四”等（如果存在）是用于区别类似的对象，而不必用于描述特定的顺序或先后次序。应该理解这样使用的数据在适当情况下可以互换，以便这里描述的本申请的实施例例如能够以除了在这里图示或描述的那些以外的顺序实施。此外，术语“包括”和“对应于”以及他们的任何变形，意图在于覆盖不排除的包含，例如，包含了一系列步骤或单元的过程、方法、系统、产品或设备不必限于清楚地列出的那些步骤或单元，而是可包括没有清楚地列出的或对于这些过程、方法、产品或设备固有的其它步骤或单元。

应理解，本申请可以应用于视频分类的场景，视频分类即可以为视频贴上“标签”，让视频归属于某一个或某多个类别。在数量庞大的视频数据库中，分类和标签是搜索视频的重要依据。视频能否被更多人看到，能否受大家欢迎，很大程度上取决于分类和标签填写是否恰当。视频分类已覆盖互联网、电视以及移动终端等，并且兼具影视、综艺以及资讯等内容形态，贯通视频内容制作、播出以及发行等环节，成为真正意义的互联网电视媒体，影响视频行业格局及全媒时代的大格局。更具体地，本申请所提供的视频分类方法可以应用于如下几类视频网站。

第一类视频网站为对等（peer-to-peer, P2P）网络电视，即采用 P2P 技术来实现节目的传输。这种技术的最大优点就是每个用户可以直接和其他用户进行连接，从而突破了服务器和带宽的瓶颈，使用的用户越多，网络互联的效果越好。此种视频网站的特点包含但不限于运营成本低，主要表现在服务器、带宽成本低；需要安装 P2P 插件或者软件；直播类节目流畅度较之点播类节目流畅度高；节目数量有限；主动发布内容，即内容寄存在自身服务器；节目观看前需要一定的缓冲时间。

第二类视频网站为 P2P 下载观看的视频网站，即实现影视文件观看与下载同步。这种技术的改善了传统 P2P 下载只能等文件下载完后再观看的缺点。此种视频网站的特点包含但不限于表现为中间平台，即内容多为网友提供，然后分享下载，此外，也有服务器自己发布内容的；需要安装 P2P 软件；点播节目为主；节目数量众多。

第三类视频网站为视频分享网站，即用户创造内容网站（user generated content, UGC），这种网站最大的特点就是采用固态存储器与动画编辑器视频（flash video, FLASH FLV）视频播放技术，并含有视频上传、分享以及评价等功能。此种视频网站的特点包含但不限于用户体验较好，不需安装软件，可时间即点即播；对服务器及带宽占用较大，运营成本较高；内容多，且多为网友上传；盈利能力较弱。

第四类视频网站为电信平台，即指电信运营商自行开设的视频网站。此种视频网站特点包含但不限于采取按次、包月等收费方式，费用支付可以通过手机、电话以及宽带费等；电信运营商提供服务器和带宽，节目提供商提供内容或者整体提供网站平台及内容；节目质量参差不齐；为早期中国网络视频表现模式。

第五类视频网站为视频搜索网站，即提供视频搜索服务的网站。此种视频网站特点包含但不限于搜索互联网的视频资源；视频搜索技术难度大，研发成本高，较先进的帧搜索需要建立庞大的索引库；部分视频网站以及搜索引擎已经研发视频搜索功能。

第六类视频网站为垂直视频网站，即以一个行业为中心，提供专业的视频服务。此种视频网站特点包含但不限于内容局限于一个行业，用户定位精确；视频播放多为载体，更多的依赖于点评模式、企业与个人（business to customer, B2C）或者个人与个人（customer to customer, C2C）等电子商务模式。

第七类视频网站为门户视频网站，即为门户设立的视频频道。此种视频网站特点包含但不限于多种视频表现模式混合；基于视频网站的盈利能力，版权购买力较强；与其门户其他产品的结合度高，具有组合优势；运营没有专业视频网站细致。

为了便于理解，本申请提出了一种视频分类的方法，该方法应用于图 1 所示的视频分类系统，请参阅图 1，图 1 为本申请实施例中视频分类系统的一个架构示意图，如图所示，视频分类装置可以部署于服务器，也可以部署于具有较高计算能力的客户端，也即是该视频分类装置可以部署于计算机设备，下面将以视频分类装置部署于服务器为例进行介绍。服务器根据视频文件的 RGB 图像帧提取外观信息特征，然后根据提取到的外观信息特征预测对应视频的光流图像的运动信息特征，在不显式计算光流图像的情况下，根据外观信息特征序列预测视频的运动信息特征序列，并结合视频的外观信息特征序列对视频进行特征描述，提升视频分类的准确性，并且提升目前视频分类模型的性能。本申请所提供的视频特征提取方法，还可以结合各种特征分类框架被应用于服务器端或云端，由此对视频数据库中的视频自动进行多类别的分类。

需要说明的是，客户端部署于终端设备上，其中，终端设备包含但不限于平板电脑、笔记本电脑、掌上电脑、手机、语音交互设备及个人电脑（personal computer, PC），此处不做限定。其中，语音交互设备包含但不限于智能音响以及智能家电。

为了便于说明，请参阅图 2，图 2 为本申请实施例中结合外观信息特征和运动信息特征的一个视频分类流程图，如图所示，本申请所提供的视频分类方法可以分为三个部分，分别为视频预处理，特征提取及预测以及特征分类。下面将针对各个部分的功能和流程进行介绍，具体地：

步骤 S11 中，获取视频文件；

步骤 S12 中，对输入的该视频文件进行编码，获取其对应的 RGB 图像帧，可以理解的是，RGB 图像帧的格式不做限定；

步骤 S21 中，以 RGB 图像帧作为图像分类网络模型的输入，由图像分类网络模型对每个 RGB 图像帧进行外观信息特征的提取，从而组成该视频文件的外观信息特征序列；

步骤 S22 中，以外观信息特征序列作为运动预测网络模型的输入，由运动预测网络模型预测得到该视频文件的运动信息特征序列；

步骤 S31 中，以运动信息特征序列以及外观信息特征序列作为视频分类网络模型的输入，对视频文件进行分类，得到视频文件的类别信息。

基于上述流程，本申请提供的视频分类的应用场景如图 3 所示，请参阅图 3，图 3 为本申请实施例中视频分类的一个应用场景示意图，如图所示，以视频分类服务为例，服务器接收用户上传的视频文件，并给视频文件进行处理，得到 RGB 图像帧。将 RGB 图像帧输入至图像分类网络模型，由此输出外观信息特征序列，再将外观信息特征序列输入至运动预测网络模型，由此输出运动信息特征序列，再将外观信息特征序列以及运动信息特征序列输入至视频分类网络模型，由此输出视频分类结果，即视频文件属于 A 类型、B 类型、C 类型以及 D 类型中的至少一类。

在上述图 2 所示实施例中，以该视频分类过程通过多个模型实现为例进行了说明，在一种可能实现方式中，上述视频分类过程还可以不通过模型实现，相应地，上述步骤中则可以直接执行相应的步骤，而非通过神经网络、图像分类网络模型、运动预测网络模型、视频分类网络模型。本申请实施例对具体采用哪种实现方式不作限定。

结合上述介绍，下面将对本申请中视频分类的方法进行介绍，请参阅图 4，本申请实施例中视频分类的方法一个实施例包括：

101、获取待分类视频文件所对应的图像帧序列，其中，图像帧序列中包括 T 帧图像，T 为大于 1 的整数；

本实施例中，视频分类装置获取待分类视频文件，其中，视频分类装置可以部署于服务器上，也可以部署于终端设备。待分类视频文件可以通过客户端上传的视频，也可以是预先存储在数据库中的视频，此处不做限定。视频分类装置对待分类视频文件进行解码处理，

从而得到图像帧序列，图像帧序列中包括 T 帧图像，T 为大于 1 的整数，T 表示图像帧序列中图像帧的总数，且一帧图像对应一个时刻，即 T 帧图像对应 T 个时刻。

102、通过图像分类网络模型获取图像帧序列所对应的外观信息特征序列，其中，外观信息特征序列中包括 T 个外观信息特征，外观信息特征与图像帧具有对应关系；

本实施例中，视频分类装置将图像帧序列中的 T 帧图像作为图像分类网络模型的输入，通过该图像分类网络模型输出外观信息特征序列，其中，外观信息特征序列包括 T 个外观信息特征，每个图像帧对应一个外观信息特征。

该步骤 102 为获取图像帧序列所对应的外观信息特征序列的过程，在该步骤 102 中仅以该外观信息特征序列通过图像分类网络模型获取得到为例进行说明，也可以不通过该图像分类网络模型实现，本申请实施例对此不作限定。

103、通过运动预测网络模型获取外观信息特征序列所对应的运动信息特征序列，其中，运动信息特征序列包括 T 个运动信息特征，运动信息特征与外观信息特征具有对应关系，运动预测网络模型用于预测外观信息特征所对应的运动信息特征；

本实施例中，视频分类装置将外观信息特征序列作为运动预测网络（Motion Prediction Network, MoNet）模型的输入，通过该运动预测网络模型输出运动信息特征序列，其中，运动信息特征序列包括 T 个运动信息特征，即每个外观信息特征对应一个运动信息特征，每个图像帧也对应一个运动信息特征。

该步骤 103 为获取外观信息特征序列所对应的运动信息特征序列的过程，在该步骤 103 中仅以该运动信息特征序列通过运动预测网络模型获取得到为例进行说明，也可以不通过该运动预测网络模型实现，本申请实施例对此不作限定。

104、根据外观信息特征序列以及运动信息特征序列确定待分类视频文件的视频分类结果。

本实施例中，视频分类装置根据外观信息特征序列以及运动信息特征序列，可以对视频文件进行分类，由此输出待分类视频的视频分类结果。通常情况下，待分类视频文件对应至少一个视频分类结果，比如，待分类视频文件可以属于 A 类型的视频，又比如，待分类视频文件可以同时属于 A 类型和 B 类型的视频。

本申请实施例中，提供了一种视频分类的方法，首先获取待分类视频文件所对应的图像帧序列，其中，图像帧序列中包括 T 帧图像，T 为大于 1 的整数，然后通过图像分类网络模型获取图像帧序列所对应的外观信息特征序列，其中，外观信息特征序列中包括 T 个外观信息特征，外观信息特征与图像帧具有对应关系，再通过运动预测网络模型获取外观信息特征序列所对应的运动信息特征序列，其中，运动信息特征序列包括 T 个运动信息特征，运动信息特征与外观信息特征具有对应关系，运动预测网络模型用于预测外观信息特征所对应的运动信息特征，最后可以根据外观信息特征序列以及运动信息特征序列确定待分类视频文件的视频分类结果。通过上述方式，能够实现对视频的自动分类，无需根据视频的外观信息特征计算光流图像，以利用光流图像计算运动信息特征。而是直接通过运动预测网络模型即可获取外观信息特征对应的运动信息特征，由此节省了计算时间和资源，从而提升了数据处理的效率。

可选地，在上述图 3 对应的实施例的基础上，本申请实施例提供的视频分类的方法第一个可选实施例中，获取待分类视频文件所对应的图像帧序列，可以包括：

获取待分类视频文件；

对待分类视频文件进行解码处理，得到图像帧序列，其中，每个帧图像具有相同的采样频率。

本实施例中，提供了一种获取图像帧序列的方法，首先，视频分类装置需要获取待分类视频文件，然后利用音视频解码工具以一定的采样频率（比如每秒 1 帧）对待分类视频文件进行解码，从而得到 RGB 图像帧序列以及音频信号序列。

可以理解的是，待分类视频文件的视频格式包含但不限于运动图像专家组（motion picture experts group, MPEG）格式、音频视频交错（audio video interleaved, AVI）、格式、高级流格式（advanced streaming format, ASF）、微软媒体视频（Windows media video, WMV）格式、第三代合作伙伴项目计划文件格式（3rd generation partnership project file format, 3GP）、多媒体容器文件格式（multimedia container file format, MKV）、流媒体格式（flash video）以及视频容器可变比特率文件格式（RealMedia variable bitrate file format, RMVB）。

可以理解的是，音视频解码工具具体可以是飞梭 MPEG（fast forward MPEG, ffmpeg），

ffmpeg 是一套可以用来记录以及转换数字音频和视频,并能将其转化为流的开源计算机程序。音视频解码工具采用的采样帧率也称为采样率或者采样速度,采样频率定义了每秒从连续信号中提取并组成离散信号的采样个数,它用赫兹(hertz, Hz)来表示。采样频率的倒数叫作采样周期或采样时间,它是采样之间的时间间隔。

可以理解的是,RGB 图像帧序列中的 RGB 图像帧格式包含但不仅限于位图(bitmap, BMP)格式、个人电脑交换(personal computer exchange, PCX)格式、标签图像文件格式(tag image file format, TIFF)、图形交换格式(graphics interchange format, GIF)、联合照片专家组(joint photographic expert group, JPEG)、已标记的图形(tagged graphics, TGA)格式、可交换的图像文件格式(exchangeable image file Format, EXIF)、闪光照片(kodak flash PiX, FPX)格式、可缩放矢量图形(scalable vector graphics, SVG)、便携式网络图形(portable network graphics, PNG)、微软图元文件格式(Windows metafile format, WMF)、封装式页描述语言(encapsulated post script, EPS)格式、高动态范围成像(high dynamic range imaging, HDRI)以及图纸交换格式(drawing exchange format, DXF)。

其次,本申请实施例中,介绍了一种视频分类装置获取待分类视频文件所对应的图像帧序列的方式,即先获取待分类视频文件,然后对待分类视频文件进行解码处理,得到图像帧序列,其中,每个帧图像具有相同的采样频率。通过上述方式,能够按照设定好的采用频率对视频文件进行解码,由此得到相应的图像帧序列,进而便于进行后续的处理,从而提升方案的可行性和可操作性。

可选地,在上述图 3 对应的实施例的基础上,本申请实施例提供的视频分类的方法第二个可选实施例中,通过图像分类网络模型获取图像帧序列所对应的外观信息特征序列,可以包括:

通过图像分类网络模型获取第  $t$  帧图像所对应的外观信息特征,其中, $t$  大于或等于 1,且小于或等于  $T$  的整数;

当获取到  $T$  帧图像中每帧图像所对应的外观信息特征时,根据  $T$  个外观信息特征生成外观信息特征序列。

本实施例中,将介绍一种通过图像分类网络模型获取图像帧序列所对应的外观信息特征序列的方法。首先将图像帧序列中的每帧图像输入至图像分类网络模型,其中,每帧图像均包括外观信息,外观信息包含但不仅限于对象的形状、大小、颜色以及类型等可以直接通过图像获得的信息。通过图像分类网络模型输出每帧图像所对应的外观信息特征, $T$  帧图像对应  $T$  个外观信息特征, $T$  个外观信息特征构成外观信息特征序列。外观信息特征指的是利用卷积神经网络对图像进行特征提取后获得的特征。由于图像中包含了外观信息,这些外观信息用特征进行表示,所得的特征即为外观信息特征。

具体地,利用目前常见的视频特征提取方法(例如卷积神经网络或递归神经网络等方法)对待分类视频文件中的图像帧机械能外观信息特征表示。对于具有  $T$  帧图像的待分类视频文件而言,能够得到外观信息特征序列  $\{x_{video}^t\}_{t=1}^T$ 。本申请使用的视频特征提取方法可以是 Inception-V4 深度神经网络模型。Inception-V4 深度神经网络模型是卷积神经网络图片分类模型。本方案利用在图像网络(ImageNet)数据集上预训练好的 Inception-V4 模型参数,给定输入图像,获得对应图像的深度神经网络特征,该特征根据网络中的参数以及其中的卷积和非线性激活函数等最后的特征输出,其中,特征维度经过主成分分析算法处理之后为 1024 维。

主成分分析算法是特征降维的算法,比如说输入特征是 4096 维,通过主成分分析获得该特征模最大的 128 个方向,取该特征在这 128 个方向的投影则得到该特征降维后 128 维的特征。

可以理解的是,在实际应用中除了可以使用 Inception-V4 深度神经网络模型以外,还可以使用其他类型的深度神经网络模型,比如残差网络(Residual Networks, ResNets)、Inception-V3 深度神经网络模型或者双通道网络结构(Dual Path Network, DPN),此处仅为一个示意,并不应理解为对本申请的限定。

其次,本申请实施例中,提供了一种通过图像分类网络模型获取外观信息特征序列的方法,即视频分类装置通过图像分类网络模型获取第  $t$  帧图像所对应的外观信息特征,当获取到  $T$  帧图像中每帧图像所对应的外观信息特征时,根据视频分类装置  $T$  个外观信息特征生成外观信息特征序列。通过上述方式,可以直接利用图像分类网络模型输出视频文件的外观信

息特征序列，不需要以人工提取特征序列，一方面可以提升特征预测的准确度，另一方面节省了人工成本，还能够提升特征提取的效率。

可选地，在上述图 3 对应的实施例的基础上，本申请实施例提供的视频分类的方法第三个可选实施例中，通过运动预测网络模型获取外观信息特征序列所对应的运动信息特征序列，可以包括：

获取第  $i$  个步骤所对应的第  $t$  个外观信息特征，其中， $i$  为大于或等于 1 的整数， $t$  大于或等于 1，且小于或等于  $T$  的整数；

