



(21) 申请号 202011546849.0

(22) 申请日 2020.12.24

(65) 同一申请的已公布的文献号

申请公布号 CN 112614571 A

(43) 申请公布日 2021.04.06

(73) 专利权人 中国科学院深圳先进技术研究院

地址 518055 广东省深圳市南山区深圳大

学城学苑大道1068号

(72) 发明人 贾富仓 夏彤

(74) 专利代理机构 深圳市威世博知识产权代理

事务所(普通合伙) 44280

专利代理师 黎坚怡

(51) Int. Cl.

G16H 30/40 (2018.01)

G06V 20/40 (2022.01)

G06V 10/764 (2022.01)

G06V 10/774 (2022.01)

G06V 10/82 (2022.01)

G06N 3/0442 (2023.01)

G06N 3/08 (2023.01)

(56) 对比文件

CN 110263697 A, 2019.09.20

US 2020237452 A1, 2020.07.30

US 2021042580 A1, 2021.02.11

审查员 石惠敏

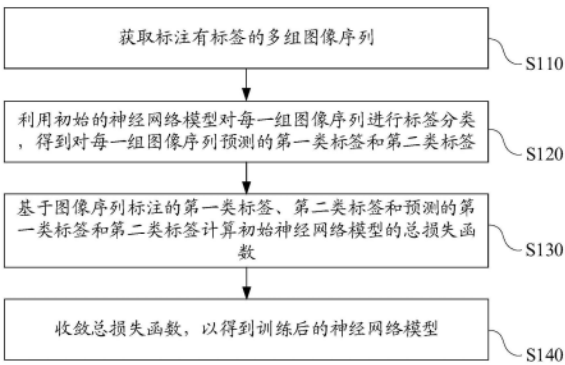
权利要求书2页 说明书9页 附图3页

(54) 发明名称

神经网络模型的训练方法、装置、图像分类方法和介质

(57) 摘要

本申请公开了一种神经网络模型的训练方法、装置、图像分类方法和介质。该方法包括如下步骤获取标注有标签的多组图像序列, 标签包括具有映射关系的第一类标签和第二类标签, 第一类标签的时间粒度大于第二类标签; 利用初始的神经网络模型对每一组图像序列进行标签分类, 得到对每一组图像序列预测的第一类标签和第二类标签; 基于图像序列标注的第一类标签、第二类标签和预测的第一类标签和第二类标签计算初始神经网络模型的总损失函数; 收敛总损失函数, 以得到训练后的神经网络模型。通过上述方式, 本申请能够训练出更精确的神经网络模型。



1. 一种神经网络模型的训练方法,所述神经网络模型至少包括主干网络、第一分支和第二分支,其中,所述第一分支包括长短时记忆网络和连接所述长短时记忆网络输出端的全连接层和映射函数,其特征在于,所述训练方法包括:

获取标注有标签的多组图像序列,所述标签包括具有映射关系的第一类标签和第二类标签,所述第一类标签的时间粒度大于所述第二类标签;

利用所述主干网络提取所述图像序列中每一图像的空间特征,得到表征所述每一图像的空间特征的空间特征向量;

利用所述长短时记忆网络对所述空间特征向量进行时间序列预测,输出包括时间序列的时空特征向量;

将所述时空特征向量输入所述全连接层进行第二类标签的分类,得到所述时空特征向量的预测的第二类标签;

将所述预测的第二类标签输入到所述映射函数,得到所述时空特征向量的预测的第一类标签;

以及基于所述空间特征向量标注的第二类标签和所述预测的第二类标签构建三元组样本,所述三元组样本包括固定样本、正样本和负样本;将所述三元组样本输入所述第二分支,以计算所述第二分支的三元组损失函数;

基于标注的第一类标签、所述标注的第二类标签和所述预测的第一类标签和所述预测的第二类标签计算所述第一分支的联合损失函数;

将所述三元组损失函数和所述联合损失函数加权求和,得到总损失函数;

收敛所述总损失函数,以得到训练后的神经网络模型;所述训练后的神经网络模型可用于对待预测图像进行两种时间粒度的分类预测。

2. 根据权利要求1所述的训练方法,其特征在于,所述第二类标签包括多个子标签,所述多个子标签包括子标签*i*和除子标签*i*以外的其他子标签,所述基于所述空间特征向量标注的第二类标签和所述预测的第二类标签构建三元组样本,包括:

记标签为所述子标签*i*的空间特征向量为所述固定样本;

比对所述空间特征向量标注的子标签和预测的子标签;

将标注为所述子标签*i*预测为所述其他子标签的空间特征向量作为所述正样本;

将标注为所述其他子标签预测为所述子标签*i*的空间特征向量作为所述负样本。

3. 根据权利要求1所述的训练方法,其特征在于,所述将所述三元组样本输入所述第二分支,以计算所述第二分支的三元组损失函数,包括:

分别计算所述正样本与所述固定样本之间的第一余弦距离、所述负样本与所述固定样本之间的第二余弦距离;

将所述第一余弦距离减去所述第二余弦距离得到所述三元组损失函数。

4. 根据权利要求3所述的训练方法,其特征在于,所述基于标注的第一类标签、所述标注的第二类标签和所述预测的第一类标签和所述预测的第二类标签计算所述第一分支的联合损失函数,包括:

计算所述标注的第一类标签和预测的第一类标签的第一交叉熵损失,以及计算所述标注的第二类标签和所述预测的第二类标签的第二交叉熵损失;

将所述第一交叉熵损失和所述第二交叉熵损失加权求和,得到所述联合损失函数。

5. 根据权利要求1所述的训练方法,其特征在于,所述基于标注的第一类标签、所述标注的第二类标签和所述预测的第一类标签和所述预测的第二类标签计算所述第一分支的联合损失函数,包括:

计算所述标注的第一类标签和所述预测的第一类标签的第一交叉熵损失,以及计算所述标注的第二类标签和所述预测的第二类标签的第二交叉熵损失;

将所述第一交叉熵损失和所述第二交叉熵损失加权求和,得到所述第一分支的联合损失函数。

6. 根据权利要求1-5任一所述的训练方法,其特征在于,所述第一类标签为手术流程的阶段标签,所述第二类标签为所述手术流程的步骤标签。

7. 一种图像分类方法,其特征在于,所述图像分类方法包括:

获取图像序列,所述图像序列包括至少两帧图像;

将所述图像序列输入神经网络模型,得到所述图像的标签;

其中,所述标签包括具有映射关系的第一类标签和第二类标签,所述第一类标签的时间粒度大于所述第二类标签;所述神经网络模型为权利要求1-6任一所述的神经网络模型的训练方法得到的训练后的神经网络模型。

8. 一种神经网络模型的训练装置,其特征在于,所述训练装置包括处理器和存储器,所述处理器耦接所述存储器,在工作时执行指令,以配合所述存储器实现如权利要求1至6任一项所述的神经网络模型的训练方法。

9. 一种计算机可读存储介质,其特征在于,所述计算机可读存储介质存储有计算机程序,所述计算机程序能够被处理器执行以实现如权利要求1至6中任一项所述的神经网络模型的训练方法或权利要求7所述的图像分类方法。

神经网络模型的训练方法、装置、图像分类方法和介质

技术领域

[0001] 本申请涉及深度学习领域,特别是涉及一种神经网络模型的训练方法、神经网络模型的训练装置、图像分类方法和计算机可读存储介质。

背景技术

[0002] 随着移动互联网与硬件处理器技术的不断发展,海量数据处理与计算能力不断提高,深度学习备受关注。卷积神经网络Convolutional Neural Network,CNN)和循环神经网络(Recurrent Neural Network,RNN)等经典神经网络模型被相继提出。

[0003] CNN通过其特有的权值共享机制输入是空间上的变化,即以图像为典型例子的空间域数据表现非常好,但对于样本序列出现的时间顺序上的变化,即时间域数据无法建模。RNN正是针对时域序列数据提出的,其特殊的网络结构使神经元的输出可以在下一个时间点作为输入直接作用到本身,实现网络的输出为该时刻的输入与历史所有时刻共同作用的结果,达到对序列建模的目的。长短时记忆神经网络(Long Short-term Memory Networks,LSTM)是一种RNN特殊的类型,可以学习长期依赖信息。常见的时间序列数据包括:语言模型、手写体识别、序列生成、机器翻译、语音、视频分析等。

[0004] 以视频分析为例,相关技术中通常使用CNN-LSTM模型范式对图像序列进行分类。然而,当图像序列帧间差异有限、存在干扰时,CNN-LSTM模型无法对抗图像序列的时空不一致的问题,进而影响模型精度。

发明内容

[0005] 本申请提供一种神经网络模型的训练方法、神经网络模型的训练装置、图像分类方法和计算机可读存储介质,以解决相关技术中神经网络模型对图像序列的识别存在时空不一致的技术问题。

[0006] 为解决上述技术问题,本申请提供一种神经网络模型的训练方法,该训练方法包括:获取标注有标签的多组图像序列,标签包括具有映射关系的第一类标签和第二类标签,第一类标签的时间粒度大于第二类标签;利用初始的神经网络模型对每一组图像序列进行标签分类,得到对每一组图像序列预测的第一类标签和第二类标签;基于图像序列标注的第一类标签、第二类标签和预测的第一类标签和第二类标签计算初始神经网络模型的总损失函数;收敛总损失函数,以得到训练后的神经网络模型。

[0007] 为解决上述技术问题,本申请提供一种图像分类方法,该分类方法包括:获取图像序列,图像序列包括至少两帧图像;将图像序列输入神经网络模型,得到图像的标签;其中,标签包括具有映射关系的第一类标签和第二类标签,第一类标签的时间粒度大于第二类标签;神经网络模型为上述的神经网络模型的训练方法得到的训练后的神经网络模型。

[0008] 为解决上述技术问题,本申请提供一种神经网络模型的训练装置。该训练装置包括处理器和存储器,处理器耦接存储器,在工作时执行指令,以配合存储器实现上述的神经网络模型的训练方法。

[0009] 为解决上述技术问题,本申请提供一种计算机可读存储介质。计算机可读存储介质存储有计算机程序,计算机程序能够被处理器执行以实现上述的神经网络模型的训练方法或上述的图像序列的分类方法。

[0010] 本申请通过利用神经网络模型对图像序列进行两种时间粒度上的分类预测,并根据预测的第一类标签和第二类标签与标注的第一类标签和第二类标签计算模型的总损失函数,基于两个时间粒度层面上对神经网络模型进行约束,以训练出能够改善对图像序列识别出现的时空不一致性的问题。

