



(12)发明专利申请

(10)申请公布号 CN 110427899 A
(43)申请公布日 2019. 11. 08

(21)申请号 201910726347.7

(22)申请日 2019.08.07

(71)申请人 网易(杭州)网络有限公司
地址 310052 浙江省杭州市滨江区网商路
599号网易大厦

(72)发明人 宋新慧 袁焱

(74)专利代理机构 北京律智知识产权代理有限
公司 11438
代理人 王辉 阚梓瑄

(51) Int. Cl.
G06K 9/00(2006.01)
G06N 3/04(2006.01)
G06N 3/08(2006.01)
G06T 7/11(2017.01)
G06T 7/136(2017.01)

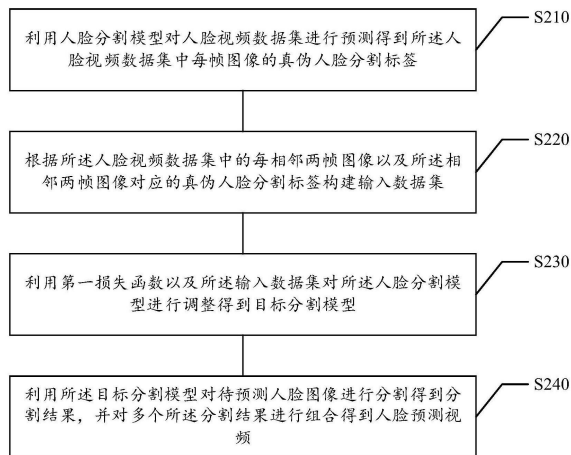
权利要求书2页 说明书12页 附图4页

(54)发明名称

基于人脸分割的视频预测方法及装置、介
质、电子设备

(57)摘要

本发明实施例是关于一种基于人脸分割的视频预测方法及装置、介质、电子设备,涉及图像处理技术领域,该方法包括:利用人脸分割模型对人脸视频数据集进行预测得到所述人脸视频数据集中每帧图像的真伪人脸分割标签;根据所述人脸视频数据集中的每相邻两帧图像以及所述相邻两帧图像对应的真伪人脸分割标签构建输入数据集;利用第一损失函数以及所述输入数据集对所述人脸分割模型进行调整得到目标分割模型;利用所述目标分割模型对待预测人脸图像进行分割得到分割结果,并对多个所述分割结果进行组合得到人脸预测视频。本发明实施例提高了人脸预测视频的稳定性。



1. 一种基于人脸分割的视频预测方法,其特征在于,包括:

利用人脸分割模型对人脸视频数据集进行预测得到所述人脸视频数据集中每帧图像的真伪人脸分割标签;

根据所述人脸视频数据集中的每相邻两帧图像以及所述相邻两帧图像对应的真伪人脸分割标签构建输入数据集;

利用第一损失函数以及所述输入数据集对所述人脸分割模型进行调整得到目标分割模型;

利用所述目标分割模型对待预测人脸图像进行分割得到分割结果,并对多个所述分割结果进行组合得到人脸预测视频。

2. 根据权利要求1所述的基于人脸分割的视频预测方法,其特征在于,在利用人脸分割模型对人脸视频数据集进行预测得到真伪人脸分割标签之前,所述基于人脸分割的视频预测方法还包括:

利用人脸图像以及人脸分割标签对深度卷积神经网络进行训练得到所述人脸分割模型。

3. 根据权利要求2所述的基于人脸分割的视频预测方法,其特征在于,利用人脸图像以及人脸分割标签对深度卷积神经网络进行训练得到所述人脸分割模型包括:

根据所述深度卷积神经网络的当前网络参数值提取所述人脸图像的第一特征向量以及所述人脸分割标签的第二特征向量;

根据所述第一特征向量以及第二特征向量计算损失函数值,并判断所述损失函数值是否小于预设阈值;

如果所述损失函数值小于所述预设阈值,则确定所述当前网络参数值对应的深度卷积神经网络为所述人脸分割模型。

4. 根据权利要求3所述的基于人脸分割的视频预测方法,其特征在于,根据所述第一特征向量以及第二特征向量计算损失函数值包括:

配置第二损失函数,并根据所述第一特征向量、第二特征向量以及所述第二损失函数计算所述损失函数值:

其中,所述第二损失函数包括:

$$J = -\sum_{c=1}^C y_c \log(p_c);$$

其中,J为第二损失函数;C为所述人脸分割标签的数量; y_c 为第一特征向量; p_c 为第二特征向量。

5. 根据权利要求4所述的基于人脸分割的视频预测方法,其特征在于,所述基于人脸分割的视频预测方法还包括:

如果所述损失函数值大于所述预设阈值,则利用梯度回传算法调整所述深度卷积神经网络的网络参数值的大小。

6. 根据权利要求4所述的基于人脸分割的视频预测方法,其特征在于,所述第一损失函数包括:

$$L=L_1(f_1-f_2)+w*(J_1+J_2);$$

其中,L为第一损失函数; f_1 为所述每相邻两帧图像中的第一帧图像在所述深度卷积神

神经网络中的特征; f_2 为所述每相邻两帧图像中的第二帧图像在所述深度卷积神经网络中的特征; L_1 为 f_1 与 f_2 之间的距离; w 为权重; J_1 为所述每相邻两帧图像中的第一帧图像的损失函数值; J_2 为所述每相邻两帧图像中的第二帧图像的损失函数值。

7. 根据权利要求2-6任一项所述的基于人脸分割的视频预测方法,其特征在于,所述人脸分割标签包括人脸、左眼眉、右眼眉、左眼、右眼、鼻子、上嘴唇、牙齿、下嘴唇、头发以及背景中的多种。

8. 一种基于人脸分割的视频预测装置,其特征在于,包括:

预测模块,用于利用人脸分割模型对人脸视频数据集进行预测得到所述人脸视频数据集中每帧图像的真伪人脸分割标签;

数据集构建模块,用于根据所述人脸视频数据集中的每相邻两帧图像以及所述相邻两帧图像对应的真伪人脸分割标签构建输入数据集;

调整模块,用于利用第一损失函数以及所述输入数据集对所述人脸分割模型进行调整得到目标分割模型;

分割模块,用于利用所述目标分割模型对待预测人脸图像进行分割得到分割结果,并对多个所述分割结果进行组合得到人脸预测视频。

9. 一种计算机可读存储介质,其上存储有计算机程序,其特征在于,所述计算机程序被处理器执行时实现权利要求1-7任一项所述的基于人脸分割的视频预测方法。

10. 一种电子设备,其特征在于,包括:

处理器;以及

存储器,用于存储所述处理器的可执行指令;

其中,所述处理器配置为经由执行所述可执行指令来执行权利要求1-7任一项所述的基于人脸分割的视频预测方法。

基于人脸分割的视频预测方法及装置、介质、电子设备

技术领域

[0001] 本发明实施例涉及图像处理技术领域,具体而言,涉及一种基于人脸分割的视频预测方法、基于人脸分割的视频预测装置、计算机可读存储介质以及电子设备。

背景技术

[0002] 视频稳像是计算机视觉中一个重要的问题,通过视频稳像可以减少整体位移,给视频观感带来很好的体验。当拍摄源不固定的情况下,很容易造成视频比较抖动,抖动的视频给人很不好的体验。其中,视频发生抖动最显著的特征是帧与帧之间发生整体位移。目前,一些视频抖动检测的方法包括块匹配法、特征点匹配法、灰度投影法以及利用深度学习预测光流的方法等等。