获取第  $(i-1)$  个步骤所对应的第  $(t+1)$  个运动信息特征以及第  $(t-1)$  个运动信息特征；

通过运动预测网络模型，根据第  $t$  个外观信息特征、第  $(t+1)$  个运动信息特征以及第  $(t-1)$  个运动信息特征，确定第  $i$  个步骤所对应的第  $t$  个运动信息特征；

当获取到  $T$  个外观信息特征中每个外观信息特征所对应的运动信息特征时，根据  $T$  个运动信息特征生成运动信息特征序列。

本实施例中，介绍了一种生成运动信息特征序列的方式，对于给定一系列属于某个视频

的  $T$  个  $D$  维外观信息特征序列  $\{x_{video}^t\}_{t=1}^T$  作为输入， $T$  为待分类视频文件的时间维度数目， $D$  维指单个特征的特征维度(长度)。比如说描述物体的位置可以用 3 维  $(x,y,z)$  的特征来描述。在这里外观信息特征是指通过卷积神经网络获得的特征，是 1024 维的。但对于不同的卷积神经网络，计算获得的维度可能不一样，可以通过设置的方式进行调整。

比如，对长度为 60 秒的待分类视频文件，以每秒一个特征的频率提取特征，则会提取得到  $T=60$  个  $D$  维特征，其时间维度按时间顺序排列。本申请提出运动预测网络模型对运动信

息特征序列  $\{s_{motion}^t\}_{t=1}^T$  进行预测，得到待分类视频文件的外观信息特征序列以及运动信息特征序列，进而进行视频分类。其中，运动信息具体可以包括物体运动的方向，运动的强度等运动相关的信息。

由于运动信息特征计算需要同时接收前一个时刻和后一个时刻的运动信息特征，假设当前时刻为  $t$ ，那么前一个时刻为  $(t-1)$  个时刻，后一个时刻为  $(t+1)$  个时刻，将第  $(t+1)$  个运动信息特征以及第  $(t-1)$  个运动信息特征作为输入，用于估计当前时刻的运动信息特征，这在实现的过程中并不能直接获得。因此本申请采取递归的方式进行估计。假设当前步骤为  $i$ ，那么上一个步骤表示为  $(i-1)$  个步骤，将第  $i$  个步骤的第  $t$  个外观信息特征、第  $(i-1)$  个步骤计算得到的第  $(t-1)$  个时刻的运动信息特征以及第  $(t-1)$  个时刻的运动信息特征共同输入至运动预测网络模型，由运动预测网络模型输出第  $i$  个步骤的第  $t$  个运动信息特征。类似地，获取  $T$  个外观信息特征中每个外观信息特征所对应的运动信息特征，由  $T$  个运动信息特征构成运动信息特征序列。

其次，本申请实施例中，提供了一种通过运动预测网络模型获取外观信息特征序列所对应的运动信息特征序列的方法，即视频分类装置先获取第  $i$  个步骤所对应的第  $t$  个外观信息特征，然后获取第  $(i-1)$  个步骤所对应的第  $(t+1)$  个运动信息特征以及第  $(t-1)$  个运动信息特征，再通过运动预测网络模型，根据第  $t$  个外观信息特征、第  $(t+1)$  个运动信息特征以及第  $(t-1)$  个运动信息特征，确定第  $i$  个步骤所对应的第  $t$  个运动信息特征，当获取到  $T$  个外观信息特征中每个外观信息特征所对应的运动信息特征时，根据  $T$  个运动信息特征生成运动信息特征序列。通过上述方式，通过运动预测网络模型能够直接获取到运动信息特征序列，进而自动对视频文件进行分类，从而完成对于视频内容的识别，由此大幅减少视频内容筛查人员的工作量。此外对于视频推荐方面的应用，可以根据视频类别进行用户个性化推荐，从而提升用户体验和丰富产品内容。

可选地，在上述图 3 对应的第三个实施例的基础上，本申请实施例提供的视频分类的方法第四个可选实施例中，根据第  $t$  个外观信息特征、第  $(t+1)$  个运动信息特征以及第  $(t-1)$  个运动信息特征，确定第  $i$  个步骤所对应的第  $t$  个运动信息特征，可以包括：

通过运动预测网络模型的更新门，根据第  $t$  个外观信息特征、第  $(t+1)$  个运动信息特征以及第  $(t-1)$  个运动信息特征，计算得到第  $i$  个步骤所对应的第  $(t+1)$  个更新门向量以及第  $(t-1)$  个更新门向量；

通过运动预测网络模型的输出门，根据第  $t$  个外观信息特征、第  $(t+1)$  个运动信息特征以

及第(t-1)个运动信息特征, 计算得到第i个步骤所对应的第(t+1)个输出门向量以及第(t-1)个输出门向量;

通过运动预测网络模型的隐藏层, 根据第t个外观信息特征、第(t+1)个运动信息特征、第(t-1)个运动信息特征、第(t+1)个更新门向量以及第(t-1)个更新门向量, 计算得到第i个步骤所对应的第t个隐藏层特征向量;

通过运动预测网络模型的最终输出层, 根据第t个隐藏层特征向量、第i个步骤所对应的第(t+1)个输出门向量以及第(t-1)个输出门向量, 计算得到第i个步骤所对应的第t个运动信息特征。

本实施例中, 对于T个D维输入外观信息特征序列  $\{x_{video}^t\}_{t=1}^T$ , 运动预测网络模型希望通过找到外观信息特征序列对运动信息特征序列的映射方法, 使得网络能够根据外观信息特征序列预测运动信息特征序列  $\{S_{motion}^t\}_{t=1}^T$ 。本方案中提供的运动预测网络模型  $f$  主要分为四个组成部分, 包括隐藏层 (hidden state)、更新门 (update gate)、输出门 (output gate) 以及最终输出层 (final output)

为了便于理解, 请参阅图5, 图5为本申请实施例中运动预测网络模型的一个结构单元示意图, 如图所示, 将第i个步骤对应的第t个外观信息特征  $x_t$  作为更新门的输入, 同时将第(i-1)个步骤对应的第(t+1)个运动信息特征  $S_{t+1}$  以及第(t-1)个运动信息特征  $S_{t-1}$  也作为更新门的输入, 由此输出第i个步骤对应的第(t-1)个更新门向量  $r_{t-1}$  以及第(t+1)个更新门向量  $r_{t+1}$ 。

将第i个步骤对应的第t个外观信息特征  $x_t$  作为输出门的输入, 同时将第(i-1)个步骤对应的第(t+1)个运动信息特征  $S_{t+1}$  以及第(t-1)个运动信息特征  $S_{t-1}$  也作为输出门的输入, 由此输出第i个步骤对应的第(t-1)个更新门向量  $z_{t-1}$  以及第(t+1)个输出门向量  $z_{t+1}$ 。

将第i个步骤对应的第t个外观信息特征  $x_t$  作为隐藏层的输入, 同时将第(i-1)个步骤对应的第(t+1)个运动信息特征  $S_{t+1}$  以及第(t-1)个运动信息特征  $S_{t-1}$  也作为隐藏层的输入, 将第(i-1)个步骤对应的第(t+1)个更新门向量  $z_{t+1}$  以及第(t-1)个更新门向量  $z_{t-1}$  也作为隐藏层的输入, 由此输出第i个步骤对应的第t个隐藏层特征向量  $h_t$ 。

将第i个步骤对应的第t个外观信息特征  $x_t$  作为最终输出层的输入, 同时将第i个步骤对应的第(t-1)个更新门向量  $z_{t-1}$  以及第(t+1)个输出门向量  $z_{t+1}$  也作为最终输出层的输入, 由此输出第i个步骤所对应的第t个运动信息特征  $S_t$ 。

应理解, 本申请采取递归的方式进行估计。将第i个步骤的第t个外观信息特征、第(i-1)个步骤计算得到的第(t-1)个时刻的运动信息特征以及第(t-1)个时刻的运动信息特征共同输入至运动预测网络模型, 由运动预测网络模型输出第i个步骤的第t个运动信息特征。为了便于理解, 请参阅图6, 图6为本申请实施例中运动预测网络模型的一个递归展开示意图, 如图所示, 在预测第i个步骤的运动信息特征  $S_t$  时, 可以采用如下运动预测网络模型:

$$S_t^i = f(x_t, S_{t+1}^{i-1}, S_{t-1}^{i-1});$$

其中,  $S_{t-1}^{i-1}$  表示第(i-1)个步骤所对应的第(t-1)个运动信息特征,  $S_{t+1}^{i-1}$  表示第(i-1)个步骤所对应的第(t+1)个运动信息特征。通过递归计算的方式, 预测得到运动信息特征序列  $\{S_{motion}^t\}_{t=1}^T$ , 在计算过程中运动信息特征的初始值设为0向量。

再次，本申请实施例中，提供了一种根据第  $t$  个外观信息特征、第  $(t+1)$  个运动信息特征以及第  $(t-1)$  个运动信息特征，确定第  $i$  个步骤所对应的第  $t$  个运动信息特征的方法。通过上述方式，在设计运动预测网络模型时主要分为四个部分，分别包括隐藏层、更新门、输出门以及最终的输出，由此，可以分别对各个部分进行计算，进而计算得到运动信息特征，从而保证方案的可行性和可操作性。

可选地，在上述图 3 对应的第四个实施例的基础上，本申请实施例提供的视频分类的方法第五个可选实施例中，通过运动预测网络模型的更新门，根据第  $t$  个外观信息特征、第  $(t+1)$  个运动信息特征以及第  $(t-1)$  个运动信息特征，计算得到第  $i$  个步骤所对应的第  $(t+1)$  个更新门向量以及第  $(t-1)$  个更新门向量，可以包括：

采用如下方式计算第  $(t+1)$  个更新门向量以及第  $(t-1)$  个更新门向量：

$$r_{t-1} = \sigma(W_r x_t + U_{r,t-1} s_{t-1}^{i-1});$$

$$r_{t+1} = \sigma(W_r x_t + U_{r,t+1} s_{t+1}^{i-1});$$

其中， $r_{t-1}$  表示第  $i$  个步骤所对应的第  $(t-1)$  个更新门向量， $t$  表示第  $t$  个时刻， $r_{t+1}$  表示第  $i$  个步骤所对应的第  $(t+1)$  个更新门向量， $\sigma(\cdot)$  表示线性整流函数， $W_r$  表示第一参数矩阵， $U_{r,t-1}$  表示第  $(t-1)$  个更新门向量所对应的第二参数矩阵， $U_{r,t+1}$  表示第  $(t+1)$  个更新门向量所对应的第三参数矩阵， $x_t$  表示第  $t$  个外观信息特征， $s_{t-1}^{i-1}$  表示第  $(i-1)$  个步骤所对应的第  $(t-1)$  个运动信息特征， $s_{t+1}^{i-1}$  表示第  $(i-1)$  个步骤所对应的第  $(t+1)$  个运动信息特征；

通过运动预测网络模型的输出门，根据第  $t$  个外观信息特征、第  $(t+1)$  个运动信息特征以及第  $(t-1)$  个运动信息特征，计算得到第  $i$  个步骤所对应的第  $(t+1)$  个输出门向量以及第  $(t-1)$  个输出门向量，可以包括：

采用如下方式计算第  $(t+1)$  个输出门向量以及第  $(t-1)$  个输出门向量：

$$[z_{t+1}, z, z_{t-1}] = \text{soft max}([\tilde{z}_{t+1}, 1, \tilde{z}_{t-1}]);$$

$$\tilde{z}_{t-1} = \sigma(W_z x_t + U_{z,t-1} s_{t-1}^{i-1});$$

$$\tilde{z}_{t+1} = \sigma(W_z x_t + U_{z,t+1} s_{t+1}^{i-1});$$

其中， $z_{t+1}$  表示第  $i$  个步骤所对应的第  $(t+1)$  个输出门向量， $z_{t-1}$  表示第  $i$  个步骤所对应的第  $(t-1)$  个输出门向量， $z$  表示预设输出门向量， $\tilde{z}_{t-1}$  表示第  $i$  个步骤所对应的第  $(t-1)$  个待处理输出门向量， $\tilde{z}_{t+1}$  表示第  $i$  个步骤所对应的第  $(t+1)$  个待处理输出门向量， $\text{soft max}(\cdot)$  表示归一化处理函数， $W_z$  表示第四参数矩阵， $U_{z,t-1}$  表示第  $(t-1)$  个输出门向量所对应的第五参数矩阵， $U_{z,t+1}$  表示第  $(t+1)$  个输出门向量所对应的第六参数矩阵；

通过运动预测网络模型的隐藏层，根据第  $t$  个外观信息特征、第  $(t+1)$  个运动信息特征、第  $(t-1)$  个运动信息特征、第  $(t+1)$  个更新门向量以及第  $(t-1)$  个更新门向量，计算得到第  $i$  个步骤所对应的第  $t$  个隐藏层特征向量，可以包括：

采用如下方式计算第  $t$  个隐藏层特征向量：

$$h_t = \phi\left(W_h x_t + U_h \left[s_{t+1}^{i-1} \circ r_{t+1}, s_{t-1}^{i-1} \circ r_{t-1}\right]^T\right);$$

其中， $h_t$  表示第  $i$  个步骤所对应的第  $t$  个隐藏层特征向量， $\phi(\cdot)$  表示 sigmoid 函数， $W_h$  表示第七参数矩阵， $U_h$  表示第八参数矩阵， $[\cdot]^T$  表示矩阵转置， $\circ$  表示点乘；

通过运动预测网络模型的最终输出层，根据第  $t$  个隐藏层特征向量、第  $i$  个步骤所对应的

第(t+1)个输出门向量以及第(t-1)个输出门向量, 计算得到第i个步骤所对应的第t个运动信息特征, 可以包括:

采用如下方式计算第t个运动信息特征:

$$s_t^i = z \circ h_t + z_{t-1} \circ s_{t-1}^{i-1} + z_{t+1} \circ s_{t+1}^{i-1};$$

其中,  $s_t^i$  表示第i个步骤所对应的第t个运动信息特征。

本实施例中, 将介绍计算第i个步骤所对应的第t个运动信息特征  $s_t^i$  的具体方式。即可以采用如下方式:

$$h_t = \phi \left( W_h x_t + U_h \left[ s_{t+1}^{i-1} \circ r_{t+1}, s_{t-1}^{i-1} \circ r_{t-1} \right]^T \right);$$

$$r_{t-1} = \sigma \left( W_r x_t + U_{r,t-1} s_{t-1}^{i-1} \right);$$

$$r_{t+1} = \sigma \left( W_r x_t + U_{r,t+1} s_{t+1}^{i-1} \right);$$

$$\tilde{z}_{t-1} = \sigma \left( W_z x_t + U_{z,t-1} s_{t-1}^{i-1} \right);$$

$$\tilde{z}_{t+1} = \sigma \left( W_z x_t + U_{z,t+1} s_{t+1}^{i-1} \right);$$

$$[z_{t+1}, z, z_{t-1}] = \text{soft max} \left( [\tilde{z}_{t+1}, 1, \tilde{z}_{t-1}] \right);$$

$$s_t^i = z \circ h_t + z_{t-1} \circ s_{t-1}^{i-1} + z_{t+1} \circ s_{t+1}^{i-1};$$

其中, 假设  $x_t$  的维度为  $D_x$ ,  $h_t$  的维度为  $D_h$ ,  $s_t$  的维度为  $D_s$ ,  $r_t$  的维度为  $D_r$ , 其中, 维度与时间t无关,  $D_s = D_r$ , 那么第七参数矩阵  $W_h$  的维度为  $D_h \times D_x$ , 第八参数矩阵  $U_h$  的维度为  $D_h \times D_s$ 。第四参数矩阵  $W_z$  的维度为  $D_z \times D_x$ , 第五参数矩阵  $U_{z,t-1}$  与第六参数矩阵  $U_{z,t+1}$  的维度为  $D_z \times D_s$ , 第二参数矩阵  $U_{r,t-1}$  与第三参数矩阵  $U_{r,t+1}$  的维度为  $D_r \times D_s$ 。

在  $s_t^i = z \circ h_t + z_{t-1} \circ s_{t-1}^{i-1} + z_{t+1} \circ s_{t+1}^{i-1}$  中,  $\circ$  表示点乘符号, 即向量的对应位置数值相乘, 比如  $[1,2,3] \circ [2,3,4] = [2,6,12]$ 。将  $z_{t+1}$ 、 $z$  和  $z_{t-1}$  分别对  $s_{t-1}^{i-1}$ 、 $h_t$  和  $s_{t+1}^{i-1}$  进行点乘, 用于控制不同上述向量对于输出  $s_t^i$  的贡献程度。

进一步地, 本申请实施例中, 提供了计算特征向量的具体方式, 包括计算第i个步骤所对应的第(t+1)个更新门向量以及第(t-1)个更新门向量的方式, 计算第i个步骤所对应的第(t+1)个输出门向量以及第(t-1)个输出门向量的方式, 计算第i个步骤所对应的第t个隐藏层特征向量的方式, 以及计算得到第i个步骤所对应的第t个运动信息特征的方式。通过上述方式, 利用递归计算的模式能够克服在计算运动信息特征时, 不能同时接收前一时刻和后一时刻的运动信息特征的问题, 因此, 采用递归的方式进行运动信息特征的估计, 可以保证方案的可行性和可操作性。