附图说明

[0011] 图1是本申请提供的神经网络模型的训练方法第一实施例的流程示意图;

[0012] 图2是本申请提供的利用初始的神经网络模型对每一组图像序列进行标签分类一实施方式的流程示意图;

[0013] 图3是本申请提供的神经网络模型的训练方法第二实施例的流程示意图;

[0014] 图4是本申请提供的一种图像分类方法一实施例的流程示意图;

[0015] 图5是本申请提供的神经网络模型的训练装置一实施例的结构示意图;

[0016] 图6是本申请提供的计算机可读存储介质一实施例的结构示意图。

具体实施方式

[0017] 为使本领域的技术人员更好地理解本发明的技术方案,下面结合附图和具体实施方式对本申请所提供的神经网络模型的训练方法、神经网络模型的训练装置、图像分类方法和计算机可读存储介质做进一步详细描述。

[0018] 目前,神经网络模型已广泛地应用于对图像序列的识别,例如用于对图像序列中的人体姿态、人脸、面部表情的识别,以及疾病图像、手术工作流的分析等等。本申请以将神经网络模型应用于手术工作流的识别分类进行说明。当然,本申请的神经网络模型的训练方法、图像分类方法还可以应用于其他类型图像序列的识别,本申请对此不做限制。

[0019] 微创手术以其创口小、恢复快、痛苦少的特点,已成为近三十年来各外科领域中的一种通用手术选择。微创手术实现了最大程度上的体贴病人,使其能尽早恢复到日常生活中,但手术人员进行微创手术前,需要进行长期训练以避免手术过程中出现不必要的失误和术后并发症。为了提高患者手术治疗的质量,现代手术室正朝着智能化方向发展。随着计算机视觉和机器人技术的发展,相关研究人员将其融合到现代微创手术中,用于辅助外科医生或手术机器人实施微创手术,从而形成一个新的领域——计算机辅助手术(Computer-Assisted Surgery, CAS)。计算机辅助手术包括术前疾病图像分析与诊断、术中手术导航和术后手术分析等方面的研究,从各方面提高微创手术的治疗效果。其中,针对手术视频进行工作流分析,是计算机辅助手术中最基本也至关重要的一项任务。

[0020] 自动手术工作流识别能在术中和术后为外科医生和全自主手术机器人提供重要信息。在术中,手术工作流能实时提供对当前手术步骤的指示,使得外科医生能清楚地意识到当前手术流程,从而避免错误操作和减少术后并发症的发生。对于缺乏经验的年轻医生,更是能进行实时地手术导航以标准化手术操作。在清晰地感知当前手术流程后,剩余手术时间能容易地计算出来,从而方便大型医院手术室的时间安排。在术后,自动识别手术工作

流能够帮助手术视频标注,自动报告生成,手术技能评估和教学等后续用途。因此,提出准确的自动手术 workflow 识别方法具有重大意义。

[0021] 由于微创手术镜头通常仅关注到患者局部且要求精细的手术操作,从而导致手术画面中呈现的场景在不同手术步骤间的类间差异十分有限。在这样的情况下,医生动作上的轻微改变和其他微小的干扰信息,都可能使得视频帧呈现出明显的与 workflow 识别任务无关的特征。一般的方法通常只依赖明显的特征,如手术器械、器官形变等来区分不同的手术步骤,因此忽略了呈现在分散在整个图片中的细粒度特征,如微小的切口和玻璃体的浑浊程度等。这些被忽略的细粒度特征与不同手术阶段有着密切的关联,在微创手术中,这些关键的细节信息也能用于对抗与任务无关的干扰。因此,手术 workflow 识别需要一种更细粒度的方法来应对这些模糊帧,并提取出综合了全局和局部的细粒度空间特征。

[0022] 手术视频的帧间时空不一致特性也是限制了目前的神经网络模型识别效果和泛化能力的原因之一。由于手术场景的细粒度特性,手术 workflow 识别任务对于相机视角的改变、主物体在场景中的位置和执行手术的姿势这些随时间变化的对象极其敏感。这对于不具有对抗时空不一致性的网络而言,进一步加重了类间差异小、类内差异大的问题。在空间上具有相似特征的图片可能来自于不同手术步骤,在时间上属于同一步骤的图片反而可能呈现较大的空间差异。

[0023] 为解决上述技术问题,本申请提供以下实施例。

[0024] 请参阅图1,图1是本申请提供的神经网络模型的训练方法第一实施例的流程示意图。本实施例包括如下步骤:

[0025] S110:获取标注有标签的多组图像序列。

[0026] 在将多组图像序列输入初始的神经网络模型之前,每组图像序列均已标注有标签。

[0027] 本实施例中,标签包括具有映射关系的第一类标签和第二类标签,第一类标签的时间粒度大于第二类标签。

[0028] 例如,第一类标签可以是手术 workflow 中的阶段,第二类标签可以是手术 workflow 中的步骤。阶段与步骤具有映射关系,每一阶段对应有至少两个步骤。

[0029] S120:利用初始的神经网络模型对每一组图像序列进行标签分类,得到对每一组图像序列预测的第一类标签和第二类标签。

[0030] 分别将图像序列输入到初始的神经网络模型中,以使得初始的神经网络模型对每组图像序列的标签进行预测分类,从而输出对每一组图像序列预测的第一类标签和第二类标签。