[0003] 具体的,以利用深度学习预测光流的方法来说,由于需要收集大量的视频数据进行训练,因此较为耗时。并且,深度学习预测带来的误差就会影响预测视频结果的稳定性。比如表情迁移,一般的数据都是图像,输入是人脸图像,输出是具有表情语义的连续标签;当训练好模型之后,应用于预测视频,由于深度学习模型会存在一定的误差,这样就引入了不稳定性因素,导致视频预测结果不稳定。

[0004] 因此,需要提供一种新的基于人脸分割的视频预测方法及装置。

[0005] 需要说明的是,在上述背景技术部分发明的信息仅用于加强对本发明的背景的理解,因此可以包括不构成对本领域普通技术人员已知的现有技术的信息。

发明内容

[0006] 本发明的目的在于提供一种基于人脸分割的视频预测方法、基于人脸分割的视频预测装置、计算机可读存储介质以及电子设备,进而至少在一定程度上克服由于相关技术的限制和缺陷而导致的视频预测结果不稳定的问题。

[0007] 根据本公开的一个方面,提供一种基于人脸分割的视频预测方法,包括:

[0008] 利用人脸分割模型对人脸视频数据集进行预测得到所述人脸视频数据集中每帧图像的真伪人脸分割标签;

[0009] 根据所述人脸视频数据集中的每相邻两帧图像以及所述相邻两帧图像对应的真伪人脸分割标签构建输入数据集;

[0010] 利用第一损失函数以及所述输入数据集对所述人脸分割模型进行调整得到目标分割模型;

[0011] 利用所述目标分割模型对待预测人脸图像进行分割得到分割结果,并对多个所述分割结果进行组合得到人脸预测视频。

[0012] 在本公开的一种示例性实施例中,在利用人脸分割模型对人脸视频数据集进行预测得到真伪人脸分割标签之前,所述基于人脸分割的视频预测方法还包括:

[0013] 利用人脸图像以及人脸分割标签对深度卷积神经网络进行训练得到所述人脸分割模型。

[0014] 在本公开的一种示例性实施例中,利用人脸图像以及人脸分割标签对深度卷积神经网络进行训练得到所述人脸分割模型包括:

[0015] 根据所述深度卷积神经网络的当前网络参数值提取所述人脸图像的第一特征向量以及所述人脸分割标签的第二特征向量;

[0016] 根据所述第一特征向量以及第二特征向量计算损失函数值,并判断所述损失函数值是否小于预设阈值;

[0017] 如果所述损失函数值小于所述预设阈值,则确定所述当前网络参数值对应的深度卷积神经网络为所述人脸分割模型。

[0018] 在本公开的一种示例性实施例中,根据所述第一特征向量以及第二特征向量计算损失函数值包括:

[0019] 配置第二损失函数,并根据所述第一特征向量、第二特征向量以及所述第二损失函数计算所述损失函数值;

[0020] 其中,所述第二损失函数包括:

$$[0021] \quad J = -\sum_{c=1}^C y_c \log(p_c);$$

[0022] 其中,J为第二损失函数;C为所述人脸分割标签的数量; y_c 为第一特征向量; p_c 为第二特征向量。

[0023] 在本公开的一种示例性实施例中,所述基于人脸分割的视频预测方法还包括:

[0024] 如果所述损失函数值大于所述预设阈值,则利用梯度回传算法调整所述卷积神经网络的网络参数值的大小。

[0025] 在本公开的一种示例性实施例中,所述第一损失函数包括:

$$[0026] \quad L=L_1(f_1-f_2)+w*(J_1+J_2);$$

[0027] 其中,L为第一损失函数; f_1 为所述每相邻两帧图像中的第一帧图像在所述深度卷积神经网络中的特征; f_2 为所述每相邻两帧图像中的第二帧图像在所述深度卷积神经网络中的特征; L_1 为 f_1 与 f_2 之间的距离; w 为权重; J_1 为所述每相邻两帧图像中的第一帧图像的损失函数值; J_2 为所述每相邻两帧图像中的第二帧图像的损失函数值。

[0028] 在本公开的一种示例性实施例中,所述人脸分割标签包括人脸、左眼眉、右眼眉、左眼、右眼、鼻子、上嘴唇、牙齿、下嘴唇、头发以及背景中的多种。

[0029] 根据本公开的一个方面,提供一种基于人脸分割的视频预测装置,包括:

[0030] 预测模块,用于利用人脸分割模型对人脸视频数据集进行预测得到所述人脸视频数据集中每帧图像的真伪人脸分割标签;

[0031] 数据集构建模块,用于根据所述人脸视频数据集中的每相邻两帧图像以及所述相邻两帧图像对应的真伪人脸分割标签构建输入数据集;

[0032] 调整模块,用于利用第一损失函数以及所述输入数据集对所述人脸分割模型进行调整得到目标分割模型;

[0033] 分割模块,用于利用所述目标分割模型对待预测人脸图像进行分割得到分割结果,并对多个所述分割结果进行组合得到人脸预测视频。

[0034] 根据本公开的一个方面,提供一种计算机可读存储介质,其上存储有计算机程序,所述计算机程序被处理器执行时实现上述任意一项所述的基于人脸分割的视频预测方法。

[0035] 根据本公开的一个方面,提供一种电子设备,包括:

[0036] 处理器;以及

[0037] 存储器,用于存储所述处理器的可执行指令;

[0038] 其中,所述处理器配置为经由执行所述可执行指令来执行上述任意一项所述的基于人脸分割的视频预测方法。

[0039] 本发明实施例一种基于人脸分割的视频预测方法及装置,一方面,通过利用人脸分割模型对人脸视频数据集进行预测得到人脸视频数据集中每帧图像的真伪人脸分割标签;并根据人脸视频数据集中的每相邻两帧图像以及相邻两帧图像对应的真伪人脸分割标签构建输入数据集;再利用第一损失函数以及输入数据集对人脸分割模型进行调整得到目标分割模型;最后利用目标分割模型对待预测人脸图像进行分割得到分割结果,并对多个分割结果进行组合得到人脸预测视频;解决了现有技术中由于深度学习模型会存在一定的误差,进而导致视频预测结果不稳定的问题,提高了人脸预测视频的稳定性;另一方面,通过利用人脸分割模型对人脸视频数据集进行预测得到人脸视频数据集中每帧图像的真伪人脸分割标签;并根据人脸视频数据集中的每相邻两帧图像以及相邻两帧图像对应的真伪人脸分割标签构建输入数据集;再利用第一损失函数以及输入数据集对人脸分割模型进行调整得到目标分割模型,解决了现有技术中需要收集大量的视频数据进行训练进而导致较为耗时的问题,提高了目标分割模型的调整速度;再一方面,通过利用目标分割模型对待预测人脸图像进行分割得到分割结果,并对多个分割结果进行组合得到人脸预测视频,提高了人脸预测视频的准确性。

[0040] 应当理解的是,以上的一般描述和后文的细节描述仅是示例性和解释性的,并不能限制本发明。

附图说明

[0041] 此处的附图被并入说明书中并构成本说明书的一部分,示出了符合本发明的实施例,并与说明书一起用于解释本发明的原理。显而易见地,下面描述中的附图仅仅是本发明的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他的附图。