可选地, 在上述图3以及图3对应的第一个至第五个实施例中任一项的基础上, 本申请实施例提供的视频分类的方法第六个可选实施例中, 根据外观信息特征序列以及运动信息特征序列确定待分类视频文件的视频分类结果, 可以包括:

获取外观信息特征序列以及运动信息特征序列所对应的视频类别向量, 其中, 视频类别向量包括C个视频类别元素, C为大于1的整数;

从视频类别向量中确定至少一个目标视频类别元素;

根据至少一个目标视频类别元素生成待分类视频文件的视频分类结果。

在一个可能实施例中, 该视频类别向量的获取过程可以通过视频分类模型实现, 本申请实施例对具体是否使用视频分类模型进行上述获取过程不作限定。

本实施例中, 将介绍一种确定视频分类结果的方式, 首先视频分类装置需要通过视频分

类模型获取外观信息特征序列以及运动信息特征序列所对应的视频类别向量，这里的视频类别向量包括 C 个视频类别元素，每个视频类别元素表示一种视频类别，比如“娱乐”、“体育”、“新闻”、“生活”以及“科技”等。然后从视频类别向量中确定至少一个目标视频类别元素，通常选择值较大的元素作为目标视频类别元素。最后视频根据至少一个目标视频类别元素生成待分类视频文件的视频分类结果。

具体地，下面将结合示例进行介绍。在获取到外观信息特征序列  $\{x_{video}^t\}_{t=1}^T$  和预测得到的运动信息特征序列  $\{s_{motion}^t\}_{t=1}^T$  之后，利用视频分类模型对待分类视频文件进行分类。其中，视频分类模型包含但不限于支持向量机 (support vector machine, SVM) 分类器、K 最近邻 (k-nearest neighbor, KNN) 分类器、神经网络分类器、多层非线性网络分类模型以及多层非线性网络混合专家分类模型。

通过视频分类模型输出以概率表示的 C 维视频类别向量，其中，C 为大于 1 的整数，C 为预先设定的视频类别元素的数量。该视频类别向量中每个元素位置的数值代表待分类视频文件为对应类别的概率，最后将该视频类别向量转换为类别，从而获得视频分类结果。以三类视频类别元素（分别依次为“娱乐”、“体育”和“生活”）为例，假设视频类别向量表示为 [0.1, 0.9, 0.7]，即代表视频分类模型确定待分类视频文件属于“娱乐”的概率为 0.1，属于“体育”的概率为 0.9，属于“生活”的概率为 0.7。可见，属于“体育”类别和属于“生活”类别的概率比较高，即确定目标视频类别元素为“体育”以及“生活”，此时，生成待分类视频文件的视频分类结果为“体育+生活”。如果只输出一个目标视频类别元素，则可以输出“体育”，即生成待分类视频文件的视频分类结果为“体育”。

需要说明的是，每个类别可以独立地出现，因此不要求视频类别向量的数值之和为 1。但也不排除用于单个类别分类的情况，即视频类别向量的数值之和为 1 的情况。

更进一步地，本申请实施例中，提供了一种确定待分类视频文件视频分类结果的方法，即视频分类装置先通过视频分类模型获取外观信息特征序列以及运动信息特征序列所对应的视频类别向量，然后从视频类别向量中确定至少一个目标视频类别元素，最后根据至少一个目标视频类别元素生成待分类视频文件的视频分类结果。通过上述方式，能够根据视频分类模型输出的视频类别向量，确定该视频文件属于某一个类别的可能性，由此，保证视频文件分类的可靠性和准确度，对于有些视频文件可能不仅仅属于一个类型，本方案还可以针对视频分类模型输出的视频类别向量，为视频文件分配多个类型，由此提升方案的多样性。

结合上述介绍，下面将对本申请中模型训练的方法进行介绍，请参阅图 7，本申请实施例中模型训练的方法一个实施例包括：

201、获取待训练视频文件所对应的光流图序列，其中，光流图序列中包括多个光流图；

本实施例中，对于运动预测网络模型的训练，首先模型训练装置可以利用全变分 L1 范数 (Total Variational L1 Norm, TVL1) 光流估计方法对待训练视频文件进行光流图的计算，TVL1 方法估计得到的光流图效果比较好，从而得到光流图序列，光流图序列中包括多个光流图。

可以理解的是，在实际应用中，还可以使用其他的光流计算方法，例如 Brox 方法。光流法实际是通过检测图像像素点的强度随时间的变化进而推断出物体移动速度及方向的方法。而光流图则是光流法计算的数值的图像显示。光流图一般分成两张单通道的灰度图像，分别代表垂直方向和水平方向的像素点运动强度。强度以图像亮度的形式表示。

正是因为这是模型训练的过程，所以可以采用光流估计方法进行光流图的提取以及光流特征的计算，一般来说，模型训练装置可以部署在性能较高和存储资源较大的计算机机器或集群中。

202、获取待训练视频文件所对应的待训练外观信息特征序列，其中，待训练外观信息特征序列包括多个待训练外观信息特征；

本实施例中，模型训练装置利用神经网络可以提取待训练视频文件所对应的待训练外观信息特征序列，其中，待训练外观信息特征序列包括多个待训练外观信息特征，待训练外观

信息特征序列可记做  $\{x_{video}^t\}_{t=1}^T$ 。

203、根据待训练视频文件所对应的光流图序列，获取真实运动信息特征序列，其中，真实运动信息特征序列包括 T 个真实运动信息特征，T 为大于 1 的整数；

本实施例中，模型训练装置根据待训练视频文件所对应的光流图序列，利用卷积神经网络进行真实运动信息特征的提取，T 个真实运动信息特征构成真实运动信息特征序列，T 为

大于 1 的整数。真实运动信息特征序列可以记为  $\{\tilde{S}_{motion}^t\}_{t=1}^T$ 。

计算外观信息特征和真实运动信息特征需要每秒 110 吉字节 (Gigabyte, G) 的浮点运算次数，而结合运动预测网络模型计算外观信息特征和运动信息特征仅需要每秒 55G 浮点运算次数。浮点运算次数为机器进行浮点数运算的次数，反映了计算资源的消耗。而存储资源则主要是光流图的存储，由于光流图是两张单通道的图片，因此近似是 RGB 图片的 2/3 大小。以 1GB 的视频为例，解压获得的 JPG 图片约 10GB，因此光流图约占 6.6GB 空间。

204、通过待训练运动预测网络模型获取待训练外观信息特征序列所对应的预测运动信息特征序列，其中，预测运动信息特征序列包括 T 个预测运动信息特征；

本实施例中，模型训练装置将待训练外观信息特征序列  $\{x_{video}^t\}_{t=1}^T$  输入至待训练运动预测网络模型，由待训练运动预测网络模型输出相应的预测运动信息特征序列，预测运动信息特征序列可以记做  $\{S_{motion}^t\}_{t=1}^T$ 。即预测运动信息特征序列  $\{S_{motion}^t\}_{t=1}^T$  中包括 T 个预测运动信息特征。

205、根据真实运动信息特征序列与预测运动信息特征序列，确定模型参数；

本实施例中，模型训练装置采用梯度下降算法对待训练运动预测网络模型中的模型参数进行更新，使用损失函数对真实运动信息特征序列与预测运动信息特征序列进行计算，得到模型参数。

206、采用模型参数对待训练运动预测网络模型进行训练，得到运动预测网络模型。

本实施例中，模型训练装置根据模型参数对待训练运动预测网络模型进行训练，得到运动预测网络模型。使得运动预测网络模型能够根据输入的外观信息特征序列  $\{x_{video}^t\}_{t=1}^T$  预测出与真实运动信息特征序列  $\{\tilde{S}_{motion}^t\}_{t=1}^T$  相近的运动信息特征序列  $\{S_{motion}^t\}_{t=1}^T$ 。而不需要对光流图进行计算，也不需要利用卷积神经网络对光流图进行真实运动信息特征提取，从而大大减少计算量和存储资源。

本申请实施例中，提供了一种模型训练的方法，即模型训练装置首先获取待训练视频文件所对应的光流图序列，然后获取待训练视频文件所对应的待训练外观信息特征序列，待训练外观信息特征序列包括多个待训练外观信息特征，模型训练装置根据待训练视频文件所对应的光流图序列，获取真实运动信息特征序列，再通过待训练运动预测网络模型获取待训练外观信息特征序列所对应的预测运动信息特征序列，其中，预测运动信息特征序列包括 T 个预测运动信息特征，根据真实运动信息特征序列与预测运动信息特征序列，确定模型参数，最后模型训练装置可以采用模型参数对待训练运动预测网络模型进行训练，得到运动预测网络模型。通过上述方式，在模型训练的过程中使用视频文件的光流图作为训练对象，从而使得训练得到的运动预测网络模型具有隐式获取视频文件中光流图的功能，在实际应用中，通过运动预测网络模型就能够直接输出运动信息特征，不需要再特意计算出光流图，从而大幅度地减少计算量和穿存储资源。

可选地，在上述图 7 对应的实施例的基础上，本申请实施例提供的模型训练的方法第一个可选实施例中，根据真实运动信息特征序列与预测运动信息特征序列，确定模型参数，可以包括：

采用损失函数对真实运动信息特征序列与预测运动信息特征序列进行计算，得到目标损失结果；

根据目标损失结果确定模型参数。

本实施例中，将介绍如何根据真实运动信息特征序列与预测运动信息特征序列确定模型参数的方法。即采用损失函数对真实值和预测值进行计算，得到目标损失结果，目标损失结果即表示真实值和预测值之间的距离，因此，损失函数是一个非负实值函数。目标损失结果越小，表示待训练运动预测网络模型的鲁棒性越好。应理解，本申请并不限定所采用的损失函数类型，下面将介绍几种本实施例中可以使用的损失函数。

第一种为均方误差 (mean-square error, MSE) 损失函数, MSE 损失函数表示真实运动信息特征与预测运动信息特征之间差值的平方和, 随着真实运动信息特征与预测运动信息特征之间的误差增大, 使得 MSE 损失结果的增加更加迅猛。

第二种为平均绝对误差 (mean absolute error, MAE) 损失函数, MAE 损失函数表示真实运动信息特征与预测运动信息特征之差绝对值的和, 即表示真实运动信息特征与预测运动信息特征的平均误差幅度, 而不需要考虑误差的方向, 范围是 0 到正无穷。

第三种为 Huber 损失函数, Huber 损失函数相比于平方损失来说对于异常值不敏感, 但它同样保持了可微的特性。它基于绝对误差但在误差很小的时候变成了平方误差, 可以使用超参数来调节这一误差的阈值。当超参数趋向于 0 时它就退化成了 MAE 损失函数, 而当超参数趋向于无穷时则退化为了 MSE 损失函数。

第四种为对数余弦变换 (Log-Cosh) 损失函数, Log-Cosh 损失函数较为平滑, 可以利用双曲余弦来计算预测误差。

第五种为分位数损失函数, 在大多数真实世界的预测问题中, 常常希望看到预测结果的不确定性。通过预测出一个取值区间而不是一个个具体的取值点对于具体业务流程中的决策至关重要。

其次, 本申请实施例, 提供了一种根据真实运动信息特征序列与预测运动信息特征序列确定模型参数的方法, 即模型训练装置先采用损失函数对真实运动信息特征序列与预测运动信息特征序列进行计算, 得到目标损失结果, 然后根据目标损失结果确定模型参数。通过上述方式, 能够采用合适的损失函数计算得到损失结果, 基于损失结果的最小值计算得到模型参数, 该模型参数能够有效地训练得到机器学习模型, 从而提升模型训练的可靠性和准确度。

可选地, 在上述图 7 对应的第一个实施例的基础上, 本申请实施例提供的模型训练的方法第二个可选实施例中, 采用损失函数对真实运动信息特征序列与预测运动信息特征序列进行计算, 得到目标损失结果, 可以包括:

采用如下方式计算目标损失结果:

$$L_{MSE} = \frac{1}{T} \sum_t \left\| \tilde{S}^t - S^t \right\|_2^2;$$

其中,  $L_{MSE}$  表示目标损失结果,  $T$  表示真实运动信息特征序列中的真实运动信息特征个数,  $\tilde{S}^t$  表示真实运动信息特征序列中第  $t$  个真实运动信息特征,  $S^t$  表示预测运动信息特征序列中第  $t$  个预测运动信息特征,  $t$  为大于或等于 1, 且小于或等于  $T$  的整数。

本实施例中, 介绍了一种利用 MSE 损失函数计算目标损失结果的具体方式, 在模型训练装置获取到真实运动信息特征序列以及预测运动信息特征序列之后, 可以利用 MSE 损失函数作为本实施例中所使用的损失函数。假设理想的运动预测网络模型是一个函数, 那么真实运动信息特征和预测运动信息特征之间的距离可以通过 MSE 损失函数来近似表示, 当距离无限缩小时, 表示模型也就越接近理想模型。

具体地, 可以采用如下方式计算目标损失结果:

$$L_{MSE} = \frac{1}{T} \sum_t \left\| \tilde{S}^t - S^t \right\|_2^2;$$

在上述式子中,  $\tilde{S}^t$  表示真实运动信息特征序列在一个批处理 (batch) 中的第  $t$  个真实运动信息特征, 而  $S^t$  表示预测运动信息特征序列中第  $t$  个预测运动信息特征, 也就是待训练运动预测网络模型输出的预测结果, MSE 损失函数是一个求 batch 的平均误差的函数。

再次, 本申请实施例中, 提供了一种计算目标损失结果的具体方式, 即利用 MSE 损失函数对真实运动信息特征序列与预测运动信息特征序列进行计算, 得到目标损失结果。通过上述方式, 采用 MSE 损失函数进行计算能够降低计算的复杂度, 即便使用固定的学习率也可以有效收敛。MSE 损失函数的梯度随损失增大而增大, 而损失趋于 0 时则会减小, 这使得在训练结束时, 使用 MSE 损失函数的结果会更精确。

应理解, 本申请在采取线性分类网络进行实验并获得较好的效果, 在大规模视频数据集

中,根据 nception-V4 方法提取的外观信息特征,通过结合运动信息特征,利用线性分类器方法对视频特征进行分类,获得比仅用外观信息特征 0.23% 全局平均精确率 (Global Average Precision, GAP) @20 的分类性能和约 0.2% 的首类命中性能。

其中, GAP@20 为多类别视频分类性能指标,首类命中指分类置信度最高的类别命中视频真实类标,结果得分最高为视频真实类标的情况的统计结果。

因此利用本申请提出的运动预测网络模型对视频运动信息特征进行预测,视频分类模型能够有效地对输入视频进行高性能的分类,获得更加准确的类别输出,因此,能够广泛地部署到各个视频应用服务中,并对如视频推荐,视频过滤等业务。

下面对本申请中的视频分类装置进行详细描述,请参阅图 8,图 8 为本申请实施例中视频分类装置一个实施例示意图,视频分类装置 30 包括:

获取模块 301,用于获取待分类视频文件所对应的图像帧序列,其中,该图像帧序列中包括 T 帧图像,该 T 为大于 1 的整数;

该获取模块 301,还用于获取该图像帧序列所对应的外观信息特征序列,其中,该外观信息特征序列中包括 T 个外观信息特征,该外观信息特征与该图像帧具有对应关系;

该获取模块 301,还用于获取该外观信息特征序列所对应的运动信息特征序列,其中,该运动信息特征序列包括 T 个运动信息特征,该运动信息特征与该外观信息特征具有对应关系;

确定模块 302,用于根据该获取模块 301 获取的该外观信息特征序列以及该运动信息特征序列确定该待分类视频文件的视频分类结果。

本实施例中,获取模块 301 获取待分类视频文件所对应的图像帧序列,其中,该图像帧序列中包括 T 帧图像,该 T 为大于 1 的整数,该获取模块 301 获取该图像帧序列所对应的外观信息特征序列,其中,该外观信息特征序列中包括 T 个外观信息特征,该外观信息特征与该图像帧具有对应关系,该获取模块 301 获取该外观信息特征序列所对应的运动信息特征序列,其中,该运动信息特征序列包括 T 个运动信息特征,该运动信息特征与该外观信息特征具有对应关系,确定模块 302 根据该获取模块 301 获取的该外观信息特征序列以及该运动信息特征序列确定该待分类视频文件的视频分类结果。

可选地,在上述图 8 所对应的实施例的基础上,本申请实施例提供的视频分类装置 30 的另一实施例中,

该获取模块 301,具体用于通过图像分类网络模型获取该图像帧序列所对应的外观信息特征序列;

该获取模块 301,具体用于通过运动预测网络模型获取该外观信息特征序列所对应的运动信息特征序列,该运动预测网络模型用于预测该外观信息特征所对应的该运动信息特征。

本申请实施例中,提供了一种视频分类装置,首先该视频分类装置获取待分类视频文件所对应的图像帧序列,其中,图像帧序列中包括 T 帧图像,T 为大于 1 的整数,然后通过图像分类网络模型获取图像帧序列所对应的外观信息特征序列,其中,外观信息特征序列中包括 T 个外观信息特征,外观信息特征与图像帧具有对应关系,再通过运动预测网络模型获取外观信息特征序列所对应的运动信息特征序列,其中,运动信息特征序列包括 T 个运动信息特征,运动信息特征与外观信息特征具有对应关系,运动预测网络模型用于预测外观信息特征所对应的运动信息特征,最后可以根据外观信息特征序列以及运动信息特征序列确定待分类视频文件的视频分类结果。通过上述方式,能够实现对视频的自动分类,无需根据视频的外观信息特征计算光流图像,以利用光流图像计算运动信息特征。而是直接通过运动预测网络模型即可获取外观信息特征对应的运动信息特征,由此节省了计算时间和资源,从而提升了数据处理的效率。