[0031] 其中,神经网络模型包括主干网络和第一分支,第一分支和第二分支连接主干网络的输出端,第一分支包括长短时记忆网络和连接长短时记忆网络输出端的全连接层和映射函数。利用主干网络实现对图像序列的空间特征的提取,利用第一分支实现对控件特征的时空融合。

[0032] 具体而言,请参阅图2,图2是本申请提供的利用初始的神经网络模型对每一组图像序列进行标签分类一实施方式的流程示意图。本实施方式包括如下步骤:

[0033] S121:利用主干网络提取图像序列中每一图像的空间特征,得到表征每一图像的空间特征的空间特征向量。

[0034] 其中,主干网络例如是分散注意力网络。注意力网络能够提取到更丰富更细粒度的全局和局部空间特征。

[0035] 注意力网络就是将注意力集中于局部关键信息的深度学习网络,可以分为两步:第一,通过扫描全局信息,发现局部有用信息;第二,增强有用信息并抑制冗余信息。换言之,注意力网络能够忽略无关的空间特征而关注重点的空间特征。

[0036] 本实施例的主干网络可以由4个50层分散注意力残差模块作为空间特征编码器。每个分散注意力残差模块由k个分组卷积组成,每个组内将输入分为r个分散通道,并进行分散注意力操作。将k个分组的结果连接后,再通过 1×1 卷积将通道数变回与输入相等的大小。主干网络部分以一个全局池化层结束,输出2048维的空间特征向量用于表征每帧图像中包含的空间特征。

[0037] S122:使用第一分支对空间特征向量进行时空融合,得到空间特征向量的预测的第一类标签和第二类标签。

[0038] 第一分支中包括有长短时记忆网络。属于循环神经网络的长短时记忆网络模型具有记忆性,能对图像序列中时间上的先后顺序建模,对于时序数据有较好的拟合效果。

[0039] 因此,将空间特征向量输入长短时记忆网络,利用长短时记忆网络对空间特征向量进行时间序列预测,能够输出包括时间序列的时空特征向量。

[0040] 进一步地,将时空特征向量输入全连接层进行第二类标签的分类,得到时空特征向量的预测的第二类标签。将预测的第二类标签输入到映射函数,得到时空特征向量的预测的第一类标签。

[0041] 其中,映射函数为表征步骤与阶段的映射关系的函数。以白内障手术 workflows 为例,其步骤可以包括切口、粘性剂注射、突破、水解剖、超声乳化、冲洗、前囊抛光、人工晶状体置入、眼内辅助器移除、伤口控制和消炎缝合等11个步骤,其阶段可包括前期准备、乳化、置入和缝合等4个阶段。映射函数可以表示如下:

[0042] 表1.白内障手术 workflows 的映射函数

[0043]	手术 workflow 类型	阶段	步骤
	白内障手术 workflow	前期准备	切口
			粘性剂注射
			突破
			水解剖
		乳化	超声乳化
			冲洗
		置入	前囊抛光
			人工晶状体置入
		缝合	眼内辅助器移除
			伤口控制
			消炎缝合

[0044] 由于手术 workflow 的阶段是沿着整个手术流程的时间顺序定义的,因此它们之间的状态转换规律通常是固定和规律的,这使得手术阶段层面的手术 workflow 识别相对容易实现,以及识别的结果相较于步骤具有更高的准确度。与之相反,手术步骤层面上的状态转移关系则相对复杂得多,从而难以捕获其中的时序关系。其次,时空不一致性问题主要出现在识别同一手术阶段内的不同手术步骤上。在一个特定的手术阶段中,手术场景呈现出更高的相似性,手术器械的使用也存在重叠,这进一步加大了神经网络模型区分混淆帧(即来自于不同手术步骤但在空间上具有相似特征的图像,在时间上属于同一步骤但呈现较大的空间特征差异的图像)的难度。因此,引入阶段层面的识别相较于直接对步骤进行识别更加可靠。

[0045] S130:基于图像序列标注的第一类标签、第二类标签和预测的第一类标签和第二类标签计算初始神经网络模型的总损失函数。

[0046] 具体而言,计算标注的第一类标签和预测的第一类标签的第一交叉熵损失,以及计算标注的第二类标签和预测的第二类标签的第二交叉熵损失。将第一交叉熵损失和第二交叉熵损失加权求和,得到总损失函数。公式表示如下:

$$[0047] \quad L_{co}(x_t; \theta_T) = \sum_{c \in C} (\lambda_1 L_{step}(\hat{p}_{t,c}) + \lambda_2 L_{phase}(\hat{p}_{t,c}))$$

[0048] 其中, $L_{co}(x_t; \theta_T)$ 表示总损失函数, $L_{step}(\hat{p}_{t,c})$ 表示第一交叉熵损失, $L_{phase}(\hat{p}_{t,c})$ 表示第二交叉熵损失。 x_t 是 t 时刻的输入图像序列, θ_T 表示第一分支的参数, λ_1 表示第一交叉熵损失的权重, λ_2 表示第二交叉熵损失的权重。

[0049] 将基于第一类标签的损失函数与基于第二类标签的损失函数进行加权求和,加入到神经网络模型整体损失函数中,能够互相修正和促进两个时间层面的预测结果。第一类标签时间的粒度大于第二类标签的时间粒度,并且第一类标签是由第二类标签映射得到

