

(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 10 月 1 日 (01.10.2020)



(10) 国际公布号
WO 2020/192704 A1

- (51) 国际专利分类号:
G06K 9/00 (2006.01) **G06T 3/00** (2006.01)
- (21) 国际申请号: PCT/CN2020/081208
- (22) 国际申请日: 2020 年 3 月 25 日 (25.03.2020)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201910228962.5 2019年3月25日 (25.03.2019) 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: 陈法圣 (CHEN, Fasheng); 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: 深圳市深佳知识产权代理事务所(普通合伙) (SHENPAT INTELLECTUAL PROPERTY AGENCY); 中国广东省深圳市罗湖区南湖街道春风路庐山大厦 B 座 18C2、18D、18E、18E2, Guangdong 518001 (CN)。
- (81) 指定国(除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, 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,

(54) Title: IMAGE PROCESSING MODEL TRAINING METHOD, IMAGE PROCESSING METHOD AND DEVICE, AND STORAGE MEDIUM

(54) 发明名称: 图像处理模型的训练方法、图像处理方法、装置及存储介质

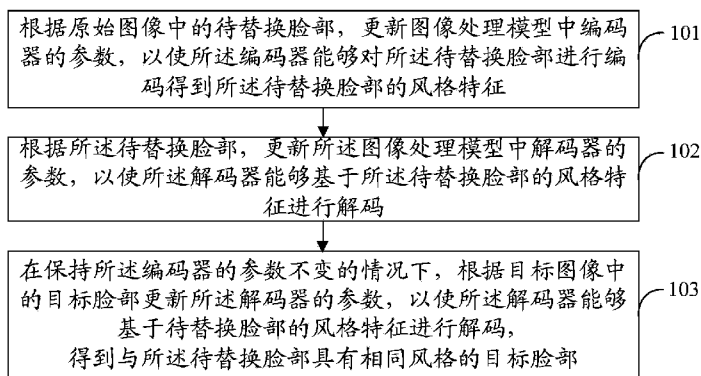


图 9

- 101 Update parameters of an encoder in an image processing model according to a face to be replaced in an original image, so that the encoder can encode said face, and style features of said face are obtained
- 102 Update parameters of a decoder in the image processing model according to said face, so that the decoder can perform decoding on the basis of the style features of said face
- 103 Under the condition that the parameters of the encoder are kept unchanged, update the parameters of the decoder according to a target face in a target image, so that the decoder can perform decoding on the basis of the style features of said face, and the target face with the same style as said face is obtained

(57) Abstract: The present application provides an image processing model training method. The method is executed by a training device, and the method comprises: updating parameters of an encoder in an image processing model according to a face to be replaced in an original image, so that the encoder can encode said face, and style features of said face are obtained; updating parameters of a decoder in the image processing model according to said face of the original image, so that the decoder can perform decoding on the basis of the style features of said face; and under the condition that the parameters of the encoder are kept unchanged, updating the parameters of the decoder according to a target face in a target image, so that the decoder can perform decoding on the basis of the

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))。

style features of said face, and the target face with the same style as said face is obtained. The present application further provides an image processing method and device for the image processing model, and a storage medium. According to the present application, the training efficiency can be improved under the condition of ensuring the recognition precision of the image processing model.

(57) 摘要: 本申请提供了图像处理模型的训练方法, 所述方法由训练装置执行, 所述方法包括: 根据原始图像中的待替换脸部, 更新图像处理模型中编码器的参数, 以使编码器能够对待替换脸部进行编码, 得到待替换脸部的风格特征; 根据原始图像的待替换脸部, 更新图像处理模型中解码器的参数, 以使解码器能够基于待替换脸部的风格特征进行解码; 在保持编码器的参数不变的情况下, 根据目标图像中目标脸部更新解码器的参数, 以使解码器能够基于待替换脸部的风格特征进行解码, 得到与待替换脸部具有相同风格的目标脸部。本申请还提供了图像处理模型的图像处理方法、装置及存储介质。本申请能够在保证图像处理模型的识别精度的情况下提升训练效率。

图像处理模型的训练方法、图像处理方法、装置及存储介质

本申请要求于 2019 年 03 月 25 日提交中国专利局、申请号为 201910228962.5、申请名称为“图像处理模型的训练方法、图像处理方法、装置及存储介质”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

本申请涉及图像处理技术，尤其涉及图像处理模型的训练和处理。

背景技术

在电影特效和互联网社交等应用中，存在保持原始图像（例如图片或者视频帧）中人物的脸部（也称为待替换脸部）的风格的情况下，将待替换脸部替换为其他人物的脸部（也称为目标脸部）的需求。为此，人工智能技术提供了训练适当的图像处理模型来支持上述应用的方案。

发明内容

有鉴于此，本申请实施例提供图像处理模型的训练方法、图像处理方法、装置及存储介质，能够保证图像处理模型的识别精度的情况下提升图像处理模型的训练效率。

本申请实施例的技术方案是这样实现的：

一方面，本申请实施例提供了一种图像处理模型的训练方法，所述方法由训练装置执行，包括：

根据原始图像中的待替换脸部，更新图像处理模型中编码器的参数，以使所述编码器能够对所述待替换脸部进行编码，得到所述待替换脸部的风格特征；

根据所述原始图像的待替换脸部，更新所述图像处理模型中解码器的参数，以使所述解码器能够基于所述待替换脸部的风格特征进行解码；

在保持所述编码器的参数不变的情况下，根据目标图像中目标脸部的风格特征更新所述解码器的参数，以使所述解码器能够基于所述待替换的风格特征进行解码，得到与所述待替换脸部具有相同风格的目标脸部。

在一种可能的实现方式中，所述方法还包括：通过所述编码器对所述待替换图像进行编码得到所述待替换图像中待替换脸部的风格特征，将所述风格特

-2-

征输入所述解码器；在所述解码器中基于所述待替换脸部的风格特征进行解码，得到与所述待替换脸部具有相同风格的目标脸部。

5 在一种可能的实现方式中，当根据目标图像中的目标脸部更新所述解码器的参数之前，所述方法还包括：对所述目标图像进行随机增广处理；将经过随机增广处理的目标图像进行随机扭曲处理，得到扭曲的所述目标图像；截取扭曲的所述目标图像中的五官图像，用于更新所述解码器的参数。

另一方面，本申请实施例提供了一种图像处理模型的图像处理方法，包括：

通过图像处理模型的编码器，对原始图像进行编码得到所述原始图像中待替换脸部的风格特征；

10 在所述图像处理模型的解码器中基于待替换脸部的风格特征进行解码，得到与所述待替换脸部具有相同风格的目标脸部；

其中，所述解码器的参数是在保持所述编码器的参数不变的情况下，根据包括所述目标脸部的目标图像进行更新得到。

另一方面，本申请实施例提供了一种图像处理模型的训练装置，包括：

15 编码器训练模块，用于根据原始图像中的待替换脸部，更新图像处理模型中编码器的参数，以使所述编码器能够对所述待替换脸部进行编码，得到所述待替换脸部的风格特征；

20 解码器训练模块，用于根据所述原始图像的待替换脸部，更新所述图像处理模型中解码器的参数，以使所述解码器能够基于所述待替换脸部的风格特征进行解码；

所述解码器训练模块，还用于在保持所述编码器的参数不变的情况下，根据目标图像中目标脸部的风格特征更新所述解码器的参数，以使所述解码器能够基于所述待替换脸部的风格特征进行解码，得到与所述待替换脸部具有相同风格的目标脸部。