可选地,在上述图 8 所对应的实施例的基础上,本申请实施例提供的视频分类装置 30 的另一实施例中,

该获取模块 301,具体用于获取该待分类视频文件;

对该待分类视频文件进行解码处理,得到该图像帧序列,其中,每个帧图像具有相同的采样频率。

其次,本申请实施例中,介绍了一种视频分类装置获取待分类视频文件所对应的图像帧序列的方式,即先获取待分类视频文件,然后对待分类视频文件进行解码处理,得到图像帧序列,其中,每个帧图像具有相同的采样频率。通过上述方式,能够按照设定好的采用频率

对视频文件进行解码，由此得到相应的图像帧序列，进而便于进行后续的处理，从而提升方案的可行性和可操作性。

可选地，在上述图 8 所对应的实施例的基础上，本申请实施例提供的视频分类装置 30 的另一实施例中，

该获取模块 301，具体用于通过该图像分类网络模型获取第  $t$  帧图像所对应的外观信息特征，其中，该  $t$  大于或等于 1，且小于或等于该  $T$  的整数；

当获取到该  $T$  帧图像中每帧图像所对应的外观信息特征时，根据该  $T$  个外观信息特征生成该外观信息特征序列。

其次，本申请实施例中，提供了一种通过图像分类网络模型获取外观信息特征序列的方法，即视频分类装置先通过图像分类网络模型获取第  $t$  帧图像所对应的外观信息特征，当获取到  $T$  帧图像中每帧图像所对应的外观信息特征时，根据视频分类装置  $T$  个外观信息特征生成外观信息特征序列。通过上述方式，可以直接利用图像分类网络模型输出视频文件的外观信息特征序列，不需要以人工提取特征序列，一方面可以提升特征预测的准确度，另一方面节省了人工成本，还能够提升特征提取的效率。

可选地，在上述图 8 所对应的实施例的基础上，本申请实施例提供的视频分类装置 30 的另一实施例中，

该获取模块 301，具体用于获取第  $i$  个步骤所对应的第  $t$  个外观信息特征，其中，该  $i$  为大于或等于 1 的整数，该  $t$  大于或等于 1，且小于或等于该  $T$  的整数；

获取第  $(i-1)$  个步骤所对应的第  $(t+1)$  个运动信息特征以及第  $(t-1)$  个运动信息特征；

通过该运动预测网络模型，根据该第  $t$  个外观信息特征、该第  $(t+1)$  个运动信息特征以及第  $(t-1)$  个运动信息特征，确定该第  $i$  个步骤所对应的第  $t$  个运动信息特征；

当获取到该  $T$  个外观信息特征中每个外观信息特征所对应的运动信息特征时，根据该  $T$  个运动信息特征生成该运动信息特征序列。

其次，本申请实施例中，提供了一种通过运动预测网络模型获取外观信息特征序列所对应的运动信息特征序列的方法，即视频分类装置先获取第  $i$  个步骤所对应的第  $t$  个外观信息特征，然后获取第  $(i-1)$  个步骤所对应的第  $(t+1)$  个运动信息特征以及第  $(t-1)$  个运动信息特征，再通过运动预测网络模型，根据第  $t$  个外观信息特征、第  $(t+1)$  个运动信息特征以及第  $(t-1)$  个运动信息特征，确定第  $i$  个步骤所对应的第  $t$  个运动信息特征，当获取到  $T$  个外观信息特征中每个外观信息特征所对应的运动信息特征时，根据  $T$  个运动信息特征生成运动信息特征序列。通过上述方式，通过运动预测网络模型能够直接获取到运动信息特征序列，进而自动对视频文件进行分类，从而完成对于视频内容的识别，由此大幅减少视频内容筛查人员的工作量。此外对于视频推荐方面的应用，可以根据视频类别进行用户个性化推荐，从而提升用户体验和丰富产品内容。

可选地，在上述图 8 所对应的实施例的基础上，本申请实施例提供的视频分类装置 30 的另一实施例中，

该获取模块 301，具体用于通过该运动预测网络模型的更新门，根据该第  $t$  个外观信息特征、该第  $(t+1)$  个运动信息特征以及第  $(t-1)$  个运动信息特征，计算得到该第  $i$  个步骤所对应的第  $(t+1)$  个更新门向量以及第  $(t-1)$  个更新门向量；

通过该运动预测网络模型的输出层，根据该第  $t$  个外观信息特征、该第  $(t+1)$  个运动信息特征以及第  $(t-1)$  个运动信息特征，计算得到该第  $i$  个步骤所对应的第  $(t+1)$  个输出层向量以及第  $(t-1)$  个输出层向量；

通过该运动预测网络模型的隐藏层，根据该第  $t$  个外观信息特征、该第  $(t+1)$  个运动信息特征、第  $(t-1)$  个运动信息特征、该第  $(t+1)$  个更新门向量以及该第  $(t-1)$  个更新门向量，计算得到该第  $i$  个步骤所对应的第  $t$  个隐藏层特征向量；

通过该运动预测网络模型的最终输出层，根据该第  $t$  个隐藏层特征向量、该第  $i$  个步骤所对应的第  $(t+1)$  个输出层向量以及第  $(t-1)$  个输出层向量，计算得到该第  $i$  个步骤所对应的第  $t$  个运动信息特征。

再次，本申请实施例中，提供了一种根据第  $t$  个外观信息特征、第  $(t+1)$  个运动信息特征以及第  $(t-1)$  个运动信息特征，确定第  $i$  个步骤所对应的第  $t$  个运动信息特征的方法。通过上述方式，在设计运动预测网络模型时主要分为四个部分，分别包括隐藏层、更新门、输出层以及最终的输出，由此，可以分别对各个部分进行计算，进而计算得到运动信息特征，从而保

证方案的可行性和可操作性。

可选地,在上述图8所对应的实施例的基础上,本申请实施例提供的视频分类装置30的另一实施例中,

该获取模块301,具体用于采用如下方式计算该第(t+1)个更新门向量以及该第(t-1)个更新门向量:

$$r_{t-1} = \sigma(W_r x_t + U_{r,t-1} s_{t-1}^{i-1});$$

$$r_{t+1} = \sigma(W_r x_t + U_{r,t+1} s_{t+1}^{i-1});$$

其中,该 $r_{t-1}$ 表示该第i个步骤所对应的该第(t-1)个更新门向量,该t表示第t个时刻,该 $r_{t+1}$ 表示该第i个步骤所对应的该第(t+1)个更新门向量,该 $\sigma(\cdot)$ 表示线性整流函数,该 $W_r$ 表示第一参数矩阵,该 $U_{r,t-1}$ 表示该第(t-1)个更新门向量所对应的第二参数矩阵,该 $U_{r,t+1}$ 表示该第(t+1)个更新门向量所对应的第三参数矩阵,该 $x_t$ 表示该第t个外观信息特征,该 $s_{t-1}^{i-1}$ 表示该第(i-1)个步骤所对应的该第(t-1)个运动信息特征,该 $s_{t+1}^{i-1}$ 表示该第(i-1)个步骤所对应的该第(t+1)个运动信息特征;

采用如下方式计算该第(t+1)个输出门向量以及该第(t-1)个输出门向量:

$$[z_{t+1}, z, z_{t-1}] = \text{soft max}([\tilde{z}_{t+1}, 1, \tilde{z}_{t-1}]);$$

$$\tilde{z}_{t-1} = \sigma(W_z x_t + U_{z,t-1} s_{t-1}^{i-1});$$

$$\tilde{z}_{t+1} = \sigma(W_z x_t + U_{z,t+1} s_{t+1}^{i-1});$$

其中,该 $z_{t+1}$ 表示该第i个步骤所对应的该第(t+1)个输出门向量,该 $z_{t-1}$ 表示该第i个步骤所对应的该第(t-1)个输出门向量,该z表示预设输出门向量,该 $\tilde{z}_{t-1}$ 表示该第i个步骤所对应的第(t-1)个待处理输出门向量,该 $\tilde{z}_{t+1}$ 表示该第i个步骤所对应的第(t+1)个待处理输出门向量,该 $\text{soft max}(\cdot)$ 表示归一化处理函数,该 $W_z$ 表示第四参数矩阵,该 $U_{z,t-1}$ 表示该第(t-1)个输出门向量所对应的第五参数矩阵,该 $U_{z,t+1}$ 表示该第(t+1)个输出门向量所对应的第六参数矩阵;

采用如下方式计算该第t个隐藏层特征向量:

$$h_t = \phi\left(W_h x_t + U_h \left[s_{t+1}^{i-1} \circ r_{t+1}, s_{t-1}^{i-1} \circ r_{t-1}\right]^T\right);$$

其中,该 $h_t$ 表示该第i个步骤所对应的该第t个隐藏层特征向量,该 $\phi(\cdot)$ 表示sigmoid函数,该 $W_h$ 表示第七参数矩阵,该 $U_h$ 表示第八参数矩阵,该 $[\cdot]^T$ 表示矩阵转置,该 $\circ$ 表示点乘;

采用如下方式计算该第t个运动信息特征:

$$s_t^i = z \circ h_t + z_{t-1} \circ s_{t-1}^{i-1} + z_{t+1} \circ s_{t+1}^{i-1};$$

其中,该 $s_t^i$ 表示该第i个步骤所对应的该第t个运动信息特征。

进一步地,本申请实施例中,提供了计算特征向量的具体方式,包括计算第i个步骤所对应的第(t+1)个更新门向量以及第(t-1)个更新门向量的方式,计算第i个步骤所对应的第(t+1)个输出门向量以及第(t-1)个输出门向量的方式,计算第i个步骤所对应的第t个隐藏层特征向量的方式,以及计算得到第i个步骤所对应的第t个运动信息特征的方式。通过上述方式,利用递归计算的模式能够克服在计算运动信息特征时,不能同时接收前一时刻和后一时刻的运动信息特征的问题,因此,采用递归的方式进行运动信息特征的估计,可以保证方案的可行

性和可操作性。

可选地，在上述图 8 所对应的实施例的基础上，本申请实施例提供的视频分类装置 30 的另一实施例中，

该确定模块 302，具体用于获取该外观信息特征序列以及该运动信息特征序列所对应的视频类别向量，其中，该视频类别向量包括 C 个视频类别元素，该 C 为大于 1 的整数；

从该视频类别向量中确定至少一个目标视频类别元素；

根据该至少一个目标视频类别元素生成该待分类视频文件的该视频分类结果。

更进一步地，本申请实施例中，提供了一种确定待分类视频文件视频分类结果的方法，即视频分类装置先通过视频分类模型获取外观信息特征序列以及运动信息特征序列所对应的视频类别向量，然后从视频类别向量中确定至少一个目标视频类别元素，最后根据至少一个目标视频类别元素生成待分类视频文件的视频分类结果。通过上述方式，能够根据视频分类模型输出的视频类别向量，确定该视频文件属于某一个类别的可能性，由此，保证视频文件分类的可靠性和准确度，对于有些视频文件可能不仅仅属于一个类型，本方案还可以针对视频分类模型输出的视频类别向量，为视频文件分配多个类型，由此提升方案的多样性。

下面对本申请中的模型训练装置进行详细描述，请参阅图 9，图 9 为本申请实施例中模型训练装置一个实施例示意图，模型训练装置 40 包括：

获取模块 401，用于获取待训练视频文件所对应的光流图序列，其中，该光流图序列中包括多个光流图；

该获取模块 401，还用于获取该待训练视频文件所对应的待训练外观信息特征序列，其中，该待训练外观信息特征序列包括多个待训练外观信息特征；

该获取模块 401，还用于根据该待训练视频文件所对应的该光流图序列，获取真实运动信息特征序列，其中，该真实运动信息特征序列包括 T 个真实运动信息特征，该 T 为大于 1 的整数；

该获取模块 401，还用于通过待训练运动预测网络模型获取该待训练外观信息特征序列所对应的预测运动信息特征序列，其中，该预测运动信息特征序列包括 T 个预测运动信息特征；

确定模块 402，用于根据该获取模块 401 获取的该真实运动信息特征序列与该获取模块 401 获取的该预测运动信息特征序列，确定模型参数；

训练模块 403，用于采用该确定模块 402 确定的该模型参数对该待训练运动预测网络模型进行训练，得到运动预测网络模型。

本实施例中，获取模块 401 获取待训练视频文件所对应的光流图序列，其中，该光流图序列中包括多个光流图，该获取模块 401 获取该待训练视频文件所对应的待训练外观信息特征序列，其中，该待训练外观信息特征序列包括多个待训练外观信息特征，该获取模块 401 根据该待训练视频文件所对应的该光流图序列，获取真实运动信息特征序列，其中，该真实运动信息特征序列包括 T 个真实运动信息特征，该 T 为大于 1 的整数，该获取模块 401 通过待训练运动预测网络模型获取该待训练外观信息特征序列所对应的预测运动信息特征序列，其中，该预测运动信息特征序列包括 T 个预测运动信息特征，确定模块 402 根据该获取模块 401 获取的该真实运动信息特征序列与该获取模块 401 获取的该预测运动信息特征序列，确定模型参数，训练模块 403 采用该确定模块 402 确定的该模型参数对该待训练运动预测网络模型进行训练，得到运动预测网络模型。

本申请实施例中，提供了一种模型训练装置，即该模型训练装置首先获取待训练视频文件所对应的光流图序列，然后获取待训练视频文件所对应的待训练外观信息特征序列，待训练外观信息特征序列包括多个待训练外观信息特征，模型训练装置根据待训练视频文件所对应的光流图序列，获取真实运动信息特征序列，再通过待训练运动预测网络模型获取待训练外观信息特征序列所对应的预测运动信息特征序列，其中，预测运动信息特征序列包括 T 个预测运动信息特征，根据真实运动信息特征序列与预测运动信息特征序列，确定模型参数，最后模型训练装置可以采用模型参数对待训练运动预测网络模型进行训练，得到运动预测网络模型。通过上述方式，在模型训练的过程中使用视频文件的光流图作为训练对象，从而使训练得到的运动预测网络模型具有隐式获取视频文件中光流图的功能，在实际应用中，通过运动预测网络模型就能够直接输出运动信息特征，不需要再特意计算出光流图，从而大幅度地减少计算量和穿存储资源。

可选地，在上述图 9 所对应的实施例的基础上，本申请实施例提供的模型训练装置 40 的另一实施例中，

该确定模块 402，具体用于采用损失函数对该真实运动信息特征序列与该预测运动信息特征序列进行计算，得到目标损失结果；

根据该目标损失结果确定该模型参数。

其次，本申请实施例中，提供了一种根据真实运动信息特征序列与预测运动信息特征序列确定模型参数的方法，即模型训练装置先采用损失函数对真实运动信息特征序列与预测运动信息特征序列进行计算，得到目标损失结果，然后根据目标损失结果确定模型参数。通过上述方式，能够采用合适的损失函数计算得到损失结果，基于损失结果的最小值计算得到模型参数，该模型参数能够有效地训练得到机器学习模型，从而提升模型训练的可靠性和准确度。

可选地，在上述图 9 所对应的实施例的基础上，本申请实施例提供的模型训练装置 40 的另一实施例中，

该确定模块 402，具体用于采用如下方式计算该目标损失结果：

$$L_{MSE} = \frac{1}{T} \sum_t \left\| \tilde{S}^t - S^t \right\|_2^2 ;$$

其中，该  $L_{MSE}$  表示该目标损失结果，该  $T$  表示该真实运动信息特征序列中的真实运动信息特征个数，该  $\tilde{S}^t$  表示该真实运动信息特征序列中第  $t$  个真实运动信息特征，该  $S^t$  表示该预测运动信息特征序列中第  $t$  个预测运动信息特征，该  $t$  为大于或等于 1，且小于或等于该  $T$  的整数。

再次，本申请实施例中，提供了一种计算目标损失结果的具体方式，即利用 MSE 损失函数对真实运动信息特征序列与预测运动信息特征序列进行计算，得到目标损失结果。通过上述方式，采用 MSE 损失函数进行计算能够降低计算的复杂度，即便使用固定的学习率也可以有效收敛。MSE 损失函数的梯度随损失增大而增大，而损失趋于 0 时则会减小，这使得在训练结束时，使用 MSE 损失函数的结果会更精确。

本申请实施例还提供了另一种视频分类装置，如图 10 所示，为了便于说明，仅示出了与本申请实施例相关的部分，具体技术细节未揭示的，请参照本申请实施例方法部分。上述计算机设备可以被提供为下述图 10 所示的终端设备，也可以为被提供为下述图 11 所示的服务器。该终端设备可以为包括手机、平板电脑、个人数字助理 (personal digital assistant, PDA)、销售终端设备 (point of sales, POS)、车载电脑等任意终端设备设备，以终端设备为手机为例：