[0042] 图1示意性示出根据本公开示例实施例的一种基于视频到视频合成方法的应用场景示例图。

[0043] 图2示意性示出根据本公开示例实施例的一种基于人脸分割的视频预测方法的流程图。

[0044] 图3示意性示出根据本公开示例实施例的一种基于人脸分割的食品预测方法的应用场景示例图。

[0045] 图4示意性示出根据本公开示例实施例的一种利用人脸图像以及人脸分割标签对深度卷积神经网络进行训练得到所述人脸分割模型的方法流程图。

[0046] 图5示意性示出根据本公开示例实施例的另一种基于人脸分割的视频预测方法的流程图。

[0047] 图6示意性示出根据本公开示例实施例的一种基于人脸分割的视频预测装置的框图。

[0048] 图7示意性示出根据本公开示例实施例的一种用于实现上述基于人脸分割的视频预测方法的电子设备。

具体实施方式

[0049] 现在将参考附图更全面地描述示例实施方式。然而，示例实施方式能够以多种形式实施，且不应被理解为限于在此阐述的范例；相反，提供这些实施方式使得本发明将更加全面和完整，并将示例实施方式的构思全面地传达给本领域的技术人员。所描述的特征、结构或特性可以以任何合适的方式结合在一个或更多实施方式中。在下面的描述中，提供许多具体细节从而给出对本发明的实施方式的充分理解。然而，本领域技术人员将意识到，可以实践本发明的技术方案而省略所述特定细节中的一个或更多，或者可以采用其它的方法、组元、装置、步骤等。在其它情况下，不详细示出或描述公知技术方案以避免喧宾夺主而使得本发明的各方面变得模糊。

[0050] 此外，附图仅为本发明的示意性图解，并非一定是按比例绘制。图中相同的附图标记表示相同或类似的部分，因而将省略对它们的重复描述。附图中所示的一些方框图是功能实体，不一定必须与物理或逻辑上独立的实体相对应。可以采用软件形式来实现这些功能实体，或在一个或多个硬件模块或集成电路中实现这些功能实体，或在不同网络和/或处理器装置和/或微控制器装置中实现这些功能实体。

[0051] 首先，对本公开示例实施例中涉及的专有名词做以下解释。

[0052] 视频抖动：视频稳像是一种基于视频系列的运动估计，其中包括全局运动估计和局部运动估计。

[0053] 深度学习：一种由大规模神经元构成的机器学习算法，由于可以很好地解决复杂的非线性问题，目前广泛的应用于在计算机视觉、语音识别、自然语言处理等多个领域。

[0054] 光流：描述相对于观察者的运动所造成的观测目标、表面或边缘的运动。

[0055] 表情迁移：根据录制的人的视频序列，得到该人的表情表达，迁移到另一个人的脸上。

[0056] 人脸分割网络：人脸分割网络指把人脸图像每个像素为左眉毛、右眉毛、左眼、右眼、鼻子、上嘴唇、牙齿、下嘴唇、脸、头发和背景等类别。

[0057] 交叉熵损失：一种在分类领域常用的损失函数。

[0058] L1损失：一维向量的距离的损失函数。 $L_1 = |f(x) - y|$ 。

[0059] 进一步的，深度学习很依赖于数据的质量，有些课题是依赖于视频序列的，现在常用的深度学习方法是利用LSTM等一些基于视频的网络结构预测结果。这样的网络结构一方面需要比较多的资源用于训练并且比较耗时；另一方面收集大量的视频数据也是耗时耗力的。所以在计算机视觉领域很多课题都是基于图像训练的，但是最终希望应用于视频上。

[0060] 在一种用于视频预测的网络模型中，可以对输入图像进行语义分割得到视频序列 $S_1^T = \{S_1, S_2, \dots, S_T\}$ ；然后再对其进行合成得到预测视频序列 $X_1^T = \{x_1, x_2, \dots, x_T\}$ 。具体的，参考图1所示，该方法可以包括如下步骤：

[0061] 首先，对根据待分割图像得到的语义分割图101进行编解码得到对应的中间图像102；并对根据待分割图像得到的真值图像103进行编解码得到对应的光流图像104；

[0062] 其次，t时刻的光流图像通过前t-L帧到t-1帧的预测图像x和分割图像以及t时刻

的分割图像 s 得到 $\tilde{w}_t = W(\tilde{X}_t^{t-1}, s_{t-L}^t)$;

[0063] 然后, t 时刻的中间图像通过前 $t-L$ 帧到 $t-1$ 帧的预测图像 x 和分割图像以及 t 时刻的分割图像 s 得到 $\tilde{h}_t = H(\tilde{X}_t^{t-1}, s_{t-L}^t)$;

[0064] 最后, 光流图像104和中间图像102通过 $[0, 1]$ 的蒙版把中间图像和光流图像进行加权得到最终的预测图像105。

[0065] 本示例实施方式中首先提供了一种基于人脸分割的视频预测方法, 该方法可以运行于服务器、服务器集群或云服务器等; 当然, 本领域技术人员也可以根据需求在其他平台运行本发明的方法, 本示例性实施例中对此不做特殊限定。参考图2所示, 该方法可以包括以下步骤:

[0066] 步骤S210. 利用人脸分割模型对人脸视频数据集进行预测得到所述人脸视频数据集中每帧图像的真伪人脸分割标签。

[0067] 步骤S220. 根据所述人脸视频数据集中的每相邻两帧图像以及所述相邻两帧图像对应的真伪人脸分割标签构建输入数据集。

[0068] 步骤S230. 利用第一损失函数以及所述输入数据集对所述人脸分割模型进行调整得到目标分割模型。

[0069] 步骤S240. 利用所述目标分割模型对待预测人脸图像进行分割得到分割结果, 并对多个所述分割结果进行组合得到人脸预测视频。

[0070] 上述基于人脸分割的视频预测方法中, 一方面, 通过利用人脸分割模型对人脸视频数据集进行预测得到人脸视频数据集中每帧图像的真伪人脸分割标签; 并根据人脸视频数据集中的每相邻两帧图像以及相邻两帧图像对应的真伪人脸分割标签构建输入数据集; 再利用第一损失函数以及输入数据集对人脸分割模型进行调整得到目标分割模型; 最后利用目标分割模型对待预测人脸图像进行分割得到分割结果, 并对多个分割结果进行组合得到人脸预测视频; 解决了现有技术中由于深度学习模型会存在一定的误差, 进而导致视频预测结果不稳定的问题, 提高了人脸预测视频的稳定性; 另一方面, 通过利用人脸分割模型对人脸视频数据集进行预测得到人脸视频数据集中每帧图像的真伪人脸分割标签; 并根据人脸视频数据集中的每相邻两帧图像以及相邻两帧图像对应的真伪人脸分割标签构建输入数据集; 再利用第一损失函数以及输入数据集对人脸分割模型进行调整得到目标分割模型, 解决了现有技术中需要收集大量的视频数据进行训练进而导致较为耗时的问题, 提高了目标分割模型的调整速度; 再一方面, 通过利用目标分割模型对待预测人脸图像进行分割得到分割结果, 并对多个分割结果进行组合得到人脸预测视频, 提高了人脸预测视频的准确性。

[0071] 以下, 将结合附图对本公开示例实施例的基于人脸分割的视频预测方法中的步骤进行详细的解释以及说明。

[0072] 首先, 本公开示例实施例提出了一种可以根据只有有标签的图像的数据和无标签的视频数据, 可以得到预测视频序列结果稳定的深度学习网络模型。比如人脸分割网络模型, 目前有的数据集包括helen-dataset和CelebAMask-HQ等。这些数据集都是基于图像的, 如果预测视频序列的话, 由于模型的误差会得到抖动的视频分割结果。进一步的, 可以把基于图像训练好的模型进行微调, 可以得到比较稳定的视频序列预测结果。收集人脸视频序