25 在一种可能的实现方式中，所述编码器训练模块具体用于：将扭曲原始图像中待替换脸部的五官图像代入损失函数，确定所述损失函数满足收敛条件时对应所述编码器的参数，其中，所述损失函数对应由所述编码器和所述解码器构成的自编码网络，所述扭曲原始图像是对所述原始图像进行扭曲处理得到。

在一种可能的实现方式中，所述解码器训练模块具体用于：将扭曲原始图

—3—

像中待替换脸部的五官图像,代入由所述编码器和所述解码器构成的自编码网络对应的损失函数,确定所述损失函数满足收敛条件时对应所述解码器的参数,其中所述扭曲原始图像是对所述原始图像进行扭曲处理得到。

5 在一种可能的实现方式中,所述解码器训练模块具体用于:将扭曲目标图像中目标脸部的五官图像,代入由所述编码器和所述解码器构成的自编码网络对应的损失函数;在保持所述损失函数中对应所述编码器的参数不变的情况下,执行以下操作:确定所述损失函数满足收敛条件时对应所述解码器的参数;其中,所述扭曲目标图像是对所述目标图像进行扭曲处理得到。

10 在一种可能的实现方式中,所述编码器训练模块还用于:当根据原始图像中的待替换脸部,更新所述像处理模型中编码器的参数之前,从样本图像集合的各个样本图像中截取包括所述待替换脸部的所述原始图像;将所截取的各个所述原始图像基于待替换脸部的特征点进行对齐。

15 在一种可能的实现方式中,所述编码器训练模块还用于:当根据原始图像中的待替换脸部,更新所述图像处理模型中编码器的参数之前,对所述原始图像进行随机增广处理;将经过随机增广处理的原始图像进行随机扭曲处理,得到扭曲的所述原始图像;截取所述扭曲的原始图像中待替换脸部的五官图像,并对所述截取的五官图像进行缩放处理以用于更新所述解码器的参数时适配所述编码器支持输入的分辨率。

20 在一种可能的实现方式中,所述编码器训练模块,具体用于:基于所述原始图像的中心进行随机旋转;将随机旋转后的所述原始图像进行随机缩放;将随机缩放后所述原始图像进行随机平移。

在一种可能的实现方式中,所述编码器训练模块,具体用于:在经过随机增广处理的原始图像中插入噪声,所述噪声包括固定噪声和动态噪声中的至少一种。

25 在一种可能的实现方式中,所述编码器训练模块,具体用于:将所述经过随机增广处理的原始图像进行网格化,并为网格中的至少部分节点分配坐标;在所述至少部分节点中添加与所分配的坐标相对应的噪声。

在一种可能的实现方式中,所述解码器训练模块,具体用于:对所述目标图像进行随机增广处理;将经过随机增广处理的目标图像进行随机扭曲处理,

—4—

得到扭曲的所述目标图像；截取扭曲的所述目标图像中的五官图像。

在一种可能的实现方式中，所述解码器训练模块，还用于调用所述编码器执行以下处理：对所述待替换图像进行编码得到所述待替换图像中待替换脸部的风格特征，将所述风格特征输入所述解码器；所述解码器训练模块，还用于
5 调用所述解码器执行以下处理：基于所述待替换脸部的风格特征进行解码，得到与所述待替换脸部具有相同风格的目标脸部。

另一方面，本申请实施例提供了一种图像处理装置，包括：

编码器模块，用于对原始图像进行编码得到所述原始图像中待替换脸部的风格特征；

10 解码器模块，用于基于所述待替换脸部的风格特征进行解码，得到与所述待替换脸部具有相同风格的目标脸部；

其中，所述解码器模块是在保持所述编码器模块的参数不变的情况下，根据包括所述目标脸部的目标图像更新得到。

15 另一方面，本申请实施例提供了一种图像处理模型的训练装置，所述训练装置包括：

存储器，用于存储可执行指令；

处理器，用于运行所述存储器存储的可执行指令时，实现上述图像处理模型的训练方法。

20 另一方面，本申请实施例提供了一种图像处理装置，所述图像处理装置包括：

存储器，用于存储可执行指令；

处理器，用于运行所述存储器存储的可执行指令时，实现上述图像处理模型的图像处理方法。

25 另一方面，本申请实施例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用于存储计算机程序，所述计算机程序用于执行上述图像处理模型的训练方法或图像处理方法。

在本申请实施例公开的一方面，一种包括指令的计算机程序产品，当其在计算机上运行时，使得所述计算机执行上述图像处理模型的训练方法或图像处理方法。

—5—

本申请实施例具有以下有益效果：

通过原始图像对图像处理模型编码器和解码器进行训练，并在保持编码器参数不变的情况下利用目标对象对相应的解码器进行训练，实现了解码器的训练与原始图像的解耦，减少训练所需要的图像的数量，进而可以有效缩短图像处理模型的训练时间。

附图说明

图1为本申请实施例提供的图像处理模型的结构示意图；

图2为本申请实施例提供的编码器的示例性结构图；

图3为本申请实施例提供的解码器的示例性结构图；

10 图4为本申请实施例提供的图像处理模型进行换脸的原理示意图；

图5为本申请实施例提供的图像处理模型进行换脸的原理示意图；

图6为本申请实施例提供的图像处理模型应用相关技术提供的训练方法的原理示意图；

15 图7为本申请实施例提供的图像处理模型的训练方法和应用方法的使用场景示意图；

图8为本申请实施例提供的用于训练图像处理模型的训练装置的结构示意图；

图9为本申请实施例提供的图像处理模型的训练方法的流程示意图；

图10为本申请实施例提供的图像处理装置的一个可选的结构示意图；

20 图11为本申请实施例提供的图像处理模型的图像处理方法的流程示意图；

图12为本申请实施例提供的用于交换人脸的图像处理模型的训练原理示意图；

图13A为本申请实施例提供的图像处理模型的训练方法的原理示意图；

图13B为本申请实施例提供的在人脸图像中绘制网格的示意图；

25 图14为本申请实施例提供的初始化编码器的原理示意图；

图15为本申请实施例对解码器的训练原理示意图；

图16为本申请实施例图像处理模型进行换脸的原理示意图；

图17为本申请实施例提供的形成具有换脸效果的视频的流程图。

具体实施方式

为了使本申请的目的、技术方案和优点更加清楚，下面将结合附图对本申请作进一步地详细描述，所描述的实施例不应视为对本申请的限制，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其它实施例，都属于本申请保护的范围。

5 在以下的描述中，涉及到“一些实施例”，其描述了所有可能实施例的子集，但是可以理解，“一些实施例”可以是所有可能实施例的相同子集或不同子集，并且可以在不冲突的情况下相互结合。

对本申请实施例进行进一步详细说明之前，对本申请实施例中涉及的名词和术语进行说明，本申请实施例中涉及的名词和术语适用于如下的解释。

10 1) 原始图像，也称为原始脸部图像，即包括待替换脸部的图像，例如图片或者视频的帧图像。

2) 目标图像，也称为目标脸部图像，即包括用于替换“待替换脸部”的图像，例如图片或者视频中的帧图像。

15 3) 编码器，用于从原始图像中待替换脸部进行特征降维（即降低特征的维度，也称为编码或特征提取）的模型，输出用于表示待替换脸部的风格的特征。

4) 风格特征，即基于编码器的中间层（隐藏层）从原始图像的待替换脸部中提取表示待替换脸部在视觉上所表现出来的具有明显的辨识性的特征，可以包括表情、神态、光照、嘴型、视线方向、是否闭眼和肤色等。