图 10 示出的是与本申请实施例提供的终端设备相关的手机的部分结构的框图。参考图 10，手机包括：射频 (radio frequency, RF) 电路 510、存储器 520、输入单元 530、显示单元 540、传感器 550、音频电路 560、无线保真 (wireless fidelity, WiFi) 模块 570、处理器 580、以及电源 590 等部件。本领域技术人员可以理解，图 10 中示出的手机结构并不构成对手机的限定，可以包括比图示更多或更少的部件，或者组合某些部件，或者不同的部件布置。

下面结合图 10 对手机的各个构成部件进行具体的介绍：

RF 电路 510 可用于收发信息或通话过程中，信号的接收和发送，特别地，将基站的下行信息接收后，给处理器 580 处理；另外，将设计上行的数据发送给基站。通常，RF 电路 510 包括但不限于天线、至少一个放大器、收发信机、耦合器、低噪声放大器 (low noise amplifier, LNA)、双工器等。此外，RF 电路 510 还可以通过无线通信与网络和其他设备通信。上述无线通信可以使用任一通信标准或协议，包括但不限于全球移动通讯系统 (global system of mobile communication, GSM)、通用分组无线服务 (general packet radio service, GPRS)、码分多址 (code division multiple access, CDMA)、宽带码分多址 (wideband code division multiple access, WCDMA)、长期演进 (long term evolution, LTE)、电子邮件、短消息服务 (short messaging service, SMS) 等。

存储器 520 可用于存储软件程序以及模块，处理器 580 通过运行存储在存储器 520 的软件程序以及模块，从而执行手机的各种功能应用以及数据处理。存储器 520 可主要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、至少一个功能所需的应用程序 (比

如声音播放功能、图像播放功能等)等;存储数据区可存储根据手机的使用所创建的数据(比如音频数据、电话本等)等。此外,存储器 520 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、或其他易失性固态存储器件。

输入单元 530 可用于接收输入的数字或字符信息,以及产生与手机的用户设置以及功能控制有关的键信号输入。具体地,输入单元 530 可包括触控面板 531 以及其他输入设备 532。触控面板 531,也称为触摸屏,可收集用户在其上或附近的触摸操作(比如用户使用手指、触笔等任何适合的物体或附件在触控面板 531 上或在触控面板 531 附近的操作),并根据预先设定的程式驱动相应的连接装置。可选的,触控面板 531 可包括触摸检测装置和触摸控制器两个部分。其中,触摸检测装置检测用户的触摸方位,并检测触摸操作带来的信号,将信号传送给触摸控制器;触摸控制器从触摸检测装置上接收触摸信息,并将它转换成触点坐标,再送给处理器 580,并能接收处理器 580 发来的命令并加以执行。此外,可以采用电阻式、电容式、红外线以及表面声波等多种类型实现触控面板 531。除了触控面板 531,输入单元 530 还可以包括其他输入设备 532。具体地,其他输入设备 532 可以包括但不限于物理键盘、功能键(比如音量控制按键、开关按键等)、轨迹球、鼠标、操作杆等中的一种或多种。

显示单元 540 可用于显示由用户输入的信息或提供给用户的信息以及手机的各种菜单。显示单元 540 可包括显示面板 541,可选的,可以采用液晶显示器(liquid crystal display, LCD)、有机发光二极管(organic light-emitting diode, OLED)等形式来配置显示面板 541。进一步的,触控面板 531 可覆盖显示面板 541,当触控面板 531 检测到在其上或附近的触摸操作后,传送给处理器 580 以确定触摸事件的类型,随后处理器 580 根据触摸事件的类型在显示面板 541 上提供相应的视觉输出。虽然在图 10 中,触控面板 531 与显示面板 541 是作为两个独立的部件来实现手机的输入和输入功能,但是在某些实施例中,可以将触控面板 531 与显示面板 541 集成而实现手机的输入和输出功能。

手机还可包括至少一种传感器 550,比如光传感器、运动传感器以及其他传感器。具体地,光传感器可包括环境光传感器及接近传感器,其中,环境光传感器可根据环境光线的明暗来调节显示面板 541 的亮度,接近传感器可在手机移动到耳边时,关闭显示面板 541 和/或背光。作为运动传感器的一种,加速计传感器可检测各个方向上(一般为三轴)加速度的大小,静止时可检测出重力的大小及方向,可用于识别手机姿态的应用(比如横竖屏切换、相关游戏、磁力计姿态校准)、振动识别相关功能(比如计步器、敲击)等;至于手机还可配置的陀螺仪、气压计、湿度计、温度计、红外线传感器等其他传感器,在此不再赘述。

音频电路 560、扬声器 561,传声器 562 可提供用户与手机之间的音频接口。音频电路 560 可将接收到的音频数据转换后的电信号,传输到扬声器 561,由扬声器 561 转换为声音信号输出;另一方面,传声器 562 将收集的声音信号转换为电信号,由音频电路 560 接收后转换为音频数据,再将音频数据输出处理器 580 处理后,经 RF 电路 510 以发送给比如另一手机,或者将音频数据输出至存储器 520 以便进一步处理。

WiFi 属于短距离无线传输技术,手机通过 WiFi 模块 570 可以帮助用户收发电子邮件、浏览网页和访问流式媒体等,它为用户提供了无线的宽带互联网访问。虽然图 10 示出了 WiFi 模块 570,但是可以理解的是,其并不属于手机的必须构成,完全可以根据需要在不改变发明的本质的范围内而省略。

处理器 580 是手机的控制中心,利用各种接口和线路连接整个手机的各个部分,通过运行或执行存储在存储器 520 内的软件程序和/或模块,以及调用存储在存储器 520 内的数据,执行手机的各种功能和处理数据,从而对手机进行整体监控。可选的,处理器 580 可包括一个或多个处理单元;可选的,处理器 580 可集成应用处理器和调制解调处理器,其中,应用处理器主要处理操作系统、用户界面和应用程序等,调制解调处理器主要处理无线通信。可以理解的是,上述调制解调处理器也可以不集成到处理器 580 中。

手机还包括给各个部件供电的电源 590(比如电池),可选的,电源可以通过电源管理系统与处理器 580 逻辑相连,从而通过电源管理系统实现管理充电、放电、以及功耗管理等功能。

尽管未示出,手机还可以包括摄像头、蓝牙模块等,在此不再赘述。

在本申请实施例中,该终端设备所包括的处理器 580 还具有以下功能:

获取待分类视频文件所对应的图像帧序列,其中,该图像帧序列中包括 T 帧图像,该 T 为大于 1 的整数;

获取该图像帧序列所对应的外观信息特征序列，其中，该外观信息特征序列中包括 T 个外观信息特征，该外观信息特征与该图像帧具有对应关系；

获取该外观信息特征序列所对应的运动信息特征序列，其中，该运动信息特征序列包括 T 个运动信息特征，该运动信息特征与该外观信息特征具有对应关系；

根据该外观信息特征序列以及该运动信息特征序列确定该待分类视频文件的视频分类结果。

在一种可能实现方式中，该处理器 580 用于执行：

通过图像分类网络模型获取该图像帧序列所对应的外观信息特征序列；

该获取该外观信息特征序列所对应的运动信息特征序列，包括：

通过运动预测网络模型获取该外观信息特征序列所对应的运动信息特征序列，该运动预测网络模型用于预测该外观信息特征所对应的该运动信息特征。

在一种可能实现方式中，该处理器 580 用于执行：

获取该待分类视频文件；

对该待分类视频文件进行解码处理，得到该图像帧序列，其中，每个帧图像具有相同的采样频率。

在一种可能实现方式中，该处理器 580 用于执行：

获取第 t 帧图像所对应的外观信息特征，其中，该 t 大于或等于 1，且小于或等于该 T 的整数；

当获取到该 T 帧图像中每帧图像所对应的外观信息特征时，根据该 T 个外观信息特征生成该外观信息特征序列。

在一种可能实现方式中，该处理器 580 用于执行：

获取第 i 个步骤所对应的第 t 个外观信息特征，其中，该 i 为大于或等于 1 的整数，该 t 大于或等于 1，且小于或等于该 T 的整数；

获取第(i-1)个步骤所对应的第(t+1)个运动信息特征以及第(t-1)个运动信息特征；

根据该第 t 个外观信息特征、该第(t+1)个运动信息特征以及第(t-1)个运动信息特征，确定该第 i 个步骤所对应的第 t 个运动信息特征；

当获取到该 T 个外观信息特征中每个外观信息特征所对应的运动信息特征时，根据该 T 个运动信息特征生成该运动信息特征序列。

在一种可能实现方式中，该处理器 580 用于执行：

通过运动预测网络模型的更新门，根据该第 t 个外观信息特征、该第(t+1)个运动信息特征以及第(t-1)个运动信息特征，计算得到该第 i 个步骤所对应的第(t+1)个更新门向量以及第(t-1)个更新门向量；

通过该运动预测网络模型的输出层，根据该第 t 个外观信息特征、该第(t+1)个运动信息特征以及第(t-1)个运动信息特征，计算得到该第 i 个步骤所对应的第(t+1)个输出门向量以及第(t-1)个输出门向量；

通过该运动预测网络模型的隐藏层，根据该第 t 个外观信息特征、该第(t+1)个运动信息特征、第(t-1)个运动信息特征、该第(t+1)个更新门向量以及该第(t-1)个更新门向量，计算得到该第 i 个步骤所对应的第 t 个隐藏层特征向量；

通过该运动预测网络模型的最终输出层，根据该第 t 个隐藏层特征向量、该第 i 个步骤所对应的第(t+1)个输出门向量以及第(t-1)个输出门向量，计算得到该第 i 个步骤所对应的第 t 个运动信息特征。

在一种可能实现方式中，该处理器 580 用于执行：

采用如下方式计算该第(t+1)个更新门向量以及该第(t-1)个更新门向量：

$$r_{t-1} = \sigma(W_r x_t + U_{r,t-1} s_{t-1}^{i-1});$$

$$r_{t+1} = \sigma(W_r x_t + U_{r,t+1} s_{t+1}^{i-1});$$

其中，该  $r_{t-1}$  表示该第 i 个步骤所对应的该第(t-1)个更新门向量，该 t 表示第 t 个时刻，该  $r_{t+1}$  表示该第 i 个步骤所对应的该第(t+1)个更新门向量，该  $\sigma(\cdot)$  表示线性整流函数，该  $W_r$  表

示第一参数矩阵, 该  $U_{r,t-1}$  表示该第(t-1)个更新门向量所对应的第二参数矩阵, 该  $U_{r,t+1}$  表示该第(t+1)个更新门向量所对应的第三参数矩阵, 该  $x_t$  表示该第 t 个外观信息特征, 该  $s_{t-1}^{i-1}$  表示该第(i-1)个步骤所对应的该第(t-1)个运动信息特征, 该  $s_{t+1}^{i-1}$  表示该第(i-1)个步骤所对应的该第(t+1)个运动信息特征;

该处理器 580 用于执行:

采用如下方式计算该第(t+1)个输出门向量以及该第(t-1)个输出门向量:

$$[z_{t+1}, z, z_{t-1}] = \text{soft max}([\tilde{z}_{t+1}, 1, \tilde{z}_{t-1}]);$$

$$\tilde{z}_{t-1} = \sigma(W_z x_t + U_{z,t-1} s_{t-1}^{i-1});$$

$$\tilde{z}_{t+1} = \sigma(W_z x_t + U_{z,t+1} s_{t+1}^{i-1});$$

其中, 该  $z_{t+1}$  表示该第 i 个步骤所对应的该第(t+1)个输出门向量, 该  $z_{t-1}$  表示该第 i 个步骤所对应的该第(t-1)个输出门向量, 该  $z$  表示预设输出门向量, 该  $\tilde{z}_{t-1}$  表示该第 i 个步骤所对应的第(t-1)个待处理输出门向量,  $\tilde{z}_{t+1}$  表示该第 i 个步骤所对应的第(t+1)个待处理输出门向量, 该  $\text{soft max}()$  表示归一化处理函数, 该  $W_z$  表示第四参数矩阵, 该  $U_{z,t-1}$  表示该第(t-1)个输出门向量所对应的第五参数矩阵, 该  $U_{z,t+1}$  表示该第(t+1)个输出门向量所对应的第六参数矩阵;

该处理器 580 用于执行:

采用如下方式计算该第 t 个隐藏层特征向量:

$$h_t = \phi(W_h x_t + U_h [s_{t+1}^{i-1} \circ r_{t+1}, s_{t-1}^{i-1} \circ r_{t-1}]^T);$$

其中, 该  $h_t$  表示该第 i 个步骤所对应的该第 t 个隐藏层特征向量, 该  $\phi()$  表示 sigmoid 函数, 该  $W_h$  表示第七参数矩阵, 该  $U_h$  表示第八参数矩阵, 该  $[\ ]^T$  表示矩阵转置, 该  $\circ$  表示点乘;

该处理器 580 用于执行:

采用如下方式计算该第 t 个运动信息特征:

$$s_t^i = z \circ h_t + z_{t-1} \circ s_{t-1}^{i-1} + z_{t+1} \circ s_{t+1}^{i-1};$$

其中, 该  $s_t^i$  表示该第 i 个步骤所对应的该第 t 个运动信息特征。

在一种可能实现方式中, 该处理器 580 用于执行:

获取该外观信息特征序列以及该运动信息特征序列所对应的视频类别向量, 其中, 该视频类别向量包括 C 个视频类别元素, 该 C 为大于 1 的整数;

从该视频类别向量中确定至少一个目标视频类别元素;

根据该至少一个目标视频类别元素生成该待分类视频文件的该视频分类结果。

图 11 是本申请实施例提供的一种服务器结构示意图, 该服务器 600 可因配置或性能不同而产生比较大的差异, 可以包括一个或一个以上中央处理器 (central processing units, CPU) 622 (例如, 一个或一个以上处理器) 和存储器 632, 一个或一个以上存储应用程序 642 或数据 644 的存储介质 630 (例如一个或一个以上海量存储设备)。其中, 存储器 632 和存储介质 630 可以是短暂存储或持久存储。存储在存储介质 630 的程序可以包括一个或一个以上模块 (图示没标出), 每个模块可以包括对服务器中的一系列指令操作。更进一步地, 中央处理器 622 可以设置为与存储介质 630 通信, 在服务器 600 上执行存储介质 630 中的一系列指令操作。

服务器 600 还可以包括一个或一个以上电源 626, 一个或一个以上有线或无线网络接口 650, 一个或一个以上输入输出接口 658, 和/或, 一个或一个以上操作系统 641, 例如 Windows ServerTM, Mac OS XTM, UnixTM, LinuxTM, FreeBSDTM 等等。

上述实施例中由服务器所执行的步骤可以基于该图 11 所示的服务器结构。

在本申请实施例中, 该服务器所包括的 CPU 622 还具有以下功能:

获取待分类视频文件所对应的图像帧序列, 其中, 该图像帧序列中包括 T 帧图像, 该 T 为大于 1 的整数;

获取该图像帧序列所对应的外观信息特征序列, 其中, 该外观信息特征序列中包括 T 个外观信息特征, 该外观信息特征与该图像帧具有对应关系;

获取该外观信息特征序列所对应的运动信息特征序列, 其中, 该运动信息特征序列包括 T 个运动信息特征, 该运动信息特征与该外观信息特征具有对应关系;

根据该外观信息特征序列以及该运动信息特征序列确定该待分类视频文件的视频分类结果。

在一种可能实现方式中, 该 CPU 622 用于执行:

通过图像分类网络模型获取该图像帧序列所对应的外观信息特征序列;

该获取该外观信息特征序列所对应的运动信息特征序列, 包括:

通过运动预测网络模型获取该外观信息特征序列所对应的运动信息特征序列, 该运动预测网络模型用于预测该外观信息特征所对应的该运动信息特征。

在一种可能实现方式中, 该 CPU 622 用于执行:

获取该待分类视频文件;

对该待分类视频文件进行解码处理, 得到该图像帧序列, 其中, 每个帧图像具有相同的采样频率。

在一种可能实现方式中, 该 CPU 622 用于执行:

获取第 t 帧图像所对应的外观信息特征, 其中, 该 t 大于或等于 1, 且小于或等于该 T 的整数;

当获取到该 T 帧图像中每帧图像所对应的外观信息特征时, 根据该 T 个外观信息特征生成该外观信息特征序列。

在一种可能实现方式中, 该 CPU 622 用于执行:

获取第 i 个步骤所对应的第 t 个外观信息特征, 其中, 该 i 为大于或等于 1 的整数, 该 t 大于或等于 1, 且小于或等于该 T 的整数;

获取第(i-1)个步骤所对应的第(t+1)个运动信息特征以及第(t-1)个运动信息特征;

根据该第 t 个外观信息特征、该第(t+1)个运动信息特征以及第(t-1)个运动信息特征, 确定该第 i 个步骤所对应的第 t 个运动信息特征;

当获取到该 T 个外观信息特征中每个外观信息特征所对应的运动信息特征时, 根据该 T 个运动信息特征生成该运动信息特征序列。

在一种可能实现方式中, 该 CPU 622 用于执行:

通过运动预测网络模型的更新门, 根据该第 t 个外观信息特征、该第(t+1)个运动信息特征以及第(t-1)个运动信息特征, 计算得到该第 i 个步骤所对应的第(t+1)个更新门向量以及第(t-1)个更新门向量;