列,不用包含任何标签;利用已训练好的人脸分割模型跑出人脸分割结果;然后每两帧作为一组输入,增加特征的L1损失再训练人脸分割模型,这样可以得到比较稳定的人脸分割模型。类比,任何基于图像训练的深度学习模型,都可以通过该方法提高视频预测的稳定性。

[0073] 在步骤S210中,利用人脸分割模型对人脸视频数据集进行预测得到所述人脸视频数据集中每帧图像的真伪人脸分割标签。

[0074] 在本示例实施例中,首先,准备人脸序列数据集D;其中,该部分需要准备无标注的人脸视频数据集。然后,通过人脸分割模型对该人脸视频数据集进行预测得到人脸视频数据集中每帧图像的真伪人脸分割标签(真人脸分割标签或者伪人脸分割标签)。

[0075] 在步骤S220中,根据所述人脸视频数据集中的每相邻两帧图像以及所述相邻两帧图像对应的真伪人脸分割标签构建输入数据集。

[0076] 在本示例实施例中,可以将上述人脸视频数据集中的每相邻两帧图像以及与该相邻两帧图像对应的真伪人脸分割标签作为一组输入数据,并根据该输入数据构建输入数据集。参考图3所示,该输入数据例如可以为连续的两帧RGB图像301。

[0077] 在步骤S230中,利用第一损失函数以及所述输入数据集对所述人脸分割模型进行调整得到目标分割模型。

[0078] 在本示例实施例中,可以将上述输入数据集中的各输入数据输入至上述人脸分割模型中,然后通过编码器302、解码303对输入数据进行编码解码得到第一分割损失304以及第二分割损失305;再根据该第一分割损失304以及第二分割损失305得到L1损失306;最后根据L1损失306对该人脸分割模型进行调整。具体的,该第一损失函数可以如下所示:

[0079] $L=L_1(f_1-f_2)+w*(J_1+J_2)$; (公式1)

[0080] 其中,L为第一损失函数; f_1 为所述每相邻两帧图像中的第一帧图像在所述深度卷积神经网络中的特征; f_2 为所述每相邻两帧图像中的第二帧图像在所述深度卷积神经网络中的特征; L_1 为 f_1 与 f_2 之间的距离; w 为权重,可以用于衡量第二损失函数值与两帧特征之间的损失值; J_1 所述每相邻两帧图像中的第一帧图像的损失函数值; J_2 为所述每相邻两帧图像中的第二帧图像的损失函数值。

[0081] 在步骤S240中,利用所述目标分割模型对待预测人脸图像进行分割得到分割结果,并对多个所述分割结果进行组合得到人脸预测视频。

[0082] 在本示例实施例中,当得到上述目标分割模型后,首先,可以对待预测人脸图像进行人脸检测;其中,通过对人脸进行检测可以减少复杂背景带来的影响或者人像太小带来的影响;然后将检测后的待预测人脸图像输入至目标分割模型中得到人脸分割结果;最后再将多帧人脸分割结果按照一定的顺序进行组合拼接即可得到人脸预测视频,例如可以是聊天过程中使用的动图等等。通过该方法,提高了人脸预测视频的稳定性,同时也提高了人脸预测视频的准确性。

[0083] 其中,对多个分割结果进行组合得到人脸预测视频可以包括:可以将各分割结果按顺序进行拼接,即可得到上述人脸预测视频。例如,第一帧分割结果为人脸的正脸,然后第二帧的分割结果为人脸稍微向左侧移动(比如移动角度为 10°);第三帧以及往后帧的分割结果为人脸依次向左移动;则对所有的分割结果依次进行拼接即可得到一个人脸向左转的预测视频。

[0084] 进一步的,为了便于利用人脸分割模型对人脸视频数据集进行预测,该基于人脸

分割的视频预测方法还可以包括：利用人脸图像以及人脸分割标签对深度卷积神经网络进行再训练得到所述人脸分割模型。具体的，参考图4所示，利用人脸图像以及人脸分割标签对深度卷积神经网络进行再训练得到所述人脸分割模型可以包括步骤S410-步骤S430，以下进行详细说明。

[0085] 在步骤S410中，根据所述深度卷积神经网络的当前网络参数值提取所述人脸图像的第一特征向量以及所述人脸分割标签的第二特征向量。

[0086] 在本示例实施例中，可以采用深度卷积神经网络模型的当前网络参数值提取人脸图像的第一特征向量以及人脸分割标签的第二特征向量。其中，该神经网络模型例如可以是VGG网络模型，也可以是其他神经网络模型，例如可以是RESNET网络模型，本示例对此不做特殊限制，进一步的，该VGG网络模型可以包括一系列的编码器、解码器、卷积层、池化层和全连接层。并且，上述人脸分割标签包括人脸、左眼眉、右眼眉、左眼、右眼、鼻子、上嘴唇、牙齿、下嘴唇、头发以及背景等等，也可以包括额头、脸颊以及颧骨等等，本示例对此不做特殊限制。

[0087] 在步骤S420中，根据所述第一特征向量以及第二特征向量计算损失函数值，并判断所述损失函数值是否小于预设阈值。

[0088] 在本示例实施例中，首先，配置第二损失函数，并根据第一特征向量、第二特征向量以及第二损失函数计算损失函数值；其中，第二损失函数可以包括：

[0089]

$$J = -\sum_{c=1}^C y_c \log(p_c); \quad (\text{公式 2})$$

[0090] 其中，J为损失函数值；C为所述人脸分割标签的数量； y_c 为第一特征向量； p_c 为第二特征向量。

[0091] 进一步的，当得到该损失函数值以后，可以判断该损失函数值是否小于预设阈值；其中，该预设阈值可以为0，也可以为其他数值，例如可以是0.1、0.2或者0.3等等，本示例对此不做特殊限制。

[0092] 在步骤S430中，如果所述损失函数值小于所述预设阈值，则确定所述当前网络参数值对应的深度卷积神经网络为所述人脸分割模型。

[0093] 进一步的，在本示例实施方法中，如果所述损失函数值大于所述预设阈值，则利用梯度回传算法调整所述卷积神经网络的网络参数值的大小。其中，使用梯度回传算法(Back Propagation)将损失函数值回传到VGG网络模型，调整VGG网络模型的网络参数值。根据损失函数，计算损失函数值相对于之前的网络层中每一个参数的梯度，使用梯度下降法更新该层的网络参数值。然后，根据上一层传递来的损失函数值，用同样的办法更新更前面一层的网络参数值。直到整个网络中的参数值都得到更新。

[0094] 以下，结合图5对本公开示例实施例中涉及的基于人脸分割的视频预测方法进行进一步的解释以及说明。参考图5所示，该基于人脸分割的视频预测方法可以包括以下步骤：

[0095] 步骤S510，根据深度卷积神经网络的当前网络参数值提取人脸图像的第一特征向量以及人脸分割标签的第二特征向量；

[0096] 步骤S520，根据第一特征向量以及第二特征向量计算损失函数值，并判断损失函

数值是否小于预设阈值;如果是,则跳转至步骤S530;如果否,则跳转至步骤S540;

[0097] 步骤S530,确定当前网络参数值对应的深度卷积神经网络为人脸分割模型;