20 5) 解码器，与目标脸部一一对应的模型，用于从输入到解码器的风格特征还原（也称为解码或重构）具有相应风格特征的目标脸部的图像。

6) 换脸，即将原始图像中的待替换脸部替换为与待替换脸部具有相同风格特征的目标脸部。

首先对本申请实施例所提供的图像处理模型进行说明，

25 图1为本申请实施例提供的图像处理模型的结构示意图，图像处理模型包括编码器和解码器。解码器是与用于替换“待替换脸部”的目标脸部是一一对应的，因此图像处理模型中解码器的数量取决于图像处理模型需要处理的不同目标脸部（例如不同人脸）的数量。例如当图像处理模型需要对视频中的待替换脸部替换为2个不同的目标人脸时，图像处理模型中需要对应2个不同的目标

人脸设置解码器。

图2为本申请实施例提供的编码器的示例性结构图，如图2所示，包括4个卷基层、2个全连接层、1个重整形层和1个上采样层，各层的示例性参数如表1所示；当然，编码器中层类型和通道数进行各种变换，不局限于图1和表1所示。

层类型	通道数	卷积核	步长	填充	激活函数
输入层	3	-	-	-	-
卷积层 1	128	(5,5)	(2,2)	相同像素	参数整流线性单元 (pReLU, Parametric Rectified Linear Unit)
卷积层 2	256	(5,5)	(2,2)	相同像素	pReLU
卷积层 3	512	(5,5)	(2,2)	相同像素	pReLU
卷积层 4	1024	(5,5)	(2,2)	相同像素	pReLU
全连接层 1	1024	(5,5)	-	-	pReLU
全连接层 2	16384	(5,5)	-	-	pReLU
重整形层	-	(4,4,1024)	-	-	-
上采样层 1	512	(3,3)	(2,2)	相同像素	pReLU

5

表 1

图3为本申请实施例提供的解码器的示例性结构图，如图3所示，包括3个上采样层和1个卷积层，解码器各层的示例性参数如表2所示，当然，解码器中的层类型和通道数进行各种变换，不局限于图3和表2所示。

层类型	通道数	卷积核	步长	填充	激活函数
输入层	512	-	-	-	-
上采样层 1	256		(2,2)	相同像素	pReLU

—8—

上采样层 2	128	(3,3)	(2,2)	相同像素	pReLU
上采样层 3	64	(3,3)	(2,2)	相同像素	pReLU
卷积层 4	3	(3,3)	(2,2)	相同像素	-

表 2

图4为本申请实施例提供的图像处理模型进行换脸的原理示意图，编码器和解码器被训练完成后，编码器从原始图像的待替换人脸中提取风格特征（也就是对待替换人脸进行编码），风格特征被输入解码器进行解码，这个解码的过程是一种人脸转换，形成新的换脸图像中包括目标人脸的五官和待替换人脸的风格，例如表情及神态。

图5为本申请实施例提供的图像处理模型进行换脸的原理示意图， x 为输入的待替换脸部，则编码器针对 x 的编码函数可以表示为： $f(h/x)$ ，对应的编码结果为 $f(h/x) * x$ ，从而将 x 编码为表示风格的隐性的特征 h ，特征 h 的维度低于 x ，因此也可以视为以一种压缩的方式来表达待替换脸部。解码器的解码函数可以表示为 $g(x/h)$ ，根据输入的特征 h 进行解码，解码得到重构的待替换脸部 r 可以表示为 $g(x/h) * h$ 。

当图像处理模型包括两个以上的解码器（图1中示例性示出了解码器A和解码器B）时，解码器共用同一个编码器。

例如，假设解码器A对应的目标脸部是对象A的脸部，解码器B对应的目标脸部是对象B的脸部，对于编码器针对同一个原始图像中的待替换脸部（对象C的脸部）编码得到的风格特征的，解码器A可以基于风格特征进行解码，以重构得到具有相同风格特征的目标脸部A，而解码器B可以基于风格特征进行解码，以重构得到具有相同风格特征的目标脸部B。

上述的对象可以是任意具有五官的生物（包括人和动物），相应的，对象的脸部可以是人脸、动物脸等。接下来以人脸为例，继续对本申请实施例提供的图像处理模型的处理过程进行说明。

图6为本申请实施例提供的图像处理模型应用相关技术提供的训练方法的原理示意图，以训练目标人脸为分别是人脸A和人脸B的解码器为例，编码器和解码器A组成自编码网络A，基于人脸A的不同样本图像（例如不同分辨率

和不同风格)的扭曲图像,对自编码网络A进行无监督地训练,以更新编码器和解码器A的参数,直至解码器能够对面脸A进行编码得到人脸A的降维后的特征,解码器A用于基于人脸A的降维后的特征进行解码,以重构人脸A的图像。同理,基于人脸B的不同样本图像的扭曲图像,对自编码网络B进行无监督地训练,以更新编码器和解码器B的参数,直至解码器能够对面脸B进行编码得到人脸B的降维后的特征,解码器B用于基于人脸B的降维后的特征进行解码,以重构人脸B的图像。

可见,对于图1示出的图像处理模型的训练来说,应用相关技术提供的训练方法涉及到对自编码网络A和自编码网络B的训练,在每个自编码网络的训练过程中都需要分别对编码器和解码器的参数进行更新,下面进行说明。

对自编码网络A的训练过程中,自编码网络A的损失函数表示为:

$\text{loss_A} = \sum (\text{decoder_A}(\text{encoder}(\text{warp}(x_1))) - x_1)^2$; 其中, decoder_A 为解码器A, warp 为扭曲人脸的函数, x_1 为人脸A, encoder 为编码器。

在迭代训练的过程中,通过将人脸A代入自编码网络A的损失函数,求解损失函数按照梯度(例如最大梯度)下降时编码器和解码器A的参数,当损失函数收敛时,结束训练。

对自编码网络B的训练过程中,自编码网络B的损失函数表示为: $\text{loss_B} = \sum (\text{decoder_B}(\text{encoder}(\text{warp}(x_2))) - x_2)^2$; 其中, decoder_B 为解码器B, warp 为扭曲人脸的函数, x_2 为人脸B, encoder 为编码器。

在迭代训练的过程中,通过将人脸B代入自编码网络B的损失函数,求解损失函数按照梯度(例如最大梯度)下降时编码器和解码器B的参数;当损失函数收敛时,结束训练。

综合上述分析,应用相关技术对本申请实施例提供的图像处理模型训练存在以下问题:

25 1、图像处理模型的调整代价大

在形成针对目标人脸是人脸A、人脸B(目标人脸)的图像处理模型之后,如果还要使图像处理模型能够处理目标人脸是人脸C、人脸D进行换脸,就需要利用人脸C、人脸D的样本,对编码器分别和人脸C、人脸D构成的自编码网络的参数进行调整,由于编码器与解码器串联,梯度从解码器传播到编码器时

会变得很小，因此需要进行多次迭代训练，消耗时间较多。

2、训练的数据量要求大

由于编码器的参数众多，需要在训练过程中为编码器的训练提供众多的训练样本，例如当需要把待替换人脸换为目标人脸时，需要采集目标人脸的大量的样本，包括不同角度、表情和光照等，往往难以获取如此丰富的样本，进而影响了图像处理模型换脸的效果。