通过该运动预测网络模型的输出层, 根据该第 t 个外观信息特征、该第(t+1)个运动信息特征以及第(t-1)个运动信息特征, 计算得到该第 i 个步骤所对应的第(t+1)个输出门向量以及第(t-1)个输出门向量;

通过该运动预测网络模型的隐藏层, 根据该第 t 个外观信息特征、该第(t+1)个运动信息特征、第(t-1)个运动信息特征、该第(t+1)个更新门向量以及该第(t-1)个更新门向量, 计算得到该第 i 个步骤所对应的第 t 个隐藏层特征向量;

通过该运动预测网络模型的最终输出层, 根据该第 t 个隐藏层特征向量、该第 i 个步骤所对应的第(t+1)个输出门向量以及第(t-1)个输出门向量, 计算得到该第 i 个步骤所对应的第 t 个运动信息特征。

在一种可能实现方式中, 该 CPU 622 用于执行:

采用如下方式计算该第(t+1)个更新门向量以及该第(t-1)个更新门向量:

$$r_{t-1} = \sigma(W_r x_t + U_{r,t-1} s_{t-1}^{i-1});$$

$$r_{t+1} = \sigma(W_r x_t + U_{r,t+1} s_{t+1}^{i-1});$$

其中, 该  $r_{t-1}$  表示该第  $i$  个步骤所对应的该第  $(t-1)$  个更新门向量, 该  $t$  表示第  $t$  个时刻, 该  $r_{t+1}$  表示该第  $i$  个步骤所对应的该第  $(t+1)$  个更新门向量, 该  $\sigma(\cdot)$  表示线性整流函数, 该  $W_r$  表示第一参数矩阵, 该  $U_{r,t-1}$  表示该第  $(t-1)$  个更新门向量所对应的第二参数矩阵, 该  $U_{r,t+1}$  表示该第  $(t+1)$  个更新门向量所对应的第三参数矩阵, 该  $x_t$  表示该第  $t$  个外观信息特征, 该  $s_{t-1}^{i-1}$  表示该第  $(i-1)$  个步骤所对应的该第  $(t-1)$  个运动信息特征, 该  $s_{t+1}^{i-1}$  表示该第  $(i-1)$  个步骤所对应的该第  $(t+1)$  个运动信息特征;

该 CPU 622 用于执行:

采用如下方式计算该第  $(t+1)$  个输出门向量以及该第  $(t-1)$  个输出门向量:

$$[z_{t+1}, z, z_{t-1}] = \text{soft max}([\tilde{z}_{t+1}, 1, \tilde{z}_{t-1}]);$$

$$\tilde{z}_{t-1} = \sigma(W_z x_t + U_{z,t-1} s_{t-1}^{i-1});$$

$$\tilde{z}_{t+1} = \sigma(W_z x_t + U_{z,t+1} s_{t+1}^{i-1});$$

其中, 该  $z_{t+1}$  表示该第  $i$  个步骤所对应的该第  $(t+1)$  个输出门向量, 该  $z_{t-1}$  表示该第  $i$  个步骤所对应的该第  $(t-1)$  个输出门向量, 该  $z$  表示预设输出门向量, 该  $\tilde{z}_{t-1}$  表示该第  $i$  个步骤所对应的第  $(t-1)$  个待处理输出门向量,  $\tilde{z}_{t+1}$  表示该第  $i$  个步骤所对应的第  $(t+1)$  个待处理输出门向量, 该  $\text{soft max}(\cdot)$  表示归一化处理函数, 该  $W_z$  表示第四参数矩阵, 该  $U_{z,t-1}$  表示该第  $(t-1)$  个输出门向量所对应的第五参数矩阵, 该  $U_{z,t+1}$  表示该第  $(t+1)$  个输出门向量所对应的第六参数矩阵;

该 CPU 622 用于执行:

采用如下方式计算该第  $t$  个隐藏层特征向量:

$$h_t = \phi\left(W_h x_t + U_h \left[s_{t+1}^{i-1} \circ r_{t+1}, s_{t-1}^{i-1} \circ r_{t-1}\right]^T\right);$$

其中, 该  $h_t$  表示该第  $i$  个步骤所对应的该第  $t$  个隐藏层特征向量, 该  $\phi(\cdot)$  表示 sigmoid 函数, 该  $W_h$  表示第七参数矩阵, 该  $U_h$  表示第八参数矩阵, 该  $[\cdot]^T$  表示矩阵转置, 该  $\circ$  表示点乘;

该 CPU 622 用于执行:

采用如下方式计算该第  $t$  个运动信息特征:

$$s_t^i = z \circ h_t + z_{t-1} \circ s_{t-1}^{i-1} + z_{t+1} \circ s_{t+1}^{i-1};$$

其中, 该  $s_t^i$  表示该第  $i$  个步骤所对应的该第  $t$  个运动信息特征。

在一种可能实现方式中, 该 CPU 622 用于执行:

获取该外观信息特征序列以及该运动信息特征序列所对应的视频类别向量, 其中, 该视频类别向量包括  $C$  个视频类别元素, 该  $C$  为大于 1 的整数;

从该视频类别向量中确定至少一个目标视频类别元素;

根据该至少一个目标视频类别元素生成该待分类视频文件的该视频分类结果。

在本申请实施例中, 该服务器所包括的 CPU 622 还具有以下功能:

获取待训练视频文件所对应的光流图序列，其中，该光流图序列中包括多个光流图；

获取该待训练视频文件所对应的待训练外观信息特征序列，其中，该待训练外观信息特征序列包括多个待训练外观信息特征；

根据该待训练视频文件所对应的该光流图序列，获取真实运动信息特征序列，其中，该真实运动信息特征序列包括  $T$  个真实运动信息特征，该  $T$  为大于 1 的整数；

通过待训练运动预测网络模型获取该待训练外观信息特征序列所对应的预测运动信息特征序列，其中，该预测运动信息特征序列包括  $T$  个预测运动信息特征；

根据该真实运动信息特征序列与该预测运动信息特征序列，确定模型参数；

采用该模型参数对该待训练运动预测网络模型进行训练，得到运动预测网络模型。

在一种可能实现方式中，该处理器用于执行：

采用损失函数对该真实运动信息特征序列与该预测运动信息特征序列进行计算，得到目标损失结果；

根据该目标损失结果确定该模型参数。

在一种可能实现方式中，该处理器用于执行：

采用如下方式计算该目标损失结果：

$$L_{MSE} = \frac{1}{T} \sum_i^T \left\| \tilde{S}^i - S^i \right\|_2^2;$$

其中，该  $L_{MSE}$  表示该目标损失结果，该  $T$  表示该真实运动信息特征序列中的真实运动信息特征个数，该  $\tilde{S}^i$  表示该真实运动信息特征序列中第  $i$  个真实运动信息特征，该  $S^i$  表示该预测运动信息特征序列中第  $i$  个预测运动信息特征，该  $i$  为大于或等于 1，且小于或等于该  $T$  的整数。

所属领域的技术人员可以清楚地了解到，为描述的方便和简洁，上述描述的系统、装置和单元的具体工作过程，可以参考前述方法实施例中的对应过程，在此不再赘述。

在本申请所提供的几个实施例中，应该理解到，所揭露的系统、装置和方法，可以通过其它的方式实现。例如，以上所描述的装置实施例仅仅是示意性的，例如，该单元的划分，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信连接可以是通过一些接口，装置或单元的间接耦合或通信连接，可以是电性，机械或其它的形式。

该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。

另外，在本申请各个实施例中的各功能单元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。

该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，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，本申请的技术方案本质上或者说对现有技术做出贡献的部分或者该技术方案的全部或部分可以以软件产品的形式体现出来，该计算机软件产品存储在一个存储介质中，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）执行本申请各个实施例该方法的全部或部分步骤。而前述的存储介质包括：U 盘、移动硬盘、只读存储器（read-only memory，ROM）、随机存取存储器（random access memory，RAM）、磁碟或者光盘等各种可以存储程序代码的介质。

以上该，以上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。

## 权利要求书

1、一种视频分类的方法，其特征在于，应用于计算机设备，包括：

获取待分类视频文件所对应的图像帧序列，其中，所述图像帧序列中包括 T 帧图像，所述 T 为大于 1 的整数；

获取所述图像帧序列所对应的外观信息特征序列，其中，所述外观信息特征序列中包括 T 个外观信息特征，所述外观信息特征与所述图像帧具有对应关系；

获取所述外观信息特征序列所对应的运动信息特征序列，其中，所述运动信息特征序列包括 T 个运动信息特征，所述运动信息特征与所述外观信息特征具有对应关系；

根据所述外观信息特征序列以及所述运动信息特征序列确定所述待分类视频文件的视频分类结果。

2、根据权利要求 1 所述的方法，其特征在于，所述获取所述图像帧序列所对应的外观信息特征序列，包括：

通过图像分类网络模型获取所述图像帧序列所对应的外观信息特征序列；

所述获取所述外观信息特征序列所对应的运动信息特征序列，包括：

通过运动预测网络模型获取所述外观信息特征序列所对应的运动信息特征序列，所述运动预测网络模型用于预测所述外观信息特征所对应的所述运动信息特征。

3、根据权利要求 1 所述的方法，其特征在于，所述获取待分类视频文件所对应的图像帧序列，包括：

获取所述待分类视频文件；

对所述待分类视频文件进行解码处理，得到所述图像帧序列，其中，每个帧图像具有相同的采样频率。

4、根据权利要求 1 所述的方法，其特征在于，所述获取所述图像帧序列所对应的外观信息特征序列，包括：

获取第 t 帧图像所对应的外观信息特征，其中，所述 t 大于或等于 1，且小于或等于所述 T 的整数；

当获取到所述 T 帧图像中每帧图像所对应的外观信息特征时，根据所述 T 个外观信息特征生成所述外观信息特征序列。

5、根据权利要求 1 所述的方法，其特征在于，所述获取所述外观信息特征序列所对应的运动信息特征序列，包括：

获取第 i 个步骤所对应的第 t 个外观信息特征，其中，所述 i 为大于或等于 1 的整数，所述 t 大于或等于 1，且小于或等于所述 T 的整数；

获取第(i-1)个步骤所对应的第(t+1)个运动信息特征以及第(t-1)个运动信息特征；

根据所述第 t 个外观信息特征、所述第(t+1)个运动信息特征以及第(t-1)个运动信息特征，确定所述第 i 个步骤所对应的第 t 个运动信息特征；

当获取到所述 T 个外观信息特征中每个外观信息特征所对应的运动信息特征时，根据所述 T 个运动信息特征生成所述运动信息特征序列。

6、根据权利要求 5 所述的方法，其特征在于，所述根据所述第 t 个外观信息特征、所述第(t+1)个运动信息特征以及第(t-1)个运动信息特征，确定所述第 i 个步骤所对应的第 t 个运动信息特征，包括：

通过运动预测网络模型的更新门，根据所述第 t 个外观信息特征、所述第(t+1)个运动信息特征以及第(t-1)个运动信息特征，计算得到所述第 i 个步骤所对应的第(t+1)个更新门向量以及第(t-1)个更新门向量；

通过所述运动预测网络模型的输出门，根据所述第 t 个外观信息特征、所述第(t+1)个运动信息特征以及第(t-1)个运动信息特征，计算得到所述第 i 个步骤所对应的第(t+1)个输出门向量以及第(t-1)个输出门向量；

通过所述运动预测网络模型的隐藏层，根据所述第 t 个外观信息特征、所述第(t+1)个运动信息特征、第(t-1)个运动信息特征、所述第(t+1)个更新门向量以及所述第(t-1)个更新门向量，计算得到所述第 i 个步骤所对应的第 t 个隐藏层特征向量；

通过所述运动预测网络模型的最终输出层，根据所述第 t 个隐藏层特征向量、所述第 i

个步骤所对应的第(t+1)个输出门向量以及第(t-1)个输出门向量,计算得到所述第i个步骤所对应的第t个运动信息特征。

7、根据权利要求6所述的方法,其特征在于,所述通过运动预测网络模型的更新门,根据所述第t个外观信息特征、所述第(t+1)个运动信息特征以及第(t-1)个运动信息特征,计算得到所述第i个步骤所对应的第(t+1)个更新门向量以及第(t-1)个更新门向量,包括:

采用如下方式计算所述第(t+1)个更新门向量以及所述第(t-1)个更新门向量:

$$\begin{aligned} r_{t-1} &= \sigma(W_r x_t + U_{r,t-1} s_{t-1}^{i-1}); \\ r_{t+1} &= \sigma(W_r x_t + U_{r,t+1} s_{t+1}^{i-1}); \end{aligned}$$

其中,所述 $r_{t-1}$ 表示所述第i个步骤所对应的所述第(t-1)个更新门向量,所述t表示第t个时刻,所述 $r_{t+1}$ 表示所述第i个步骤所对应的所述第(t+1)个更新门向量,所述 $\sigma(\cdot)$ 表示线性整流函数,所述 $W_r$ 表示第一参数矩阵,所述 $U_{r,t-1}$ 表示所述第(t-1)个更新门向量所对应的第二参数矩阵,所述 $U_{r,t+1}$ 表示所述第(t+1)个更新门向量所对应的第三参数矩阵,所述 $x_t$ 表示所述第t个外观信息特征,所述 $s_{t-1}^{i-1}$ 表示所述第(i-1)个步骤所对应的所述第(t-1)个运动信息特征,所述 $s_{t+1}^{i-1}$ 表示所述第(i-1)个步骤所对应的所述第(t+1)个运动信息特征;

所述通过所述运动预测网络模型的输出门,根据所述第t个外观信息特征、所述第(t+1)个运动信息特征以及第(t-1)个运动信息特征,计算得到所述第i个步骤所对应的第(t+1)个输出门向量以及第(t-1)个输出门向量,包括:

采用如下方式计算所述第(t+1)个输出门向量以及所述第(t-1)个输出门向量:

$$\begin{aligned} [z_{t+1}, z, z_{t-1}] &= \text{soft max}([\tilde{z}_{t+1}, 1, \tilde{z}_{t-1}]); \\ \tilde{z}_{t-1} &= \sigma(W_z x_t + U_{z,t-1} s_{t-1}^{i-1}); \\ \tilde{z}_{t+1} &= \sigma(W_z x_t + U_{z,t+1} s_{t+1}^{i-1}); \end{aligned}$$

其中,所述 $z_{t+1}$ 表示所述第i个步骤所对应的所述第(t+1)个输出门向量,所述 $z_{t-1}$ 表示所述第i个步骤所对应的所述第(t-1)个输出门向量,所述z表示预设输出门向量,所述 $\tilde{z}_{t-1}$ 表示所述第i个步骤所对应的第(t-1)个待处理输出门向量,所述 $\tilde{z}_{t+1}$ 表示所述第i个步骤所对应的第(t+1)个待处理输出门向量,所述 $\text{soft max}(\cdot)$ 表示归一化处理函数,所述 $W_z$ 表示第四参数矩阵,所述 $U_{z,t-1}$ 表示所述第(t-1)个输出门向量所对应的第五参数矩阵,所述 $U_{z,t+1}$ 表示所述第(t+1)个输出门向量所对应的第六参数矩阵;

所述通过所述运动预测网络模型的隐藏层,根据所述第t个外观信息特征、所述第(t+1)个运动信息特征、第(t-1)个运动信息特征、所述第(t+1)个更新门向量以及所述第(t-1)个更新门向量,计算得到所述第i个步骤所对应的第t个隐藏层特征向量,包括:

采用如下方式计算所述第t个隐藏层特征向量:

$$h_t = \phi\left(W_h x_t + U_h \left[s_{t+1}^{i-1} \circ r_{t+1}, s_{t-1}^{i-1} \circ r_{t-1}\right]^T\right);$$

其中,所述 $h_t$ 表示所述第i个步骤所对应的所述第t个隐藏层特征向量,所述 $\phi(\cdot)$ 表示sigmoid函数,所述 $W_h$ 表示第七参数矩阵,所述 $U_h$ 表示第八参数矩阵,所述 $[\cdot]^T$ 表示矩阵转

置, 所述 $\circ$ 表示点乘;

所述通过所述运动预测网络模型的最终输出层, 根据所述第  $t$  个隐藏层特征向量、所述第  $i$  个步骤所对应的第  $(t+1)$  个输出门向量以及第  $(t-1)$  个输出门向量, 计算得到所述第  $i$  个步骤所对应的第  $t$  个运动信息特征, 包括:

采用如下方式计算所述第  $t$  个运动信息特征:

$$s_t^i = z \circ h_t + z_{t-1} \circ s_{t-1}^{i-1} + z_{t+1} \circ s_{t+1}^{i-1};$$

其中, 所述  $s_t^i$  表示所述第  $i$  个步骤所对应的所述第  $t$  个运动信息特征。

8、根据权利要求 1 至 7 中任一项所述的方法, 其特征在于, 所述根据所述外观信息特征序列以及所述运动信息特征序列确定所述待分类视频文件的视频分类结果, 包括:

获取所述外观信息特征序列以及所述运动信息特征序列所对应的视频类别向量, 其中, 所述视频类别向量包括  $C$  个视频类别元素, 所述  $C$  为大于 1 的整数;