[0098] 步骤S540,利用梯度回传算法调整卷积神经网络的网络参数值的大小,直至损失函数值小于预设阈值;

[0099] 步骤S550,利用人脸分割模型对人脸视频数据集进行预测得到人脸视频数据集中每帧图像的真伪人脸分割标签;

[0100] 步骤S560,根据人脸视频数据集中的每相邻两帧图像以及相邻两帧图像对应的真伪人脸分割标签构建输入数据集;

[0101] 步骤S570,利用第一损失函数以及输入数据集对人脸分割模型进行调整得到目标分割模型;

[0102] 步骤S580,利用目标分割模型对待预测人脸图像进行分割得到分割结果,并对多个分割结果进行组合得到人脸预测视频。

[0103] 此处需要进一步补充说明的是,本公开涉及的技术方案不仅可以用于人脸视频预测(譬如关于人脸的动图),也适用于其他视频的预测。譬如,针对于其他的视频预测,可以直接将脸部图像更换成其他部位,譬如整个人的视频预测或者动物视频的预测等等;仅仅只是需要对分别标签进行替换即可。

[0104] 综上所述,本公开示例实施例提供的基于人脸分割的视频预测方法的主要优势在于:一方面,解决了基于图像训练的深度学习网络模型用于视频序列预测时带来的抖动问题,解决了视频序列的预测就需要收集对应的有标注的视频序列的数据的问题。提高了基于图像的深度学习预测的稳定性。该方法是一个普适的方法,不止可以用在人脸分割网络结构中,只要是基于深度学习的网络预测模型都可以用该方法再训练提高预测视频序列的稳定性。

[0105] 另一方面,本公开示例实施例不需要构建复杂的网络结构,只是在原有的网络结构上再训练达到稳定视频预测结果的目的。

[0106] 再一方面,本公开示例实施例通过对基于图像的深度学习网络结构再训练达到不需要大量有标注的视频序列数据的目的。通过基于图像的标注数据训练了网络模型之后,只利用无标注的视频序列既可以对网络结构再训练,从而达到防止网络模型预测视频序列抖动的问题。

[0107] 本公开还提供了一种基于人脸分割的视频预测装置。参考图6所示,该基于人脸分割的视频预测装置可以包括预测模块610、数据集构建模块620、调整模块630以及分割模块640。其中:

[0108] 预测模块610可以用于利用人脸分割模型对人脸视频数据集进行预测得到所述人脸视频数据集中每帧图像的真伪人脸分割标签。

[0109] 数据集构建模块620可以用于根据所述人脸视频数据集中的每相邻两帧图像以及所述相邻两帧图像对应的真伪人脸分割标签构建输入数据集。

[0110] 调整模块630可以用于利用第一损失函数以及所述输入数据集对所述人脸分割模型进行调整得到目标分割模型。

[0111] 分割模块640可以用于利用所述目标分割模型对待预测人脸图像进行分割得到分割结果,并对多个所述分割结果进行组合得到人脸预测视频。

[0112] 在本公开的一种示例性实施例中,所述基于人脸分割的视频预测装置还包括:

[0113] 人脸分割模型训练模块,可以用于利用人脸图像以及人脸分割标签对深度卷积神经网络进行训练得到所述人脸分割模型。

[0114] 在本公开的一种示例性实施例中,利用人脸图像以及人脸分割标签对深度卷积神经网络进行训练得到所述人脸分割模型包括:

[0115] 根据所述深度卷积神经网络的当前网络参数值提取所述人脸图像的第一特征向量以及所述人脸分割标签的第二特征向量;

[0116] 根据所述第一特征向量以及第二特征向量计算损失函数值,并判断所述损失函数值是否小于预设阈值;

[0117] 如果所述损失函数值小于所述预设阈值,则确定所述当前网络参数值对应的深度卷积神经网络为所述人脸分割模型。

[0118] 在本公开的一种示例性实施例中,根据所述第一特征向量以及第二特征向量计算损失函数值包括:

[0119] 配置第二损失函数,并根据所述第一特征向量、第二特征向量以及所述第二损失函数计算所述损失函数值:

[0120] 其中,所述损失函数包括:

[0121]

$$J = -\sum_{c=1}^C y_c \log(p_c); \quad (\text{公式 2})$$

[0122] 其中,J为损失函数值;C为所述人脸分割标签的数量; y_c 为第一特征向量; p_c 为第二特征向量。

[0123] 在本公开的一种示例性实施例中,所述基于人脸分割的视频预测装置还包括:

[0124] 调整子模块,用于如果所述损失函数值大于所述预设阈值,则利用梯度回传算法调整所述卷积神经网络的网络参数值的大小。

[0125] 在本公开的一种示例性实施例中,所述第一损失函数包括:

[0126] $L=L_1(f_1-f_2)+w*(J_1+J_2)$; (公式1)

[0127] 其中,L为第一损失函数; f_1 为所述每相邻两帧图像中的第一帧图像在所述深度卷积神经网络中的特征; f_2 为所述每相邻两帧图像中的第二帧图像在所述深度卷积神经网络中的特征; L_1 为 f_1 与 f_2 之间的距离; w 为权重; J_1 为所述每相邻两帧图像中的第一帧图像的损失函数值; J_2 为所述每相邻两帧图像中的第二帧图像的损失函数值。

[0128] 在本公开的一种示例性实施例中,所述人脸分割标签包括人脸、左眼眉、右眼眉、左眼、右眼、鼻子、上嘴唇、牙齿、下嘴唇、头发以及背景中的多种。

[0129] 上述基于人脸分割的视频预测装置中各模块的具体细节已经在对应的基于人脸分割的视频预测方法中进行了详细的描述,因此此处不再赘述。

[0130] 应当注意,尽管在上文详细描述中提及了用于动作执行的设备的若干模块或者单元,但是这种划分并非强制性的。实际上,根据本发明的实施方式,上文描述的两个或更多模块或者单元的特征和功能可以在一个模块或者单元中具体化。反之,上文描述的一个模块或者单元的特征和功能可以进一步划分为由多个模块或者单元来具体化。

[0131] 此外,尽管在附图中以特定顺序描述了本发明中方法的各个步骤,但是,这并非要

求或者暗示必须按照该特定顺序来执行这些步骤,或是必须执行全部所示的步骤才能实现期望的结果。附加的或备选的,可以省略某些步骤,将多个步骤合并为一个步骤执行,以及/或者将一个步骤分解为多个步骤执行等。

[0132] 通过以上的实施方式的描述,本领域的技术人员易于理解,这里描述的示例实施方式可以通过软件实现,也可以通过软件结合必要的硬件的方式来实现。因此,根据本发明实施方式的技术方案可以以软件产品的形式体现出来,该软件产品可以存储在一个非易失性存储介质(可以是CD-ROM,U盘,移动硬盘等)中或网络上,包括若干指令以使得一台计算设备(可以是个人计算机、服务器、移动终端、或者网络设备等)执行根据本发明实施方式的方法。

[0133] 在本发明的示例性实施例中,还提供了一种能够实现上述方法的电子设备。

[0134] 所属技术领域的技术人员能够理解,本发明的各个方面可以实现为系统、方法或程序产品。因此,本发明的各个方面可以具体实现为以下形式,即:完全的硬件实施方式、完全的软件实施方式(包括固件、微代码等),或硬件和软件方面结合的实施方式,这里可以统称为“电路”、“模块”或“系统”。

[0135] 下面参照图7来描述根据本发明的这种实施方式的电子设备700。图7显示的电子设备700仅仅是一个示例,不应对本发明实施例的功能和使用范围带来任何限制。