针对上述问题，本申请实施例提供了一种图像处理模型的训练方法和训练装置，以及图像处理模型的图像处理方法和图像处理装置。作为示例，图7为本申请实施例提供的图像处理模型的训练方法和应用方法的使用场景示意图；为实现支撑一个示例性应用，实现本申请实施例图像处理模型的训练装置30可以是服务器，实现本申请实施例的图像处理装置10可以是运行各种客户端的终端，两者通过网络20连接，其中网络20可以是广域网或者局域网，又或者是二者的组合，使用无线链路实现数据传输。图像处理装置10提交针对不同目标脸部的训练任务，由训练装置30训练相应的图像处理模型，并下发图像处理模型

的参数到图像处理装置10。例如，图像处理装置10的图形界面中可以运行视频客户端，视频客户端能够根据用户在播放界面通过各种人机交互方式（例如手势、语音等）指示的待替换人脸120以及目标人脸110，向训练装置30提交对应的训练任务，由训练装置30完成图像处理模型的训练任务并向视频客户端下发对应的参数，视频客户端通过运行图像处理模型重构与待替换人脸120具有相同风格的目标人脸130，在解码后出现待替换人脸120的帧图像中，利用重构的目标人脸130替换视频中的待替换人脸120，以形成视频的换脸效果。

可以理解地，上述的视频解码的处理也可以迁移到服务器中，借助服务器的硬件资源对替换后的帧图像进行重新编码，形成具有换脸效果的视频。

继续参考附图描述实现本申请实施例的训练装置的硬件结构，实现本申请实施例的训练装置可以为服务器或各种类型的终端。参考图8，图8为本申请实施例提供的用于训练图像处理模型的训练装置30的结构示意图，结合图8示出的结构进行说明。

本申请实施例提供的训练装置30可以包括处理装置301（例如中央处理器、

图形处理器等), 其可以根据存储在只读存储器 (ROM, Read Only Memory) 302 中的程序或者从存储装置 308 加载到随机访问存储器 (RAM, Random Access Memory) 303 中的程序而执行各种适当的动作和处理。作为程序的示例, 可以包括编码器训练模块 3081 和解码器训练模块 3082。

5 在 RAM 303 中, 还存储有训练装置 30 操作所需的各种程序和数据。处理装置 301、ROM 302 以及 RAM 303 通过总线 304 彼此相连。输入/输出 (I/O) 接口 306 也连接至总线 304。

10 通常, 以下装置可以连接至 I/O 接口 305: 包括例如触摸屏、触摸板、键盘、鼠标、摄像头、麦克风、加速度计、陀螺仪等的输入装置 306; 包括例如液晶显示器 (LCD, Liquid Crystal Display)、扬声器、振动器等的输出装置 307; 包括例如磁带、硬盘等的存储装置 308。

通信装置 309 可以允许训练装置 30 与其他设备进行无线或有线通信以交换数据。虽然图 8 示出了具有各种装置的训练装置 30, 但是应理解的是, 并不要求实施或具备所有示出的装置。可以替代地实施或具备更多或更少的装置。

15 结合图 8 示出的训练装置 30 说明本申请实施例提供的图像处理模型的图像处理方法, 参见图 9, 图 9 为本申请实施例提供的图像处理模型的训练方法的流程图示意图, 可以理解地, 图 9 所示的步骤可以由运行于终端中的训练装置 30 执行, 例如训练装置 30 可以是终端, 或者是可以耦合到终端内部/外部接口中的功能模块; 图 9 所示的步骤也可以由运行于服务器中的训练装置 30 执行, 例如
20 训练装置 30 可以是服务器, 或者是可以耦合到服务器内部/外部接口的功能模块。下面针对图 9 示出的步骤进行说明。

步骤 101, 根据原始图像中的待替换脸部, 更新图像处理模型中编码器的参数, 以使编码器能够对待替换脸部进行编码, 得到待替换脸部的风格特征。

25 图像处理模型在未进行任何训练之前, 其初始结构仅包括一个编码器和一个解码器, 当解码器未针对目标脸部进行训练前解码器也称为临时解码器。当针对目标脸部在图像处理模型中训练对应的解码器之前, 需要首先对编码器进行初始化 (即更新编码器的参数)。

在一些实施例中, 将编码器和临时解码器组成自编码网络, 通过对自编码网络进行无监督训练的方式来更新编码器的参数, 由于自编码网络具有这样的

特性：由编码器对输入的脸部图像进行编码（也就是压缩）较输入的图像更低维度的特征，这些特征来表示脸部本身在风格上的视觉特性而不是噪声；然后由解码器基于特征来最大程度还原（也就是重构）输入的脸部图像。因此，通过训练自编码网络可以使编码器具有从原始图像的待替换脸部中提取风格特征的能力。

例如，自编码网络的损失函数用于量化表示以下两者的差异：扭曲原始图像中的五官图像；解码器重构的五官图像。从而，将扭曲原始图像（扭曲原始图像是对原始图像进行扭曲处理得到）中待替换脸部的五官图像，代入由编码器和解码器构成的自编码网络对应的损失函数时，可以确定损失函数满足收敛条件时对应编码器的参数，收敛条件可以是损失函数的低于损失函数阈值，或者，迭代训练自编码网络的次数达到迭代次数阈值。

步骤102，根据原始图像的待替换脸部，更新图像处理模型中解码器的参数，以使解码器能够基于待替换脸部的风格特征进行解码。

如前所述，在编码器的初始化阶段，由于是通过将编码器和临时解码器构成的自编码网络进行训练，因此在自编码网络的训练过程中，解码器的参数也同步进行了更新，此时，解码器具有基于待替换脸部的风格特征进行解码时得到重构的待替换脸部的性能。也就是将扭曲原始图像中待替换脸部的五官图像，代入由编码器和解码器构成的自编码网络对应的损失函数，确定损失函数满足收敛条件时对应编码器的参数时，还可以同步确定损失函数满足收敛条件时临时解码器的参数。

在编码器和临时解码器组成的自编码网络训练完成后，编码器已经具有从输入的脸部图像提取风格特征的能力，即使需要训练临时解码器具有从风格特征还原不同目标脸部的能力时，编码器的参数也不需要继续进行参数调整，样本和训练时间的节约从整体上提升了后续利用目标脸部的样本训练解码器的效率。

为了提升模型训练的效率，当根据原始图像中的待替换脸部，更新图像处理模型中编码器的参数之前，还可以针对原始图像的待替换脸部进行各种初始化处理。

在一些实施例中，初始化处理可以包括：在数据采集阶段，对于采集的包

括待替换脸部的样本图像集合,从各个样本图像中截取包括待替换脸部的原始图像;将所截取的各个原始图像基于待替换脸部的特征点进行对齐。从而能够使编码器能够准确学习脸部的风格特征。

5 在一些实施例中,初始化处理可以包括对原始图像进行随机增广处理:将经过随机增广处理的原始图像进行随机扭曲处理,得到扭曲的原始图像;截取扭曲的原始图像中待替换脸部的五官图像,并对截取的五官图像进行缩放处理,以用于更新所述编码器的参数时适配编码器支持输入的分辨率。

10 作为示例,对原始图像进行随机增广处理,可以采用以下方式:基于原始图像的中心进行随机旋转;将随机旋转后的原始图像进行随机缩放;将随机缩放后原始图像进行随机平移。

作为示例,将经过随机增广处理的原始图像进行随机扭曲处理,可以采用以下方式:在经过随机增广处理的原始图像中插入噪声,噪声包括固定噪声(即固定的颜色值)和动态噪声(即变化的颜色值)中的至少一种。