从所述视频类别向量中确定至少一个目标视频类别元素;

根据所述至少一个目标视频类别元素生成所述待分类视频文件的所述视频分类结果。

9、一种模型训练的方法, 其特征在于, 应用于服务器, 包括:

获取待训练视频文件所对应的光流图序列, 其中, 所述光流图序列中包括多个光流图;

获取所述待训练视频文件所对应的待训练外观信息特征序列, 其中, 所述待训练外观信息特征序列包括多个待训练外观信息特征;

根据所述待训练视频文件所对应的所述光流图序列, 获取真实运动信息特征序列, 其中, 所述真实运动信息特征序列包括  $T$  个真实运动信息特征, 所述  $T$  为大于 1 的整数;

通过待训练运动预测网络模型获取所述待训练外观信息特征序列所对应的预测运动信息特征序列, 其中, 所述预测运动信息特征序列包括  $T$  个预测运动信息特征;

根据所述真实运动信息特征序列与所述预测运动信息特征序列, 确定模型参数;

采用所述模型参数对所述待训练运动预测网络模型进行训练, 得到运动预测网络模型。

10、根据权利要求 9 所述的方法, 其特征在于, 所述根据所述真实运动信息特征序列与所述预测运动信息特征序列, 确定模型参数, 包括:

采用损失函数对所述真实运动信息特征序列与所述预测运动信息特征序列进行计算, 得到目标损失结果;

根据所述目标损失结果确定所述模型参数。

11、根据权利要求 10 所述的方法, 其特征在于, 所述采用损失函数对所述真实运动信息特征序列与所述预测运动信息特征序列进行计算, 得到目标损失结果, 包括:

采用如下方式计算所述目标损失结果:

$$L_{MSE} = \frac{1}{T} \sum_t \left\| \tilde{S}^t - S^t \right\|_2^2;$$

其中, 所述  $L_{MSE}$  表示所述目标损失结果, 所述  $T$  表示所述真实运动信息特征序列中的真实运动信息特征个数, 所述  $\tilde{S}^t$  表示所述真实运动信息特征序列中第  $t$  个真实运动信息特征, 所述  $S^t$  表示所述预测运动信息特征序列中第  $t$  个预测运动信息特征, 所述  $t$  为大于或等于 1, 且小于或等于所述  $T$  的整数。

12、一种计算机设备, 其特征在于, 包括: 存储器、收发器、处理器以及总线系统;

其中, 所述存储器用于存储程序;

所述处理器用于执行所述存储器中的程序, 包括如下步骤:

获取待分类视频文件所对应的图像帧序列, 其中, 所述图像帧序列中包括  $T$  帧图像, 所述  $T$  为大于 1 的整数;

获取所述图像帧序列所对应的外观信息特征序列, 其中, 所述外观信息特征序列中包括  $T$  个外观信息特征, 所述外观信息特征与所述图像帧具有对应关系;

获取所述外观信息特征序列所对应的运动信息特征序列, 其中, 所述运动信息特征序列

包括 T 个运动信息特征, 所述运动信息特征与所述外观信息特征具有对应关系;

根据所述外观信息特征序列以及所述运动信息特征序列确定所述待分类视频文件的视频分类结果;

所述总线系统用于连接所述存储器以及所述处理器, 以使所述存储器以及所述处理器进行通信。

13、根据权利要求 12 所述的计算机设备, 其特征在于, 所述处理器用于执行:

通过图像分类网络模型获取所述图像帧序列所对应的外观信息特征序列;

所述获取所述外观信息特征序列所对应的运动信息特征序列, 包括:

通过运动预测网络模型获取所述外观信息特征序列所对应的运动信息特征序列, 所述运动预测网络模型用于预测所述外观信息特征所对应的所述运动信息特征。

14、根据权利要求 12 所述的计算机设备, 其特征在于, 所述处理器用于执行:

获取所述待分类视频文件;

对所述待分类视频文件进行解码处理, 得到所述图像帧序列, 其中, 每个帧图像具有相同的采样频率。

15、根据权利要求 12 所述的计算机设备, 其特征在于, 所述处理器用于执行:

获取第 t 帧图像所对应的外观信息特征, 其中, 所述 t 大于或等于 1, 且小于或等于所述 T 的整数;

当获取到所述 T 帧图像中每帧图像所对应的外观信息特征时, 根据所述 T 个外观信息特征生成所述外观信息特征序列。

16、根据权利要求 12 所述的计算机设备, 其特征在于, 所述处理器用于执行:

获取第 i 个步骤所对应的第 t 个外观信息特征, 其中, 所述 i 为大于或等于 1 的整数, 所述 t 大于或等于 1, 且小于或等于所述 T 的整数;

获取第(i-1)个步骤所对应的第(t+1)个运动信息特征以及第(t-1)个运动信息特征;

根据所述第 t 个外观信息特征、所述第(t+1)个运动信息特征以及第(t-1)个运动信息特征, 确定所述第 i 个步骤所对应的第 t 个运动信息特征;

当获取到所述 T 个外观信息特征中每个外观信息特征所对应的运动信息特征时, 根据所述 T 个运动信息特征生成所述运动信息特征序列。

17、根据权利要求 16 所述的计算机设备, 其特征在于, 所述处理器用于执行:

通过运动预测网络模型的更新门, 根据所述第 t 个外观信息特征、所述第(t+1)个运动信息特征以及第(t-1)个运动信息特征, 计算得到所述第 i 个步骤所对应的第(t+1)个更新门向量以及第(t-1)个更新门向量;

通过所述运动预测网络模型的输出层, 根据所述第 t 个外观信息特征、所述第(t+1)个运动信息特征以及第(t-1)个运动信息特征, 计算得到所述第 i 个步骤所对应的第(t+1)个输出层向量以及第(t-1)个输出层向量;

通过所述运动预测网络模型的隐藏层, 根据所述第 t 个外观信息特征、所述第(t+1)个运动信息特征、第(t-1)个运动信息特征、所述第(t+1)个更新门向量以及所述第(t-1)个更新门向量, 计算得到所述第 i 个步骤所对应的第 t 个隐藏层特征向量;

通过所述运动预测网络模型的最终输出层, 根据所述第 t 个隐藏层特征向量、所述第 i 个步骤所对应的第(t+1)个输出层向量以及第(t-1)个输出层向量, 计算得到所述第 i 个步骤所对应的第 t 个运动信息特征。

18、根据权利要求 17 所述的计算机设备, 其特征在于, 所述处理器用于执行:

采用如下方式计算所述第(t+1)个更新门向量以及所述第(t-1)个更新门向量:

$$r_{t-1} = \sigma(W_r x_t + U_{r,t-1} s_{t-1}^{i-1});$$

$$r_{t+1} = \sigma(W_r x_t + U_{r,t+1} s_{t+1}^{i-1});$$

其中, 所述  $r_{t-1}$  表示所述第 i 个步骤所对应的所述第(t-1)个更新门向量, 所述 t 表示第 t 个时刻, 所述  $r_{t+1}$  表示所述第 i 个步骤所对应的所述第(t+1)个更新门向量, 所述  $\sigma(\cdot)$  表示线

性整流函数, 所述  $W_r$  表示第一参数矩阵, 所述  $U_{r,t-1}$  表示所述第(t-1)个更新门向量所对应的第二参数矩阵, 所述  $U_{r,t+1}$  表示所述第(t+1)个更新门向量所对应的第三参数矩阵, 所述  $x_t$  表示所述第 t 个外观信息特征, 所述  $s_{t-1}^{i-1}$  表示所述第(i-1)个步骤所对应的所述第(t-1)个运动信息特征, 所述  $s_{t+1}^{i-1}$  表示所述第(i-1)个步骤所对应的所述第(t+1)个运动信息特征;

所述处理器用于执行:

采用如下方式计算所述第(t+1)个输出门向量以及所述第(t-1)个输出门向量:

$$[z_{t+1}, z, z_{t-1}] = \text{soft max}([\tilde{z}_{t+1}, 1, \tilde{z}_{t-1}]);$$

$$\tilde{z}_{t-1} = \sigma(W_z x_t + U_{z,t-1} s_{t-1}^{i-1});$$

$$\tilde{z}_{t+1} = \sigma(W_z x_t + U_{z,t+1} s_{t+1}^{i-1});$$

其中, 所述  $z_{t+1}$  表示所述第 i 个步骤所对应的所述第(t+1)个输出门向量, 所述  $z_{t-1}$  表示所述第 i 个步骤所对应的所述第(t-1)个输出门向量, 所述  $z$  表示预设输出门向量, 所述  $\tilde{z}_{t-1}$  表示所述第 i 个步骤所对应的第(t-1)个待处理输出门向量,  $\tilde{z}_{t+1}$  表示所述第 i 个步骤所对应的第(t+1)个待处理输出门向量, 所述  $\text{soft max}(\cdot)$  表示归一化处理函数, 所述  $W_z$  表示第四参数矩阵, 所述  $U_{z,t-1}$  表示所述第(t-1)个输出门向量所对应的第五参数矩阵, 所述  $U_{z,t+1}$  表示所述第(t+1)个输出门向量所对应的第六参数矩阵;

所述处理器用于执行:

采用如下方式计算所述第 t 个隐藏层特征向量:

$$h_t = \phi\left(W_h x_t + U_h \left[s_{t+1}^{i-1} \circ r_{t+1}, s_{t-1}^{i-1} \circ r_{t-1}\right]^T\right);$$

其中, 所述  $h_t$  表示所述第 i 个步骤所对应的所述第 t 个隐藏层特征向量, 所述  $\phi(\cdot)$  表示 sigmoid 函数, 所述  $W_h$  表示第七参数矩阵, 所述  $U_h$  表示第八参数矩阵, 所述  $[\cdot]^T$  表示矩阵转置, 所述  $\circ$  表示点乘;

所述处理器用于执行:

采用如下方式计算所述第 t 个运动信息特征:

$$s_t^i = z \circ h_t + z_{t-1} \circ s_{t-1}^{i-1} + z_{t+1} \circ s_{t+1}^{i-1};$$

其中, 所述  $s_t^i$  表示所述第 i 个步骤所对应的所述第 t 个运动信息特征。

19、根据权利要求 12 至 18 中任一项所述的计算机设备, 其特征在于, 所述处理器用于执行:

获取所述外观信息特征序列以及所述运动信息特征序列所对应的视频类别向量, 其中, 所述视频类别向量包括 C 个视频类别元素, 所述 C 为大于 1 的整数;

从所述视频类别向量中确定至少一个目标视频类别元素;

根据所述至少一个目标视频类别元素生成所述待分类视频文件的所述视频分类结果。

20、一种服务器, 其特征在于, 包括: 存储器、收发器、处理器以及总线系统;

其中, 所述存储器用于存储程序;

所述处理器用于执行所述存储器中的程序, 包括如下步骤:

获取待训练视频文件所对应的光流图序列, 其中, 所述光流图序列中包括多个光流图;

获取所述待训练视频文件所对应的待训练外观信息特征序列, 其中, 所述待训练外观信

息特征序列包括多个待训练外观信息特征；

根据所述待训练视频文件所对应的所述光流图序列，获取真实运动信息特征序列，其中，所述真实运动信息特征序列包括 T 个真实运动信息特征，所述 T 为大于 1 的整数；

通过待训练运动预测网络模型获取所述待训练外观信息特征序列所对应的预测运动信息特征序列，其中，所述预测运动信息特征序列包括 T 个预测运动信息特征；

根据所述真实运动信息特征序列与所述预测运动信息特征序列，确定模型参数；

采用所述模型参数对所述待训练运动预测网络模型进行训练，得到运动预测网络模型；

所述总线系统用于连接所述存储器以及所述处理器，以使所述存储器以及所述处理器进行通信。

21、根据权利要求 20 所述的服务器，其特征在于，所述处理器用于执行：

采用损失函数对所述真实运动信息特征序列与所述预测运动信息特征序列进行计算，得到目标损失结果；

根据所述目标损失结果确定所述模型参数。

22、根据权利要求 21 所述的服务器，其特征在于，所述处理器用于执行：

采用如下方式计算所述目标损失结果：

$$L_{MSE} = \frac{1}{T} \sum_t \left\| \tilde{S}^t - S^t \right\|_2^2;$$

其中，所述  $L_{MSE}$  表示所述目标损失结果，所述 T 表示所述真实运动信息特征序列中的真实运动信息特征个数，所述  $\tilde{S}^t$  表示所述真实运动信息特征序列中第 t 个真实运动信息特征，所述  $S^t$  表示所述预测运动信息特征序列中第 t 个预测运动信息特征，所述 t 为大于或等于 1，且小于或等于所述 T 的整数。

23、一种计算机可读存储介质，包括指令，当其在计算机上运行时，使得计算机执行如权利要求 1 至 8 中任一项所述的方法，或执行如权利要求 9 至 11 中任一项所述的方法。