[0136] 如图7所示,电子设备700以通用计算设备的形式表现。电子设备700的组件可以包括但不限于:上述至少一个处理单元710、上述至少一个存储单元720、连接不同系统组件(包括存储单元720和处理单元710)的总线730。

[0137] 其中,所述存储单元存储有程序代码,所述程序代码可以被所述处理单元710执行,使得所述处理单元710执行本说明书上述“示例性方法”部分中描述的根据本发明各种示例性实施方式的步骤。例如,所述处理单元710可以执行如图2中所示的步骤S210:利用人脸分割模型对人脸视频数据集进行预测得到所述人脸视频数据集中每帧图像的真伪人脸分割标签;步骤S220:根据所述人脸视频数据集中的每相邻两帧图像以及所述相邻两帧图像对应的真伪人脸分割标签构建输入数据集;步骤S230:利用第一损失函数以及所述输入数据集对所述人脸分割模型进行调整得到目标分割模型;步骤S240:利用所述目标分割模型对待预测人脸图像进行分割得到分割结果,并对多个所述分割结果进行组合得到人脸预测视频。

[0138] 存储单元720可以包括易失性存储单元形式的可读介质,例如随机存取存储单元(RAM) 7201和/或高速缓存存储单元7202,还可以进一步包括只读存储单元(ROM) 7203。

[0139] 存储单元720还可以包括具有一组(至少一个)程序模块7205的程序/实用工具7204,这样的程序模块7205包括但不限于:操作系统、一个或者多个应用程序、其它程序模块以及程序数据,这些示例中的每一个或某种组合中可能包括网络环境的实现。

[0140] 总线730可以为表示几类总线结构中的一种或多种,包括存储单元总线或者存储单元控制器、外围总线、图形加速端口、处理单元或者使用多种总线结构中的任意总线结构的局域总线。

[0141] 电子设备700也可以与一个或多个外部设备770(例如键盘、指向设备、蓝牙设备等)通信,还可与一个或者多个使得用户能与该电子设备700交互的设备通信,和/或与使得该电子设备700能与一个或多个其它计算设备进行通信的任何设备(例如路由器、调制解调

器等等)通信。这种通信可以通过输入/输出(I/O)接口750进行。并且,电子设备700还可以通过网络适配器760与一个或者多个网络(例如局域网(LAN),广域网(WAN)和/或公共网络,例如因特网)通信。如图所示,网络适配器760通过总线730与电子设备700的其它模块通信。应当明白,尽管图中未示出,可以结合电子设备700使用其它硬件和/或软件模块,包括但不限于:微代码、设备驱动器、冗余处理单元、外部磁盘驱动阵列、RAID系统、磁带驱动器以及数据备份存储系统等。

[0142] 通过以上的实施方式的描述,本领域的技术人员易于理解,这里描述的示例实施方式可以通过软件实现,也可以通过软件结合必要的硬件的方式来实现。因此,根据本发明实施方式的技术方案可以以软件产品的形式体现出来,该软件产品可以存储在一个非易失性存储介质(可以是CD-ROM,U盘,移动硬盘等)中或网络上,包括若干指令以使得一台计算设备(可以是个人计算机、服务器、终端装置、或者网络设备等)执行根据本发明实施方式的方法。

[0143] 在本发明的示例性实施例中,还提供了一种计算机可读存储介质,其上存储有能够实现本说明书上述方法的程序产品。在一些可能的实施方式中,本发明的各个方面还可以实现为一种程序产品的形式,其包括程序代码,当所述程序产品在终端设备上运行时,所述程序代码用于使所述终端设备执行本说明书上述“示例性方法”部分中描述的根据本发明各种示例性实施方式的步骤。

[0144] 根据本发明的实施方式的用于实现上述方法的程序产品,其可以采用便携式紧凑盘只读存储器(CD-ROM)并包括程序代码,并可以在终端设备,例如个人电脑上运行。然而,本发明的程序产品不限于此,在本文件中,可读存储介质可以是任何包含或存储程序的有形介质,该程序可以被指令执行系统、装置或者器件使用或者与其结合使用。

[0145] 所述程序产品可以采用一个或多个可读介质的任意组合。可读介质可以是可读信号介质或者可读存储介质。可读存储介质例如可以为但不限于电、磁、光、电磁、红外线、或半导体的系统、装置或器件,或者任意以上的组合。可读存储介质的更具体的例子(非穷举的列表)包括:具有一个或多个导线的电连接、便携式盘、硬盘、随机存取存储器(RAM)、只读存储器(ROM)、可擦式可编程只读存储器(EPROM或闪存)、光纤、便携式紧凑盘只读存储器(CD-ROM)、光存储器件、磁存储器件、或者上述的任意合适的组合。

[0146] 计算机可读信号介质可以包括在基带中或者作为载波一部分传播的数据信号,其中承载了可读程序代码。这种传播的数据信号可以采用多种形式,包括但不限于电磁信号、光信号或上述的任意合适的组合。可读信号介质还可以是可读存储介质以外的任何可读介质,该可读介质可以发送、传播或者传输用于由指令执行系统、装置或者器件使用或者与其结合使用的程序。

[0147] 可读介质上包含的程序代码可以用任何适当的介质传输,包括但不限于无线、有线、光缆、RF等等,或者上述的任意合适的组合。

[0148] 可以以一种或多种程序设计语言的任意组合来编写用于执行本发明操作的程序代码,所述程序设计语言包括面向对象的程序设计语言—诸如Java、C++等,还包括常规的过程式程序设计语言—诸如“C”语言或类似的设计语言。程序代码可以完全地在用户计算设备上执行、部分地在用户设备上执行、作为一个独立的软件包执行、部分在用户计算设备上部分在远程计算设备上执行、或者完全在远程计算设备或服务器上执行。在涉及远

程计算设备的情形中,远程计算设备可以通过任意种类的网络,包括局域网(LAN)或广域网(WAN),连接到用户计算设备,或者,可以连接到外部计算设备(例如利用因特网服务提供商来通过因特网连接)。

[0149] 此外,上述附图仅是根据本发明示例性实施例的方法所包括的处理的示意性说明,而不是限制目的。易于理解,上述附图所示的处理并不表明或限制这些处理的时间顺序。另外,也易于理解,这些处理可以是例如在多个模块中同步或异步执行的。

[0150] 本领域技术人员在考虑说明书及实践这里发明的发明后,将容易想到本发明的其他实施例。本申请旨在涵盖本发明的任何变型、用途或者适应性变化,这些变型、用途或者适应性变化遵循本发明的一般性原理并包括本发明未发明的本技术领域中的公知常识或惯用技术手段。说明书和实施例仅被视为示例性的,本发明的真正范围和精神由权利要求指出。