的,第一类标签的预测结果相较于第二类标签更加准确,从而对第二类标签的预测进行约束,能够加快第二交叉熵损失的收敛速度。

[0050] S140:收敛总损失函数,以得到训练后的神经网络模型。

[0051] 收敛总损失函数,使得总损失函数的值在收敛迭代的过程中越来越小。当迭代达到一定次数,或者总损失函数输出的值小于预期时,可以停止迭代,此时神经网络模型中各参数作为训练好的神经网络模型的参数,完成神经网络模型的训练。

[0052] 本实施例通过利用神经网络模型从两个时间颗粒层面上对标签进行分类,并且分别从两个时间颗粒层面上计算交叉熵损失,以构建总损失函数,使得两个时间颗粒层面上的预测结果能够相互修正和促进,从而改善时空不一致性问题。

[0053] 为了进一步提高模型对抗时空不一致性问题,神经网络模型还可以包括第二分支,第二分支连接主干网络的输出端。在第一分支输出对空间向量特征的预测的第一标签和第二标签后,还可以进一步基于第一分支的预测结果,利用第二分支计算三元组损失,从而引导着整个神经网络模型学到一个既具有细粒度识别能力又能对抗时空不一致性的时空特征表达。请参阅图3,图3是本申请提供的神经网络模型的训练方法第二实施例的流程示意图。本实施例是基于神经网络模型的训练方法第一实施例,相同的步骤在此不再赘述。本实施例包括如下步骤:

[0054] S310:获取标注有标签的多组图像序列。

[0055] S320:利用初始的神经网络模型对每一组图像序列进行标签分类,得到对每一组图像序列预测的第一类标签和第二类标签。

[0056] S330:基于空间特征向量标注的第二类标签和预测的第二类标签构建三元组样本。

[0057] 其中,三元组样本包括固定样本、正样本和负样本。由于第二类标签的时间粒度更小,拉近或拉远第二类型标签的空间特征能够更加有效地训练出高精度的神经网络模型,因此本实施例基于第二类标签构建三元组样本,并基于三元组样本加强和修正混淆帧在空间上的表达。

[0058] 基于标注的第二类标签和预测的第二类标签构建三元组样本的具体过程如下:

[0059] 第二类标签包括多个子标签,多个子标签包括子标签 i 和除子标签 i 以外的其他子标签。记标签为子标签 i 的空间特征向量为固定样本 x_i^a 。比对空间特征向量标注的子标签和预测的子标签。将标注为子标签 i 预测为其他子标签的空间特征向量作为正样本 x_i^p ,将标注为其他子标签预测为子标签 i 的空间特征向量作为负样本 x_i^n 。固定样本、正样本和负样本构成三元组样本。

[0060] S340:基于图像序列标注的第一类标签、第二类标签和预测的第一类标签和第二类标签计算初始神经网络模型的总损失函数。

[0061] 本实施例中,计算标注的第一类标签和预测的第一类标签的第一交叉熵损失,以及计算标注的第二类标签和预测的第二类标签的第二交叉熵损失。将第一交叉熵损失和第二交叉熵损失加权求和,得到联合损失函数。公式表示如下:

[0062]
$$L_{co}(x_i; \theta_T) = \sum_{c \in C} (\lambda_1 L_{step}(\hat{p}_{i,c}) + \lambda_2 L_{phase}(\hat{p}_{i,c}))$$

[0063] 其中, $L_{co}(x_t; \theta_T)$ 表示联合损失函数, $L_{step}(\hat{p}_{t,c})$ 表示第一交叉熵损失, $L_{phase}(\hat{p}_{t,c})$ 表示第二交叉熵损失。 x_t 是 t 时刻的输入图像序列, θ_T 表示第一分支的参数, λ_1 表示第一交叉熵损失的权重, λ_2 表示第二交叉熵损失的权重。 λ_1 的取值范围可以是 0.7~0.9, 具体例如是 0.7、0.8 或 0.9 等。相应地, λ_2 的取值范围可以是 0.1~0.3, 具体例如是 0.3、0.2 或 0.1 等。 λ_1 与 λ_2 之和等于 1。

[0064] 在第一分支输出对空间向量特征的预测的第一标签和第二标签后, 除计算第一分支的联合损失函数外, 还可以进一步基于第一分支的预测结果, 利用第二分支计算三元组函数, 以拉近近同一类别样本 (标注标签一致的图像序列) 的特征距离, 拉远不同类别样本 (标注标签不一致的图像序列) 的距离, 从而引导着整个神经网络模型学到一个既具有细粒度识别能力又能对抗时空不一致性的时空特征表达。

[0065] 将三元组样本输入第二分支, 以计算第二分支的三元组损失函数。

[0066] 针对所有的混淆帧, 第二分支都即时的加强和修正他们在特征空间上的表达, 以最终搜索到一个具有强表征能力的隐式空间 f , f 满足:

$$[0067] \quad \|f(x_i^a) - f(x_i^p)\| \ll \|f(x_i^a) - f(x_i^n)\|$$

[0068] 其中, $\|f(x_i^a) - f(x_i^p)\|$ 表示固定样本与正样本之间的余弦距离, $\|f(x_i^a) - f(x_i^n)\|$ 表示固定样本与负样本之间的余弦距离。