15 对于在经过随机增广处理的原始图像中插入动态噪声来说,可以将经过随机增广处理的原始图像进行网格化,并为至少部分或者全部节点分配坐标,将坐标根据高斯分布添加噪声;并在节点之外的区域进行图像插值(例如线性插值或非线性插值)以放大原始图像的尺寸,当采用线性插值时,插值的像素的取值可以与坐标的取值正相关。

20 作为示例,从扭曲的原始图像中待替换脸部的五官图像,可以以扭曲的原始图像的中心为几何中心,以各种形状(例如圆形、椭圆形和长方形等)截取待替换脸部的一定比例的图像作为五官图像(通常,截取比例超出62.5%时能够保证截取到完整的五官图像)。

当编码器初始化完成后,可以继续针对不同的目标脸部有针对性地训练临时解码器,将在步骤103中说明。

25 步骤103,在保持编码器的参数不变的情况下,根据目标图像中的目标脸部更新解码器的参数,以使解码器能够基于待替换脸部的风格特征进行解码,得到与待替换脸部具有相同风格的目标脸部。

可以理解地,根据上文记载的针对原始图像的初始化处理,可以针对目标图像进行类似的处理,例如对目标图像进行类似的初始化处理,进行随机增广

处理,将经过随机增广处理的目标图像进行随机扭曲处理,得到扭曲的目标图像;截取扭曲的目标图像中的五官图像。

在一些实施例中,当编码器初始化完成后,如果需要训练对应目标脸部的解码器,在保持临时编码器的参数不变的情况下,通过扭曲目标图像(即针对目标图像进行扭曲处理得到)中目标脸部的五官图像训练编码器和临时解码器组成的自编码网络,以继续更新临时解码器的参数,直至自编码网络的损失函数满足收敛条件。

需要指出地,针对不同目标脸部训练对应的解码器时,都是在编码器初始化完成后临时解码器的参数的基础上进行更新,以使临时解码器不会学习同时恢复多个目标脸部的能力,从而确保重构目标脸部的精度。

以训练目标为图1示出的图像处理模型中的解码器A为例,解码器A是用于重构目标脸部A的,将与扭曲目标脸部A中目标脸部的五官图像,代入由编码器和临时解码器构成的自编码网络对应的损失函数,在保持编码器的参数不变的情况下,确定损失函数满足收敛条件时临时解码器的参数,从而作为与目标脸部对应的解码器A的参数,由于解码器A的结构是先验的,因此确定了解码器A的参数也就相当于确定了解码器A。

同理,训练目标为图1示出的图像处理模型中的解码器B时,解码器B是用于重构目标脸部B的,将与扭曲目标脸部B中目标脸部的五官图像,代入由编码器和临时解码器构成的自编码网络对应的损失函数,在保持编码器的参数不变的情况下,确定损失函数满足收敛条件时临时解码器的参数,从而作为与目标脸部对应的解码器B的参数。

在一些实施例中,当图像处理模型训练完成之后,还可以应用图像处理模型进行如图7示出的换脸的操作,通过编码器对待替换人脸120的图像编码得到待替换人脸120的风格特征,将待替换人脸120的风格特征输入到针对目标人脸110训练的解码器;在目标人脸110的解码器中基于待替换人脸120的风格特征进行解码,得到与待替换人脸120具有相同风格的目标人脸130,从而实现了换脸的效果。

继续参考附图描述实现本申请实施例的训练装置的结构,实现本申请实施例的训练装置可以为服务器或各种类型的终端。

参考图10,图10为本申请实施例提供的图像处理装置10的一个可选的结构示意图。根据上文针对图8的说明,可以理解本申请实施例中的图像处理装置10包括的处理装置101、ROM 102、RAM 103、总线104、I/O接口105、输入装置106和输出装置107。此外,作为存储装置108中存储的程序的示例,可以包括编码器模块1081和解码器模块1082。

结合图10示出的图像处理装置10说明本申请实施例提供的图像处理模型的图像处理方法,参见图11,图11为本申请实施例提供的图像处理模型的图像处理方法的流程示意图,可以理解地,图11所示的步骤可以由运行于终端中的图像处理装置10执行,其中图像处理装置10可以是终端,或者是可以耦合到终端内部/外部接口中的功能模块;图11所示的步骤也可以由运行于服务器中的图像处理装置10执行,其中图像处理装置10可以是服务器,或者是可以耦合到服务器内部/外部接口的功能模块。下面针对图11示出的步骤进行说明。

步骤201,通过图像处理模型的编码器,对原始图像进行编码得原始图像中待替换脸部的风格特征。

步骤202,在图像处理模型的解码器中基于待替换脸部的风格特征进行解码,得到与待替换脸部具有相同风格的目标脸部。

根据上文的图像处理模型的训练方案可以看出,由于解码器的参数是在保持编码器的参数不变的情况下,根据目标图像风格特征进行更新得到的,因此当需要图像处理模型中需要增加对应新的目标脸部的解码器时,不需要继续训练编码器,一方面不需要针对编码器而设置对应目标脸部的大量样本图像,降低了图像处理模型的训练时间;另一方面,针对目标脸部只针对性训练对应的解码器,进一步节约了训练时间。

继续说明如图8示出的编码器训练模块3081和解码器训练模块3082。

编码器训练模块3081,用于根据原始图像中的待替换脸部,更新图像处理模型中编码器的参数,以使编码器能够对待替换脸部进行编码,得到待替换脸部的风格特征。

解码器训练模块3082,用于根据原始图像的待替换脸部,更新图像处理模型中解码器的参数,以使解码器能够基于待替换脸部的风格特征进行解码;保持编码器的参数不变,并根据目标图像中目标脸部的风格特征更新解码器的参

数,以使解码器能够基于待替换脸部的风格特征进行解码得到与所述待替换脸部具有相同风格的目标脸部。

在一些实施例中,编码器训练模块3081具体用于:将扭曲原始图像中待替换脸部的五官图像,代入自编码网络对应的损失函数,确定损失函数满足收敛条件时对应编码器的参数,此时解码器基于待替换脸部的风格特征进行解码时能够得到重构的待替换脸部。扭曲原始图像是对原始图像进行扭曲处理得到。

在一些实施例中,解码器训练模块3082具体用于:将扭曲原始图像中待替换脸部的五官图像,代入自编码网络对应的损失函数,确定损失函数满足收敛条件时对应解码器的参数,其中扭曲原始图像是对原始图像进行扭曲处理得到。

在一些实施例中,解码器训练模块3082具体用于:当图像处理模型中包括至少两个解码器,且解码器各自与不同的目标脸部对应时,通过编码器对目标图像中目标脸部的五官图像进行编码,得到与解码器对应的目标脸部的风格特征;将与解码器对应的目标脸部的风格特征,代入由编码器和解码器构成的自编码网络对应的损失函数,确定损失函数满足收敛条件时对应解码器的参数。

在一些实施例中,编码器训练模块3081还用于:当根据原始图像中的待替换脸部,更新图像处理模型中编码器的参数之前,从样本图像集合的各个样本图像中截取包括待替换脸部的原始图像;将所截取的各个原始图像基于待替换脸部的特征点进行对齐。

在一些实施例中,编码器训练模块3081还用于:当根据原始图像中的待替换脸部,更新图像处理模型中编码器的参数之前,对原始图像进行随机增广处理;将经过随机增广处理的原始图像进行随机扭曲处理,得到扭曲的原始图像;截取扭曲的原始图像中待替换脸部的五官图像,并对截取的五官图像进行缩放处理以用于更新所述编码器的参数时适配编码器支持输入的分辨率。