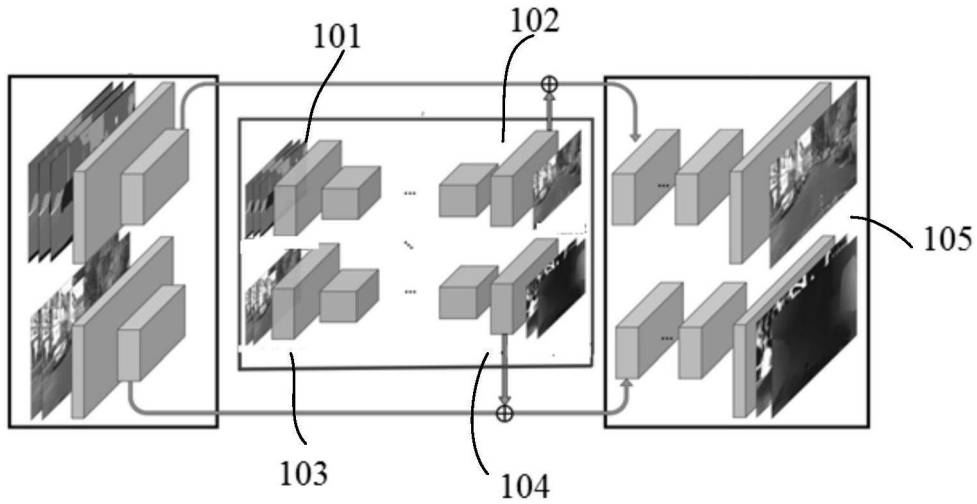


图1

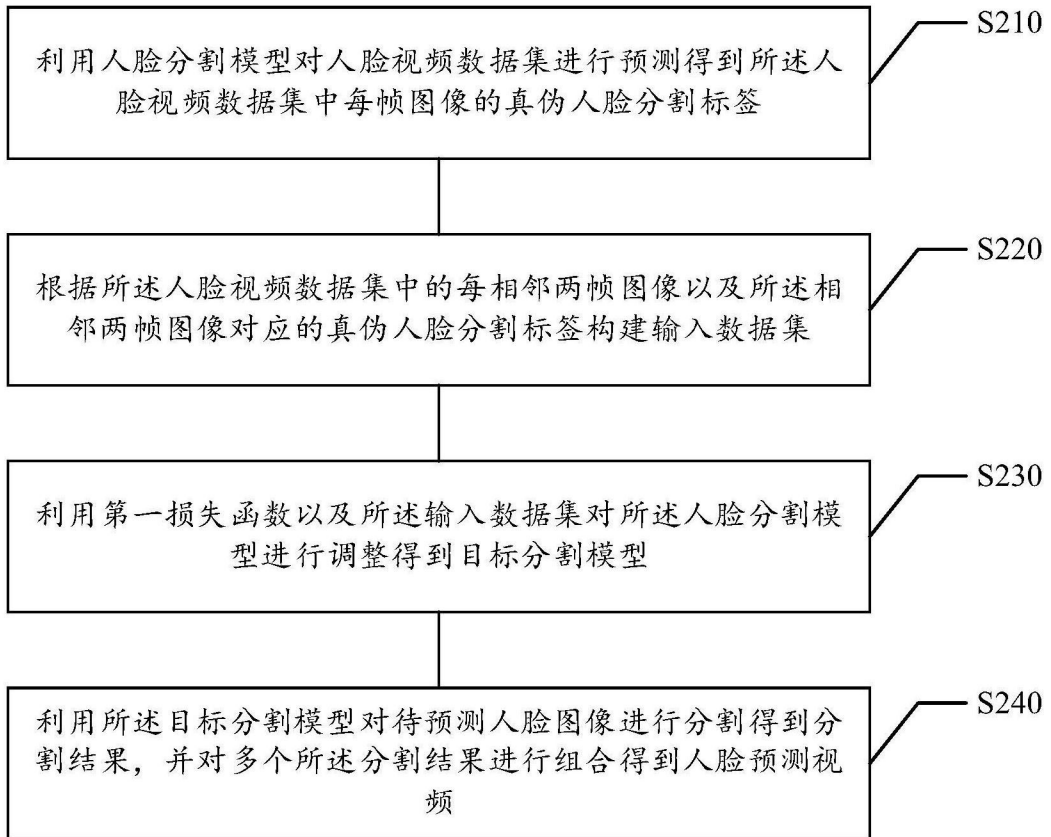


图2

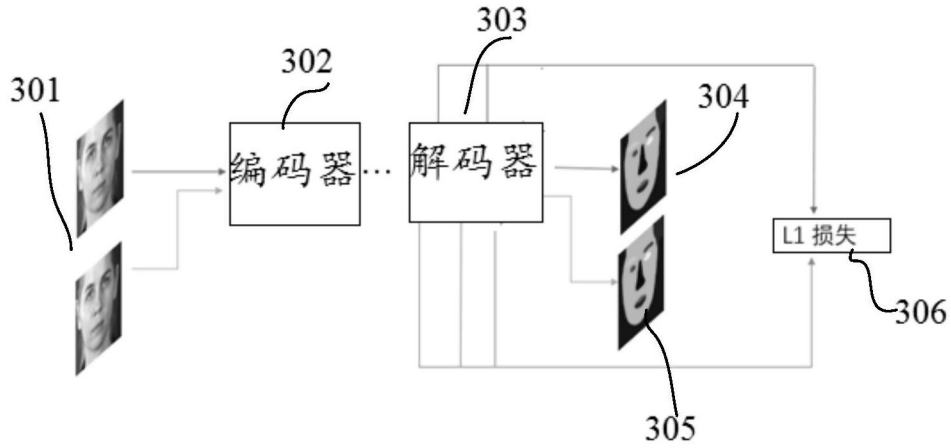


图3

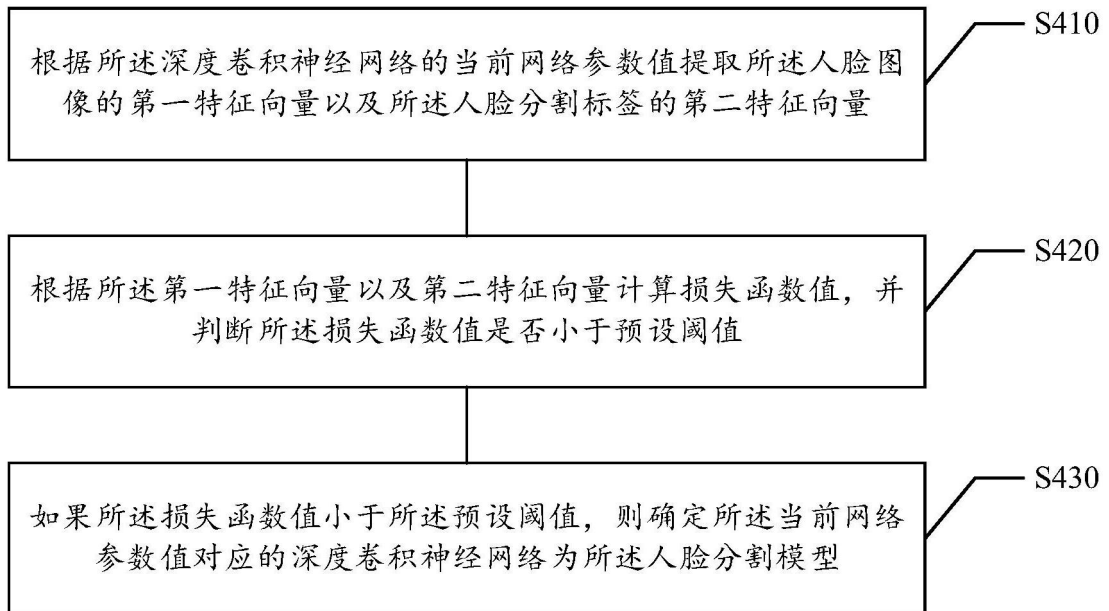