[0069] 分别计算正样本与固定样本之间的第一余弦距离、负样本与固定样本之间的第二余弦距离。将第一余弦距离减去第二余弦距离得到三元组损失函数。公式表示如下:

$$[0070] \quad L_{con}(x_i; \theta_C) = distance(x_i^a, x_i^p) - distance(x_i^a, x_i^n)$$

[0071] 其中, $L_{con}(x_t; \theta_C)$ 表示三元组损失函数, $distance(x_i^a, x_i^p)$ 表示第一余弦距离, $distance(x_i^a, x_i^n)$ 表示第二余弦距离, θ_C 表示第二分支的参数。

[0072] 将三元组损失函数和联合损失函数加权求和, 得到总损失函数。公式表示如下:

$$[0073] \quad L(x_t; \theta_T; \theta_C) = \lambda_T L_{co}(x_t; \theta_T) + \lambda_C L_{con}(x_t; \theta_C)$$

[0074] 其中, $L(x_t; \theta_T; \theta_C)$ 表示总损失函数, λ_T 表示第一分支的权重, λ_C 表示第二分支的权重。 λ_T 的取值范围可以是 0.5~0.7, 具体例如是 0.5、0.6 或 0.7 等。相应地, λ_C 的取值范围可以是 0.3~0.5, 具体例如是 0.5、0.4 或 0.3 等。 λ_T 与 λ_C 之和等于 1。

[0075] S350: 收敛总损失函数, 以得到训练后的神经网络模型。

[0076] 本实施例通过余弦距离来度量正样本、负样本和固定样本之间的距离, 并把它加之神经网络模型整体损失函数中, 从而使整个网络被迫学习一种特征表达, 拉近正样本与固定样本之间的空间距离, 拉远负样本与固定样本之间的空间距离, 从而找到区分不同类别样本的最关键信息。在这样的约束条件下, 得以获得对一段连续帧序列的时空融合特征, 进而训练出对时空特征识别精度更高的神经网络模型, 解决时空不一致性问题。

[0077] 请参阅图4, 图4是本申请提供的一种图像分类方法一实施例的流程示意图。本实施例包括如下步骤:

[0078] S410: 获取图像序列, 图像序列包括至少两帧图像。

[0079] 其中, 图像序列例如是视频, 图像序列由时间连续的至少两帧图像组成。

[0080] S420: 将图像序列输入神经网络模型, 得到图像的标签。

[0081] 利用上述的神经网络模型的训练方法实施例训练得到的训练后的神经网络模型对图像序列进行分类,得到图像的标签。

[0082] 其中,标签包括具有映射关系的第一类标签和第二类标签,第一类标签的时间粒度大于第二类标签。

[0083] 这两种不同层面的标签描述中蕴含了一些值得利用的特性,基于上述方法得到的训练后的神经网络模型能够更加准确的预测的第一类标签和第二类标签,进而使用户更加清楚地了解到当前图像处于整个时间流的哪个流程。

[0084] 上述神经网络模型的训练方法的第一实施例由神经网络模型的训练装置实现,因而本申请还提出神经网络模型的训练装置,请参阅图5,图5是本申请提供的神经网络模型的训练装置一实施例的结构示意图。本实施例神经网络模型的训练装置500可以包括相互连接的处理器501和存储器502。其中,存储器502用于存储初始的神经网络模型和图像序列,其中,图像序列标注有标签,标签包括具有映射关系的第一类标签和第二类标签,第一类标签的时间粒度大于第二类标签。处理器501用于从存储器502获取标注有标签的多组图像序列,利用初始的神经网络模型对每一组图像序列进行标签分类,得到对每一组图像序列预测的第一类标签和第二类标签;基于图像序列标注的第一类标签、第二类标签和预测的第一类标签和第二类标签计算初始神经网络模型的总损失函数;收敛总损失函数,以得到训练后的神经网络模型。

[0085] 其中,处理器501可以是一种集成电路芯片,具有信号的处理能力。处理器501还可以是通用处理器、数字信号处理器(DSP)、专用集成电路(ASIC)、现场可编程门阵列(FPGA)或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件。通用处理器可以是微处理器或者该处理器也可以是任何常规的处理器等。

[0086] 对于上述实施例的方法,其可以计算机程序的形式存在,因而本申请提出一种计算机可读存储介质,请参阅图6,图6是本申请提供的计算机可读存储介质一实施例的结构示意图。本实施例计算机可读存储介质600中存储有计算机程序601,其可被执行以实现上述实施例中的方法。

[0087] 本实施例计算机可读存储介质600可以是U盘、移动硬盘、只读存储器(ROM,Read-Only Memory)、随机存取存储器(RAM,Random Access Memory)、磁碟或者光盘等可以存储程序指令的介质,或者也可以为存储有该程序指令的服务器,该服务器可将存储的程序指令发送给其他设备运行,或者也可以自运行该存储的程序指令。

[0088] 在本申请所提供的几个实施例中,应该理解到,所揭露的方法和装置,可以通过其它的方式实现。例如,以上所描述的装置实施方式仅仅是示意性的,例如,模块或单元的划分,仅仅为一种逻辑功能划分,实际实现时可以有另外的划分方式,例如多个单元或组件可以结合或者可以集成到另一个系统,或一些特征可以忽略,或不执行。另一点,所显示或讨论的相互之间的耦合或直接耦合或通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。