在一些实施例中,编码器训练模块3081,具体用于:基于原始图像的中心进行随机旋转;将随机旋转后的原始图像进行随机缩放;将随机缩放后原始图像进行随机平移。

在一些实施例中,编码器训练模块3081,具体用于:在经过随机增广处理的原始图像中插入噪声,噪声包括固定噪声和动态噪声中的至少一种。

在一些实施例中，编码器训练模块3081，具体用于：将经过随机增广处理的原始图像进行网格化，并为网格中的至少部分节点分配坐标；在至少部分节点中添加与所分配的坐标相对应的噪声。

5 在一些实施例中，解码器训练模块3082，具体用于：对目标图像进行随机增广处理；将经过随机增广处理的目标图像进行随机扭曲处理，得到扭曲的目标图像；截取扭曲的目标图像中的五官图像。

10 在一些实施例中，解码器训练模块3082，还用于调用编码器执行以下处理：对待替换图像进行编码，得到待替换图像中待替换脸部的风格特征，将风格特征输入解码器；解码器训练模块3082，还用于调用解码器执行以下处理：基于待替换脸部的风格特征进行解码，得到与待替换脸部具有相同风格的目标脸部。

15 继续说明如图10示出的编码器模块1081和解码器模块1082。编码器模块1081，用于对原始图像进行编码得到所述原始图像中待替换脸部的风格特征。解码器模块1082，用于基于所述待替换脸部的风格特征进行解码，得到与所述待替换脸部具有相同风格的目标脸部；其中，解码器模块1082是在保持编码器模块1081的参数不变的情况下，根据包括目标脸部的目标图像更新得到。

20 本申请实施例还提供一种计算机可读存储介质，所述计算机可读存储介质用于存储计算机程序。计算机可读介质可以是上述训练装置30/图像处理装置10中所包含的；也可以是单独存在，而未装配入该训练装置30/图像处理装置10中。上述计算机可读介质承载有一个或者多个计算机程序，当上述一个或者多个计算机程序被训练装置30/图像处理装置10执行时，使得训练装置30/图像处理装置10能够实现本申请实施例提供的图像处理模型的训练方法/图像处理方法。

25 下面以训练并应用交换人脸的图像处理模型（也可以称为换脸模型）为例说明。

参见图12，图12为本申请实施例提供的用于交换人脸的图像处理模型的训练原理示意图。图像处理模型在未进行任何训练之前，其初始结构仅包括一个编码器和一个临时解码器。

训练时首先进行编码器的初始化。对于包括人脸的图像，通过人脸检测的

方式来截取人脸图像，利用人脸图像作为样本，训练由编码器和临时解码器组成的自编码网络，使得自编码网络对人脸图像的样本编码（即特征提取/降维）后，能够利用编码得到的风格特征重构人脸图像。

在编码器初始化的过程中，临时解码器的参数也得以同步更新。在编码器初始化后，从人脸图像中识别出的不同人脸可以作为目标人脸以训练对应的解码器，不同目标人脸对应的解码器都是以编码器初始化后临时解码器的参数为基础训练得到。

以训练人脸 i 对应的解码器 i 为例（ i 为大于或等于1的整数），通过保持编码器的参数不变，以人脸 i 图像作为样本，对编码器和临时解码器组成的自编码网络进行训练，以更新临时解码器的参数，训练完成后临时解码器的参数即作为解码器 i 的参数，也即得到解码器 i 。

下面再分别对编码器初始化、训练针对目标人脸的解码器以及相应的应用进行说明。

参考图13A，图13A为本申请实施例提供的图像处理模型的训练方法的原理示意图，结合图13A进行说明。

初始化编码器之前首先需要进行人脸图像数据的获取，可以采用如下方式：采用人脸检测算法，框选人脸位置；使用五官定位算法、标出人眼、嘴、鼻、等脸部的特征点；根据检测到的人脸位置截取人脸图像，并将截取的人脸图像基于特征点（例如人眼）进行对齐，人脸图像的示例性分辨率可以为512*512（像素）。

人脸图像在初始化编码器之前还进行各种初始化处理，包括随机增广、随机扭曲和五官截取，下面分别进行说明。

随机增广处理，包括：将人脸图像以图像中心为中心，做随机旋转，旋转的角度范围为 α ，例如 α 的取值范围为 0° 至 30° ；将旋转后的人脸图像的宽高进行随机拉伸，拉伸的倍数在 $[1-\beta, 1+\beta]$ 之间，例如取值为0.01；随机平移，将人脸图像在竖直方向和水平方向对应随机平移 θ_1 和 θ_2 ，例如 θ_1 、 θ_2 的值对应人脸图像的宽高的1%。

随机增广处理后的人脸图像还进行随机扭曲处理，可以采用如下方式：绘制网格；添加坐标噪声；图像重映射。下面分别进行说明。

绘制网格时, 参见图13B, 图13B为本申请实施例提供的在人脸图像40中绘制网格50的示意图, 假设截取的人脸图像40为正方形图像, 记为I, 宽高均为W, 在人脸图像40中绘制密度为N (表示网格中节点的整数) *N的均匀网格50, 示例的网格的密度为5*5 (即宽度方向5个节点, 高度方向5个节点), 在图13B中示例性示出了节点51至节点55。

添加坐标噪声时, 将每个节点分配坐标 (x, y), 记第i行第j列 (在图13B中, i和j的取值范围均在1至5之间) 的坐标分别为 (x_{ij}, y_{ij}), x_{ij}=i/N, y_{ij}=j/N, 则所有的x_{ij}(i=1,...,5, j=1,...,5)、y_{ij}(i=1,...,5, j=1,...,5)组成矩阵 (X, Y)。将 (X, Y) 的每个元素分别加上独立分布的高斯噪声, 高斯噪声的期望为0, 方差为σ², 方差的示例性取值为0.02, 将矩阵 (X, Y) 使用图像插值的方法 (例如使用线性插值), 放大矩阵 (X, Y) 的宽高到W, 得到新的图像 (即矩阵) (X', Y')。构造宽高均为W的矩阵 (X2, Y2), 其中矩阵X2的元素x_{2ij}=i/W, 矩阵Y2中元素y_{2ij}=j/W, 其中i=1,...,W, j=1,...,W。

图像重映射时, 对添加坐标噪声的人脸图像进行插值得到扭曲后的人脸图像Z。记Z第a行第b列的像素为z_{ab}, 查找矩阵X2中比x_{ab}小的元素, 记为x_{2c}, 查找矩阵Y2中比y_{ij}小的元素记为x_{2d}, 则可以构造如下四个点: A (x_{2c,d}, y_{2c,d}), B (x_{2c+1,d}, y_{2c+1,d}), C (x_{2c,d+1}, y_{2c,d+1}), D (x_{2c+1,d+1}, y_{2c+1,d+1})。分别计算点 (x_{ab}, y_{ab}) 到A、B、C、D的距离dA、dB、dC和dD, 根据这四个距离加权得到z_{ab}的值: z_{ab}=dA/(dA+dB+dC+dD)×I_{c,d}+ dB/(dA+dB+dC+dD)×I_{c+1,d}+ dC/(dA+dB+dC+dD)×I_{c,d+1}+ dD/(dA+dB+dC+dD)×I_{c+1,d+1}。