图4

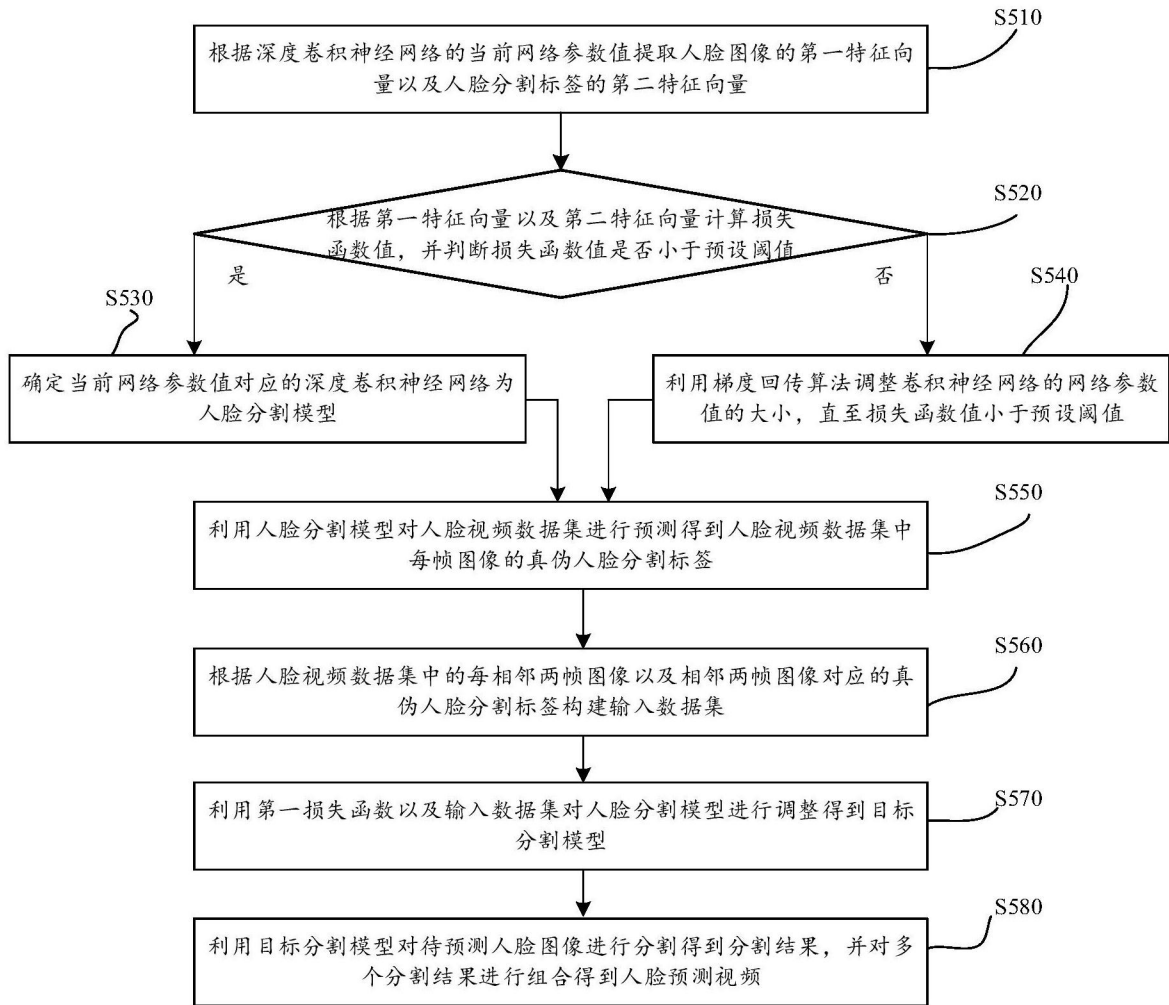


图5

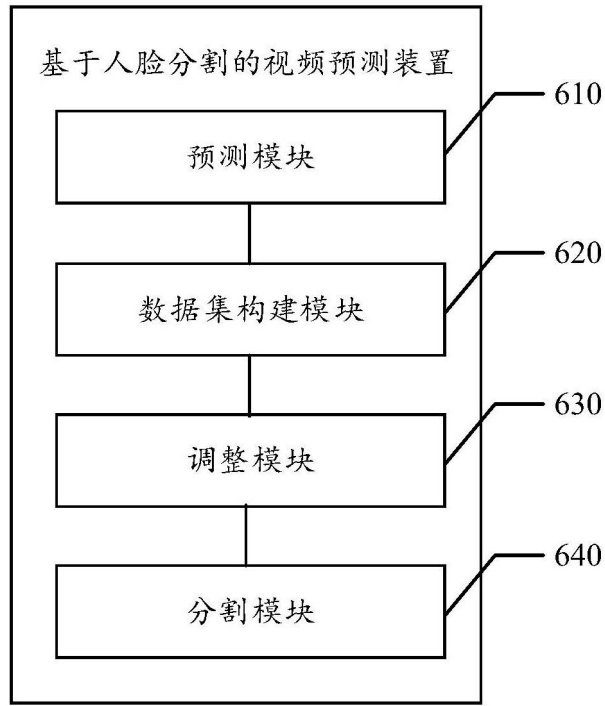


图6

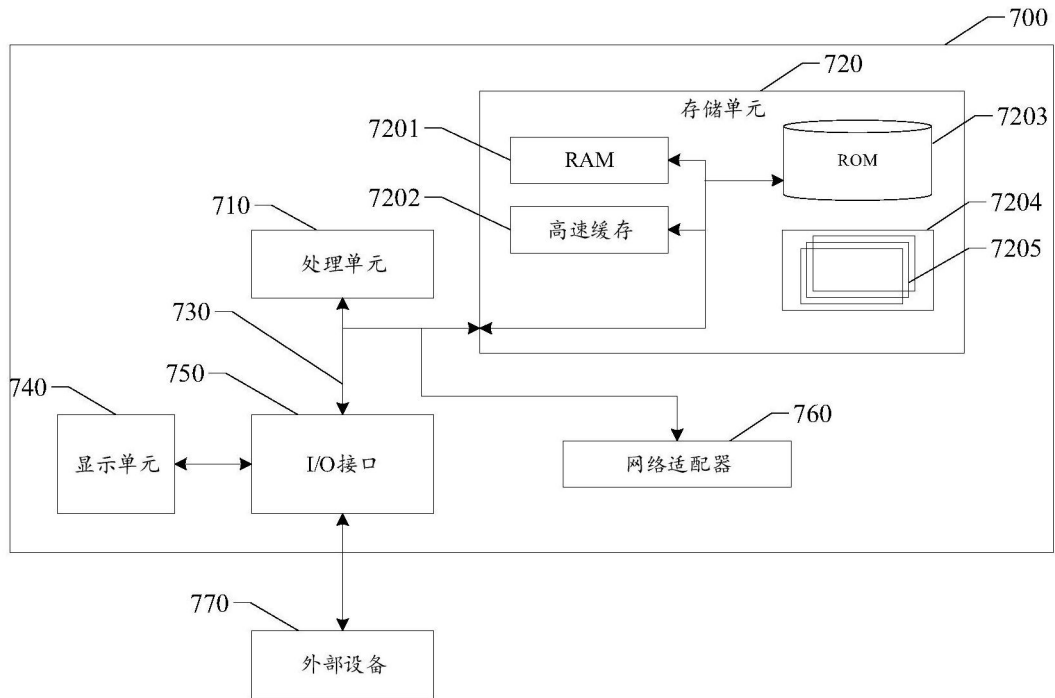


图7