[0089] 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方式方案的目的。

[0090] 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。

[0091] 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)或处理器(processor)执行本申请各个实施方式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(ROM, Read-Only Memory)、随机存取存储器(RAM, Random Access Memory)、磁碟或者光盘等各种可以存储程序代码的介质。

[0092] 以上所述仅为本申请的实施方式,并非因此限制本申请的专利范围,凡是利用本申请说明书及附图内容所作的等效结构或等效流程变换,或直接或间接运用在其他相关的技术领域,均同理包括在本申请的专利保护范围内。

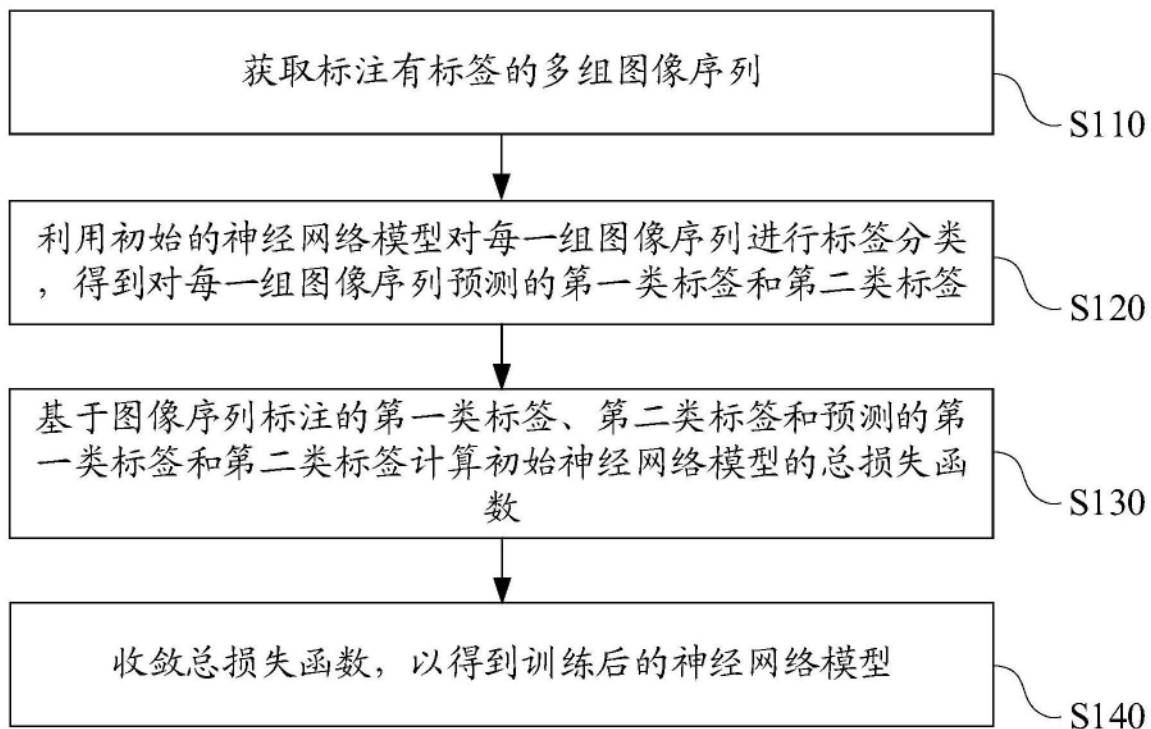


图1

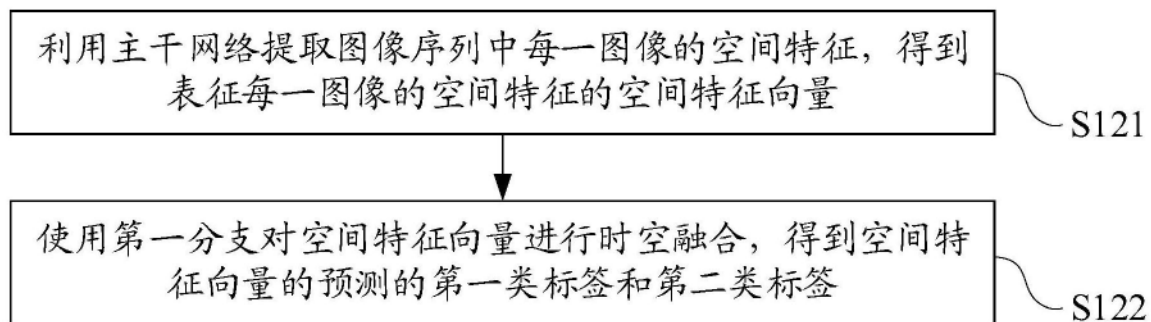


图2

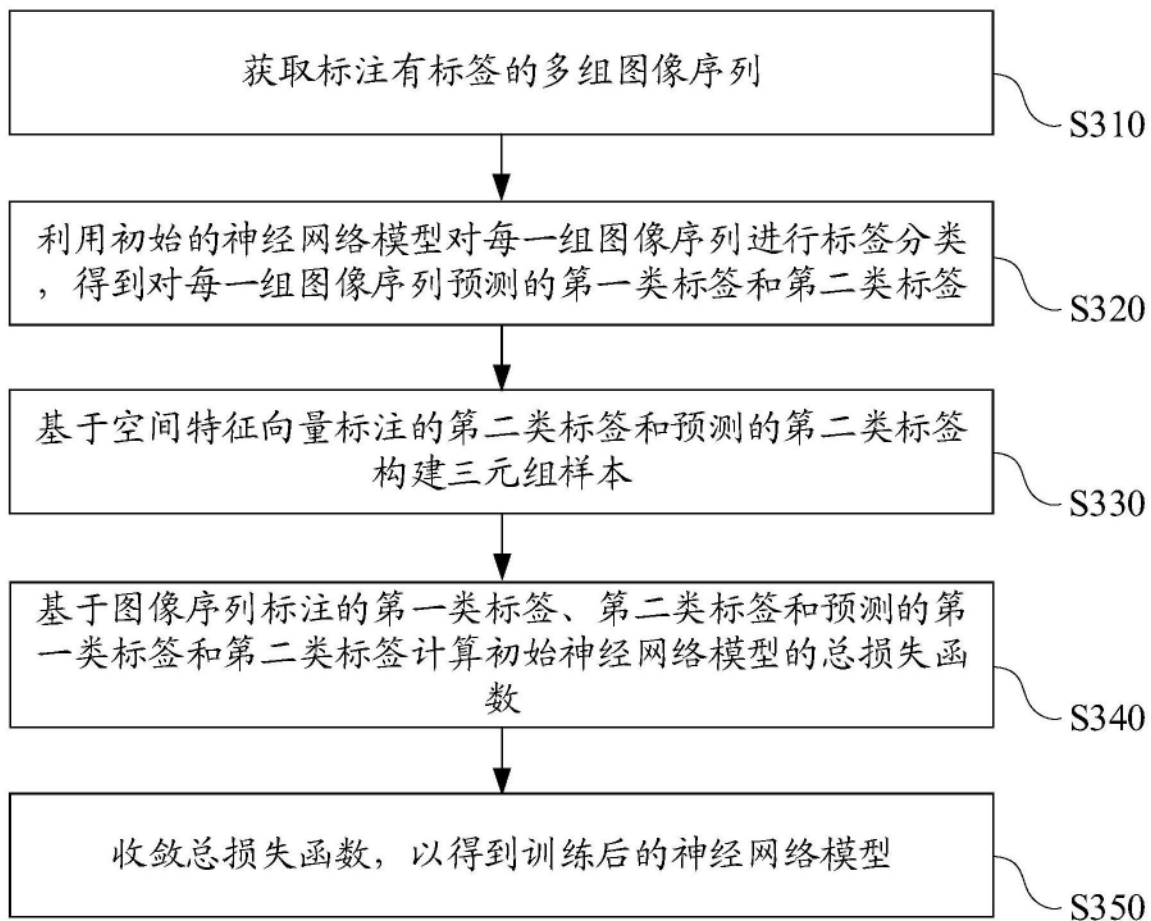


图3

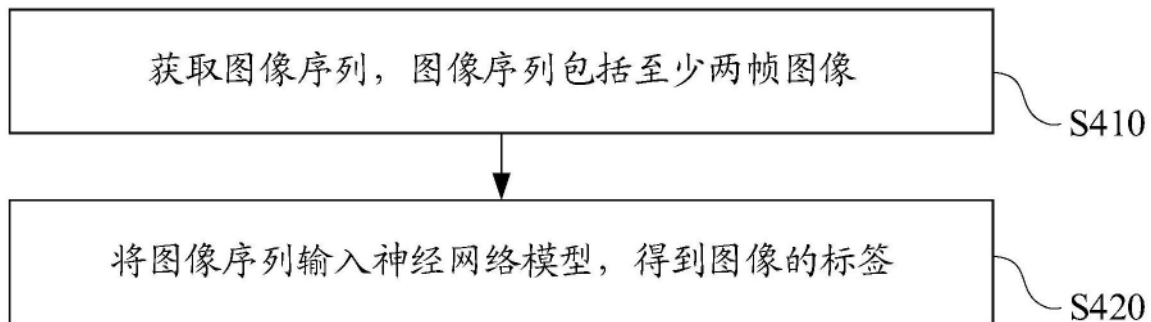


图4

500

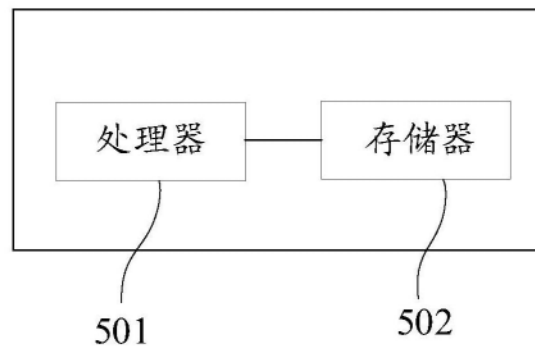


图5

600

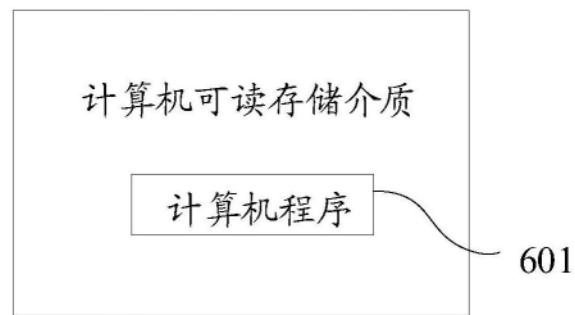


图6