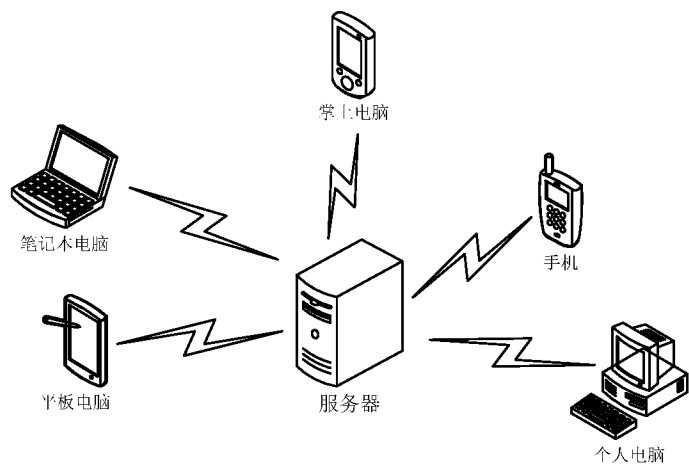


图 1

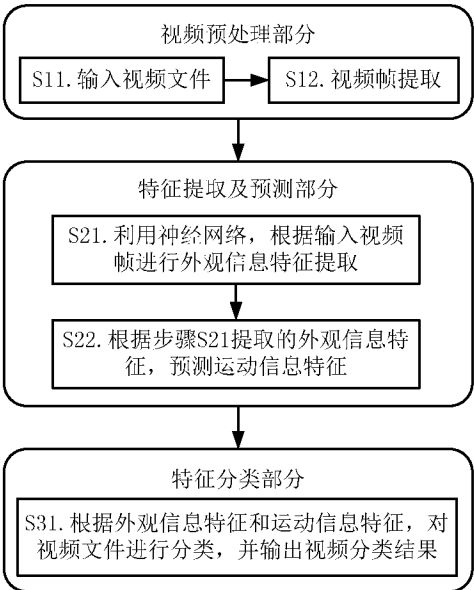


图 2

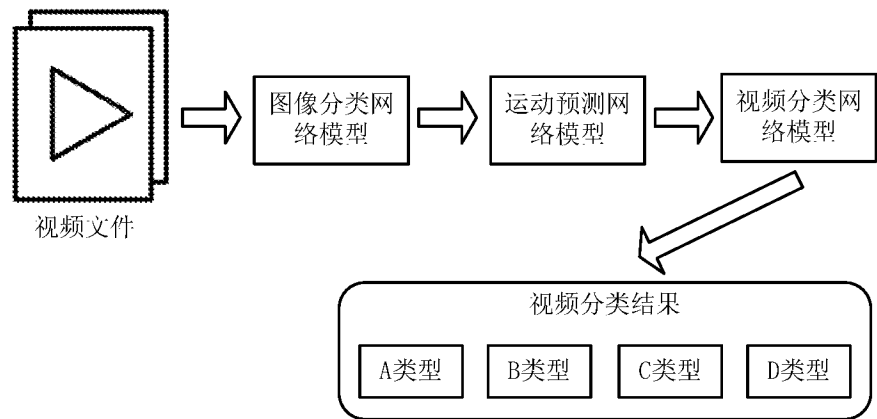


图 3

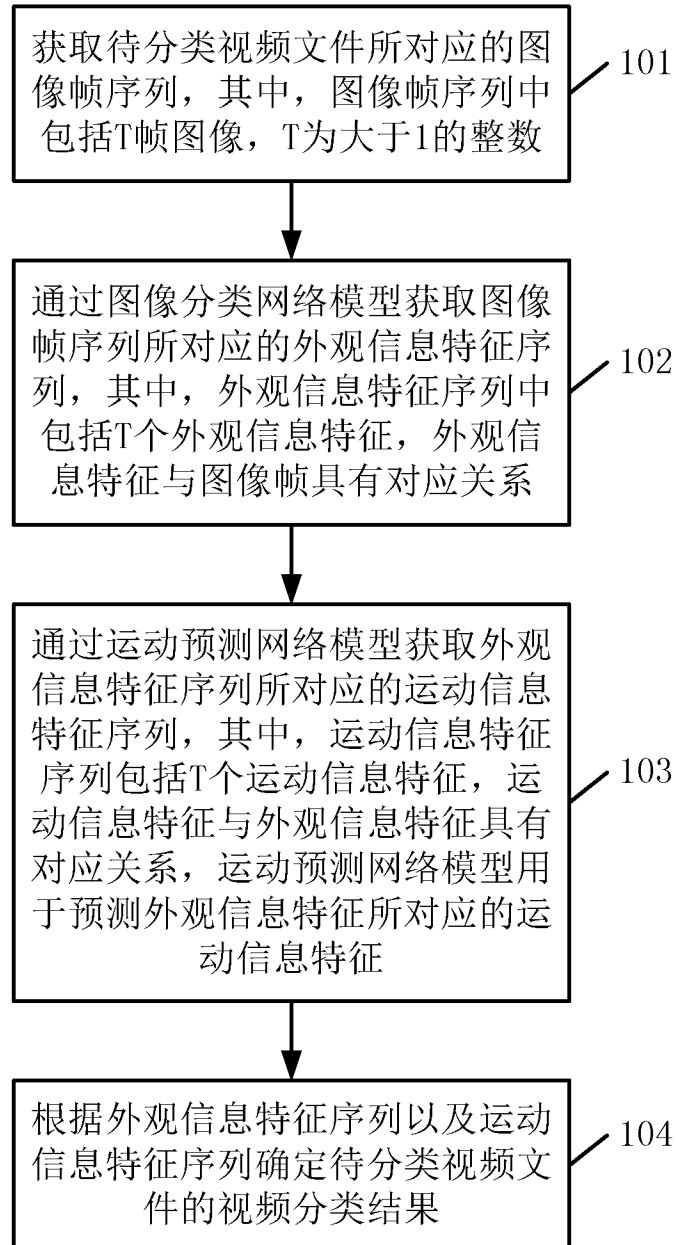


图 4

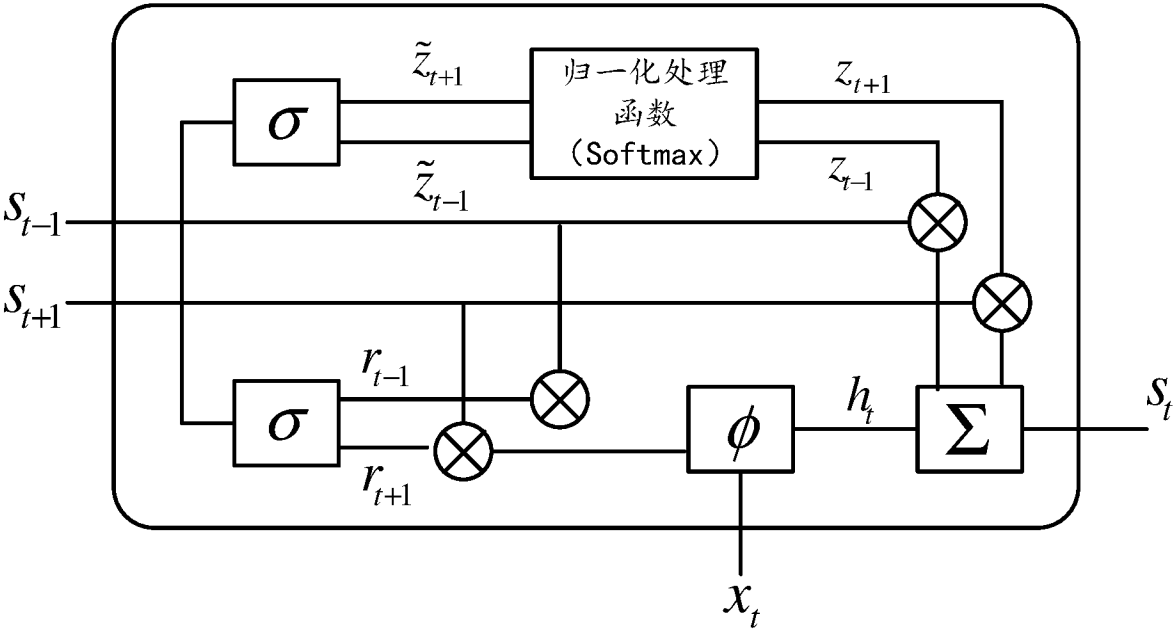


图 5

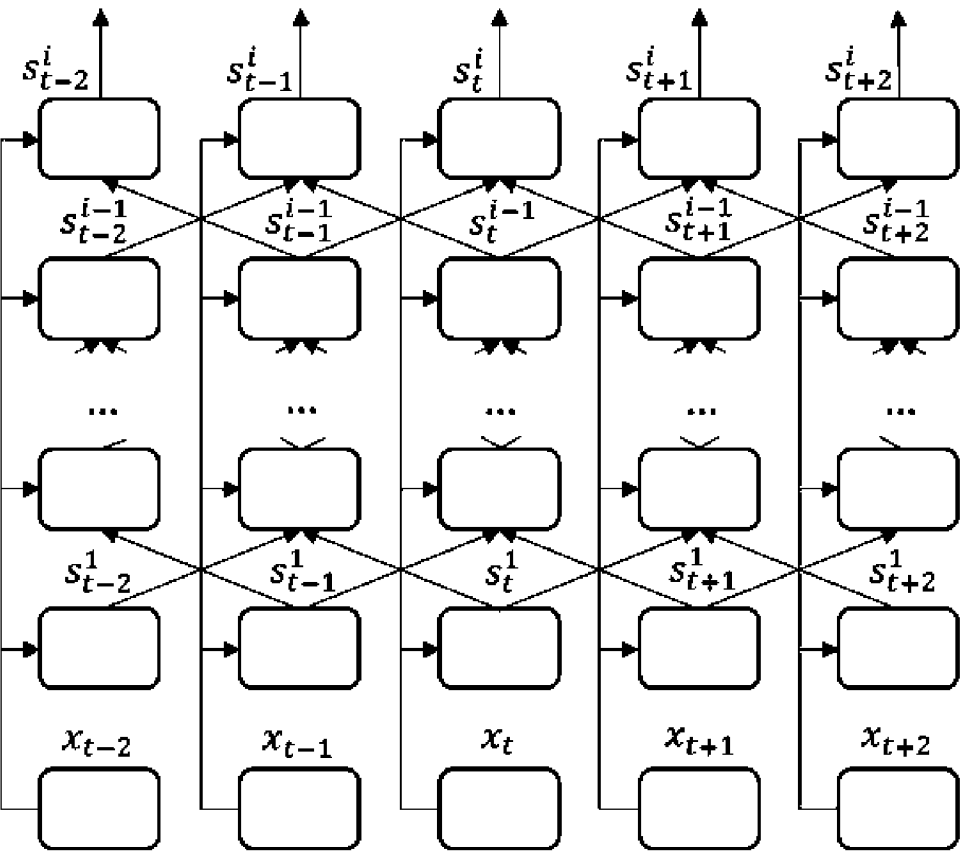


图 6

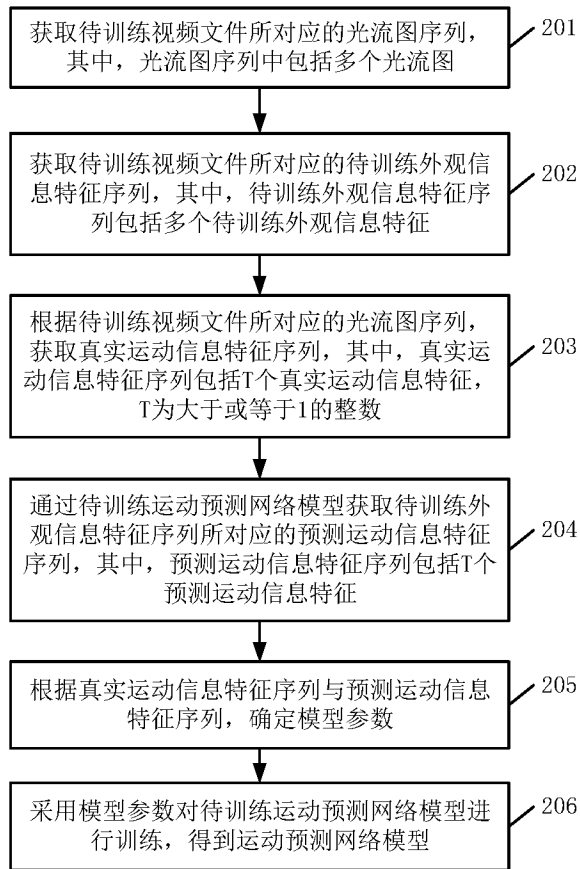


图 7

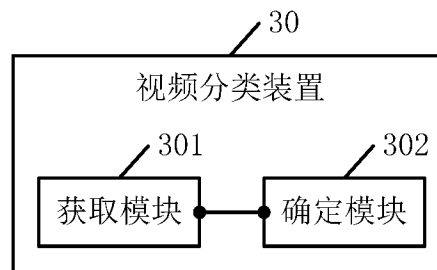


图 8

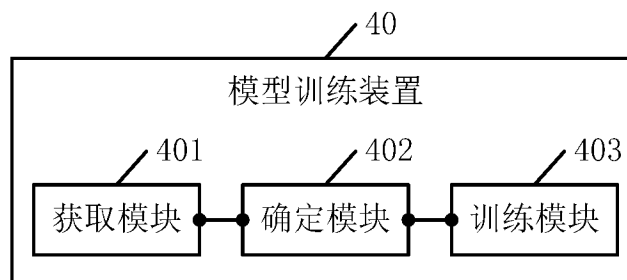


图 9

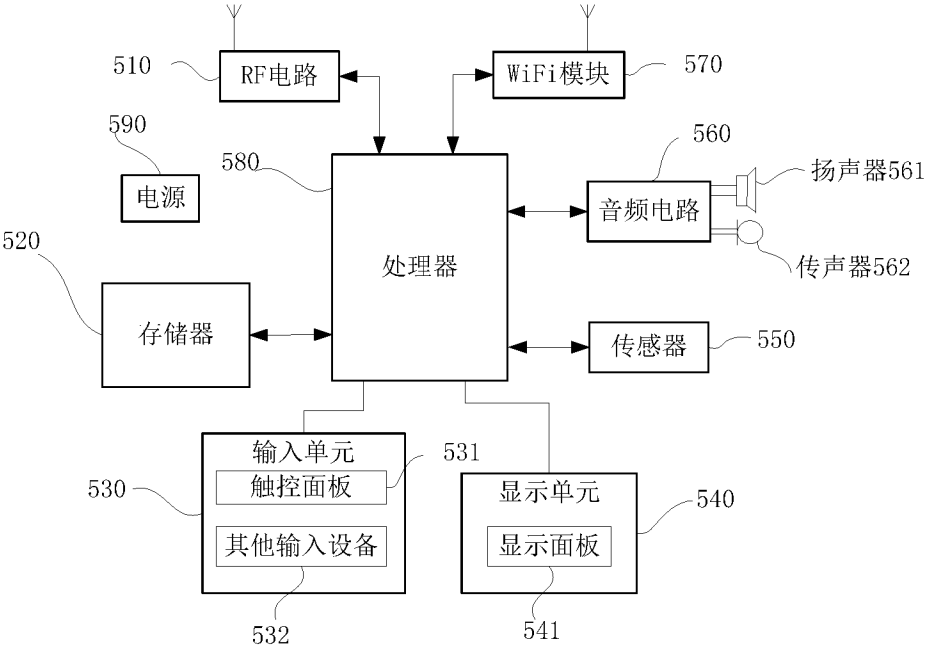


图 10

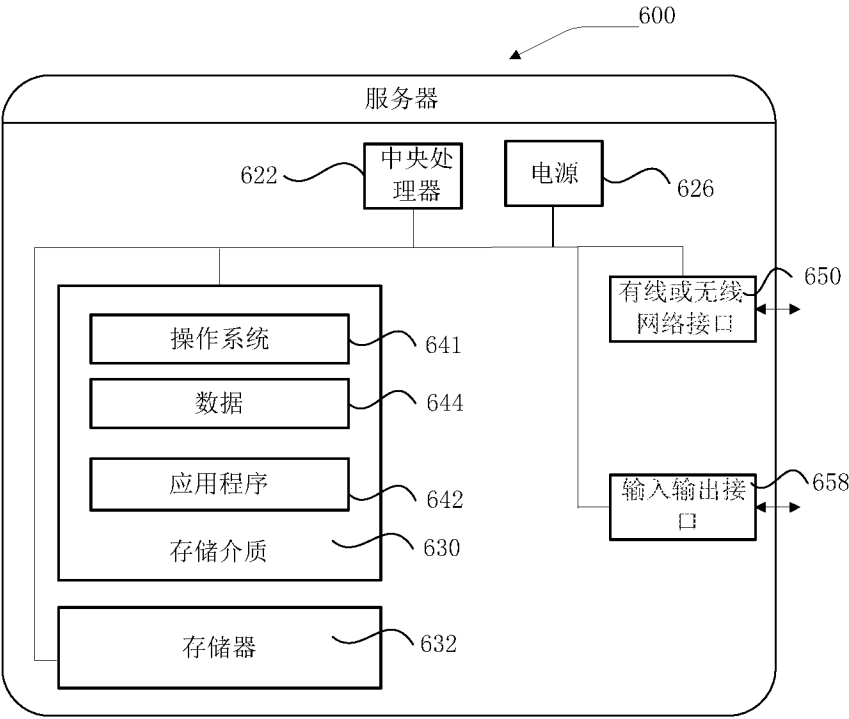


图 11

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2020/077809

**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)

CNABS, CNTXT, CNKI, VEN, USTXT, EPTXT, WOTXT: 视频, 分类, 图像, 颜色, 外观, 运动, 特征, 序列, video, classify, image, colour, appearance, motion, characteristic, sequence

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

| Category* | Citation of document, with indication, where appropriate, of the relevant passages                                                                   | Relevant to claim No. |
|-----------|------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------|
| PX        | CN 109919087 A (TENCENT TECHNOLOGY SHENZHEN CO., LTD.) 21 June 2019 (2019-06-21)<br>entire document                                                  | 1-23                  |
| A         | CN 105426883 A (SHENZHEN INSTITUTES OF ADVANCED TECHNOLOGY, CHINESE ACADEMY OF SCIENCES) 23 March 2016 (2016-03-23)<br>description, paragraphs 25-72 | 1-23                  |
| A         | CN 108875456 A (BEIJING KUANGSHI TECHNOLOGY CO., LTD. et al.) 23 November 2018 (2018-11-23)<br>description, paragraphs 5-26                          | 1-23                  |
| A         | CN 108960031 A (INSTITUTE OF SOFTWARE CHINESE ACADEMY OF SCIENCES) 07 December 2018 (2018-12-07)<br>entire document                                  | 1-23                  |
| A         | US 2017255831 A1 (XEROX CORPORATION) 07 September 2017 (2017-09-07)<br>entire document                                                               | 1-23                  |



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

04 June 2020

Date of mailing of the international search report

11 June 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/CN2020/077809**

| Patent document<br>cited in search report |            |    | Publication date<br>(day/month/year) | Patent family member(s) | Publication date<br>(day/month/year) |
|-------------------------------------------|------------|----|--------------------------------------|-------------------------|--------------------------------------|
| CN                                        | 109919087  | A  | 21 June 2019                         | None                    |                                      |
| CN                                        | 105426883  | A  | 23 March 2016                        | CN 105426883            | B 01 March 2019                      |
| CN                                        | 108875456  | A  | 23 November 2018                     | None                    |                                      |
| CN                                        | 108960031  | A  | 07 December 2018                     | None                    |                                      |
| US                                        | 2017255831 | A1 | 07 September 2017                    | US 9977968              | B2 22 May 2018                       |

## 国际检索报告

国际申请号

PCT/CN2020/077809

## 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))

CNABS, CNTXT, CNKI, VEN, USTXT, EPTXT, WOTXT: 视频, 分类, 图像, 颜色, 外观, 运动, 特征, 序列, video, classify, image, colour, appearance, motion, characteristic, sequence

## 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 109919087 A (腾讯科技深圳有限公司) 2019年 6月 21日 (2019 - 06 - 21)<br>全文             | 1-23    |
| A    | CN 105426883 A (中国科学院深圳先进技术研究院) 2016年 3月 23日 (2016 - 03 - 23)<br>说明书第25-72段 | 1-23    |
| A    | CN 108875456 A (北京旷视科技有限公司 等) 2018年 11月 23日 (2018 - 11 - 23)<br>说明书第5-26段   | 1-23    |
| A    | CN 108960031 A (中国科学院软件研究所) 2018年 12月 7日 (2018 - 12 - 07)<br>全文             | 1-23    |
| A    | US 2017255831 A1 (XEROX CORPORATION) 2017年 9月 7日 (2017 - 09 - 07)<br>全文     | 1-23    |

☐ 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

2020年 6月 4日

国际检索报告邮寄日期

2020年 6月 11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

授权官员

杨越松

电话号码 86-(20)-28950385

国际检索报告  
关于同族专利的信息

国际申请号

PCT/CN2020/077809

| 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          | 109919087  | A  | 2019年 6月 21日   | 无            |                 |
| CN          | 105426883  | A  | 2016年 3月 23日   | CN 105426883 | B 2019年 3月 1日   |
| CN          | 108875456  | A  | 2018年 11月 23日  | 无            |                 |
| CN          | 108960031  | A  | 2018年 12月 7日   | 无            |                 |
| US          | 2017255831 | A1 | 2017年 9月 7日    | US 9977968   | B2 2018年 5月 22日 |