随机扭曲后的人脸图像需要进行五官图像截取才能初始化编码器, 以截取扭曲的人脸图像中间固定位置的图像, 截取的示例性的百分比为62.5%。之后将图像缩放到编码器支持输入的分辨率。以上述随机扭曲图像宽高为512*512为例, 截取宽高行数在 (48, 208) 内的五官图像, 然后将得到的161*161的五官图像缩小到64*64的五官图像。

在初始化编码器的阶段, 对于数据集中检测到的所有人脸图像进行上述随机增广、随机扭曲和五官截取的初始化处理以得到五官图像的样本, 用来训练如图13A示出的自编码网络来初始化编码器, 在图13A中编码器的示例性的结构如图2和表1所示, 当然不限于此, 例如可以对卷积层、全连接层和上采样层

的通道以及数量进行灵活变化。

自编码网络的训练目标是将输入的扭曲的五官图像的特征进行编码后,能够利用编码后的特征重构扭曲的五官图像。例如,编码器的网络的输出只有1024的神经元,而输入的扭曲人脸有12288个维度,通过训练,编码器能够尽可能保留输入的扭曲五官图像的特征,使得在之后训练换脸模型时,编码器的参数无需再进行微调。

记编码器为 $\text{Encoder}(x)$, x 为输入的扭曲的五官图像,临时解码器为 $\text{Decoder_tmp}(f)$, f 为 x 经编码器编码后得到的输出(即 $\text{Encoder}(x)$),则损失函数 Loss_init 可以表示为: $\text{loss_init} = \sum (\text{Decoder_tmp}(\text{Encoder}(x)) - x)^2$ 。

当 loss_init 再优化时不再有显著下降(即小于损失函数阈值)时,或者当迭代次数达到迭代次数阈值时,编码器初始化完成,编码器已经具有从输入的任意人脸图像中编码得到风格特征的能力,从而后续不需要进行参数的进一步调整。

参见图14,图14为本申请实施例提供的初始化编码器的原理示意图,人脸图像61、人脸图像63进行上述的初始化处理后,人脸图像61、人脸图像63的扭曲的五官图像被用于训练编码器和临时解码器组成的自编码网络,在训练完成后,临时解码器能够基于编码器编码得到的特征重构人脸图像61、人脸图像63中被扭曲的五官图像。

下面参考图15,图15为本申请实施例对解码器的训练原理示意图,结合图14说明解码器的训练过程。

以训练针对用户 i 的解码器为例,首先获取用户 i 的人脸数据。可以采用算法自动,或者人工分类的方法,选取用户 i 的150张以上的人脸图像。根据上文所说明的初始化方案进行随机增广、随机扭曲和五官图像截取,得到扭曲的五官图像。

使用用户 i 的扭曲五官图像训练编码器与解码器 i (解码器 i 的初始参数与临时解码器的参数相同)组成的自编码网络,在训练的过程中,由于无需再对编码器的参数进行更新。

记编码器为 $\text{Encoder}(x)$, x 为输入的扭曲的五官图像,解码器 i 为 $\text{Decoder}_i(f)$,其中 f 为 x 经过编码器编码后得到的特征,使用如下的损失函数

-21-

loss_i, 更新编码器Decoder_i的参数: $\text{loss}_i = \sum (\text{Decoder}_i(\text{Encoder}(x)) - x)^2$ 。

当loss_i不再有显著下降(即小于损失函数阈值)时, 或者当迭代次数达到迭代次数阈值时, 解码器i训练完成。

参见图16, 图16为本申请实施例图像处理模型进行换脸的原理示意图, 图
5 图像处理模型中包括针对用户A训练的解码器A以及针对用户B训练的解码器B。

当用户A的人脸图像71被输入到编码器时, 编码器能够从中提取用户A的人脸在图像71中的风格特征, 当风格特征被输入到解码器B时, 由于解码器B已经通过训练得到了基于风格特征重构用户B的人脸图像的能力, 因此解码器
10 B能够基于风格特征重构具有相同风格特征的用户B的人脸图像74。同理, 当用户B的图像73被输入到编码器时, 编码器能够从中提取用户B在图像73中的风格特征, 当风格特征被输入到解码器A时, 由于解码器A已经通过训练得到了基于风格特征重构用户A的人脸图像的能力, 解码器A能够基于风格特征重构具有相同风格特征的用户A的人脸图像72。

15 继续参见图7所示的应用场景, 图像处理装置10可以运行于终端中, 训练装置30可以运行于用以执行训练任务的服务器中。用户希望将终端的视频客户端中播放的视频中某一人物的脸部(即待替换人脸120)替换用户自己的脸部(即目标人脸110), 在视频客户端中播放具有换脸效果(即目标人脸130)的视频。

20 为此, 用户已经向服务器上传了自己的脸部图像(即目标人脸), 并在一个视频中选定了角色, 期望替换该角色的脸部(即待替换人脸), 服务器通过本申请实施例提供的图像处理模型(也即换脸模型)的训练方法训练了编码器以及对应目标对象的解码器。

参见图17, 图17为本申请实施例提供的形成具有换脸效果的视频的流程
25 图, 结合示出的步骤说明服务器制作具有换脸效果的视频的过程。

步骤301, 视频解码。

视频会被解码得到一系列的帧图像, 下面以第n(n为大于或等于1的整数)帧图像为例说明。

步骤302, 提取第n帧图像。

- 22 -

步骤303, 对第n帧图像进行人脸检测, 如果有人脸, 则执行步骤304; 如果没有人脸, 则返回提取第n+1帧图像继续进行人脸检测。

步骤304, 对第n帧图像进行人脸识别, 以判断是否包括待替换人脸, 如果是则执行步骤305; 否则跳转到步骤308。

5 步骤305, 截取得替换人脸。

步骤306, 调用换脸模型重构目标人脸。

换脸模型中编码器对待替换人脸进行编码得到待替换人脸的风格特征, 目标人脸的编码器对风格特征进行解码, 得到具有风格特征的目标人脸。

步骤307, 将重构的目标人脸回贴第n帧中的待替换人脸。

10 步骤308, 编码视频。

当视频中的帧图像处理完成后, 将各帧视频按照视频的原始格式进行编码, 当然也可以编码为其他格式。

可以理解地, 上述制作视频的处理也可以迁移到终端中, 根据终端的硬件配置可以采取相适配的解码/编码参数。

15 通过本申请实施例所示的对图像处理模型的训练方法, 当用户希望调用图像处理模型交换视频图像帧中的人脸图像时, 仅需要重新对图像处理模型中的代替换图像的解码器进行训练, 当采用上文表1与表2所示的图像处理模型的数量时, 图像处理模型所要的编码器数量和解码器数量如表3所示, 其中, 编码器的参数量是解码器参数量的约11.24倍, 因此, 在保持编码器的参数不变, 20 并根据目标图像中目标脸部的特征更新解码器的参数的过程中, 能够有效节省对于图像处理模型的训练时间。

图像处理模型	编码器数量	解码器数量
参数量	69662976	6199747

表 3

综上所述, 本申请各实施例具有以下有益效果:

25 1、编码器只要一次初始化, 后续只需要针对性训练解码器, 从而需要优化的参数量大大降低, 训练速度变快。

2、由于需要的优化的参数量大大降低, 使得图像处理模型更不容易过拟

合，从而依赖的样本的数量降低，例如使用150张人脸即可训练换脸模型（无需原有模型需要的至少300张）。

3、由于图像处理模型的训练过程中，仅需要训练相应的待替换脸部的解码器，不需要重复训练编码器，因此，在相同的硬件条件下，能够同时训练更多的图像处理模型。

以上所述，仅为本申请的实施例而已，并非用于限定本申请的保护范围，凡在本申请的精神和原则之内所作的任何修改、等同替换和改进等，均应包含在本申请的保护范围之内。

本申请实施例还提供了一种包括指令的计算机程序产品，当其在服务器上运行时，使得服务器执行上述实施例提供的图像处理模型的训练方法或图像处理方法。

权 利 要 求

1、一种图像处理模型的训练方法，所述方法由训练装置执行，所述方法包括：

根据原始图像中的待替换脸部，更新图像处理模型中编码器的参数，以使
5 所述编码器能够对所述待替换脸部进行编码，得到所述待替换脸部的风格特征；

根据所述原始图像的待替换脸部，更新所述图像处理模型中解码器的参数，以使所述解码器能够基于所述待替换脸部的风格特征进行解码；

在保持所述编码器的参数不变的情况下，根据目标图像中的目标脸部更新
10 所述解码器的参数，以使所述解码器能够基于所述待替换脸部的风格特征进行解码，得到与所述待替换脸部具有相同风格的目标脸部。

2、根据权利要求 1 所述的方法，所述根据原始图像中的待替换脸部，更新图像处理模型中编码器的参数，包括：

将扭曲原始图像中待替换脸部的五官图像，代入自编码网络对应的损失函
15 数，确定所述损失函数满足收敛条件时对应所述编码器的参数，其中，所述自编码网络由所述编码器和所述解码器构成，所述扭曲原始图像是对所述原始图像进行扭曲处理得到。

3、根据权利要求 1 所述的方法，所述根据所述待替换脸部，更新所述图像处理模型中解码器的参数，包括：

20 将扭曲原始图像中待替换脸部的五官图像，代入所述自编码网络对应的损失函数，确定所述损失函数满足收敛条件时对应所述解码器的参数，其中所述扭曲原始图像是对所述原始图像进行扭曲处理得到。

4、根据权利要求 1 所述的方法，所述在保持所述编码器的参数不变的情况下，根据目标图像中目标脸部的风格特征更新所述解码器的参数，包括：

25 将扭曲目标图像中目标脸部的五官图像，代入所述自编码网络对应的损失函数；

在保持所述损失函数中对应所述编码器的参数不变的情况下，确定所述损失函数满足收敛条件时对应所述解码器的参数；其中，所述扭曲目标图像是对所述目标图像进行扭曲处理得到。

5、根据权利要求 1 所述的方法，在所述根据原始图像中的待替换脸部，更新所述图像处理模型中编码器的参数之前，所述方法还包括：

从样本图像集合的样本图像中截取包括所述待替换脸部的所述原始图像；
将所截取的多个所述原始图像基于待替换脸部的特征点进行对齐。

5 6、根据权利要求 1 所述的方法，在所述根据原始图像中的待替换脸部，更新所述图像处理模型中编码器的参数之前，所述方法还包括：

对所述原始图像进行随机增广处理；

将经过随机增广处理的原始图像进行随机扭曲处理，得到扭曲的所述原始图像；

10 截取扭曲的所述原始图像中待替换脸部的五官图像，并对所述截取的五官图像进行缩放处理，以用于更新所述编码器的参数时适配所述编码器支持输入的分辨率。

7、根据权利要求 6 所述的方法，所述对所述原始图像进行随机增广处理，包括：

15 基于所述原始图像的中心进行随机旋转；

将随机旋转后的所述原始图像进行随机缩放；

将随机缩放后所述原始图像进行随机平移。

8、根据权利要求 6 所述的方法，所述将经过随机增广处理的原始图像进行随机扭曲处理，包括：

20 在经过随机增广处理的原始图像中插入噪声，所述噪声包括固定噪声和动态噪声中的至少一种。

9、根据权利要求 8 所述的方法，所述在经过随机增广处理的原始图像中插入噪声，包括：

25 将所述经过随机增广处理的原始图像进行网格化，并为网格中的至少部分节点分配坐标；

在所述至少部分节点中添加与所分配的坐标相对应的噪声。

10、一种图像处理模型的图像处理方法，所述方法由图像处理装置执行，所述方法包括：

通过图像处理模型的编码器，对原始图像进行编码得到所述原始图像中待

替换脸部的风格特征;

在所述图像处理模型的解码器中基于待替换脸部的风格特征进行解码,得到与所述待替换脸部具有相同风格的目标脸部;

其中,所述解码器的参数是在保持所述编码器的参数不变的情况下,根据
5 包括所述目标脸部的目标图像进行更新得到。

11、一种图像处理模型的训练装置,所述训练装置包括:

编码器训练模块,用于根据原始图像中的待替换脸部,更新图像处理模型中编码器的参数,以使所述编码器能够对所述待替换脸部进行编码,得到所述待替换脸部的风格特征;

10 解码器训练模块,用于根据所述原始图像的待替换脸部,更新所述图像处理模型中解码器的参数,以使所述解码器能够基于所述待替换脸部的风格特征进行解码;

所述解码器训练模块,还用于在保持所述编码器的参数不变的情况下,根据目标图像中目标脸部的风格特征更新所述解码器的参数,以使所述解码器能够
15 够基于所述待替换脸部的风格特征进行解码,得到与所述待替换脸部具有相同风格的目标脸部。

12、一种图像处理装置,所述图像处理装置包括:

编码器模块,用于对原始图像进行编码得到所述原始图像中待替换脸部的风格特征;

20 解码器模块,用于基于所述待替换脸部的风格特征进行解码,得到与所述待替换脸部具有相同风格的目标脸部;

其中,所述解码器模块是在保持所述编码器模块的参数不变的情况下,根据包括所述目标脸部的目标图像更新得到。

13、一种图像处理模型的训练装置,所述训练装置包括:

25 存储器,用于存储可执行指令;

处理器,用于运行所述存储器存储的可执行指令时,实现权利要求1至9任一项所述的图像处理模型的训练方法。

14、一种图像处理装置,所述图像处理装置包括:

存储器,用于存储可执行指令;

处理器，用于运行所述存储器存储的可执行指令时，实现权利要求 10 所述的图像处理模型的图像处理方法。

15、一种计算机可读存储介质，所述计算机可读存储介质用于存储计算机程序，所述计算机程序用于执行权利要求 1 至 9 任一项所述的图像处理模型的
5 训练方法，或者执行权利要求 10 所述的图像处理模型的图像处理方法。

16、一种包括指令的计算机程序产品，当其在计算机上运行时，使得所述计算机权利要求 1 至 9 任一项所述的图像处理模型的训练方法，或者执行权利要求 10 所述的图像处理模型的图像处理方法。

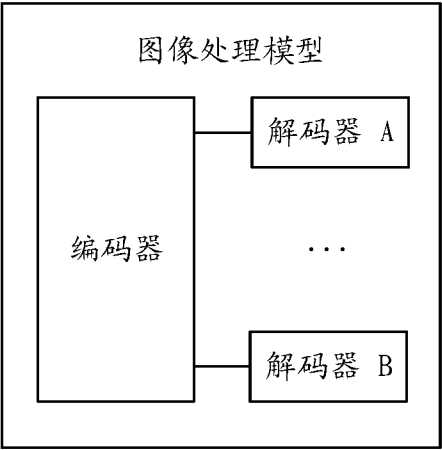


图 1

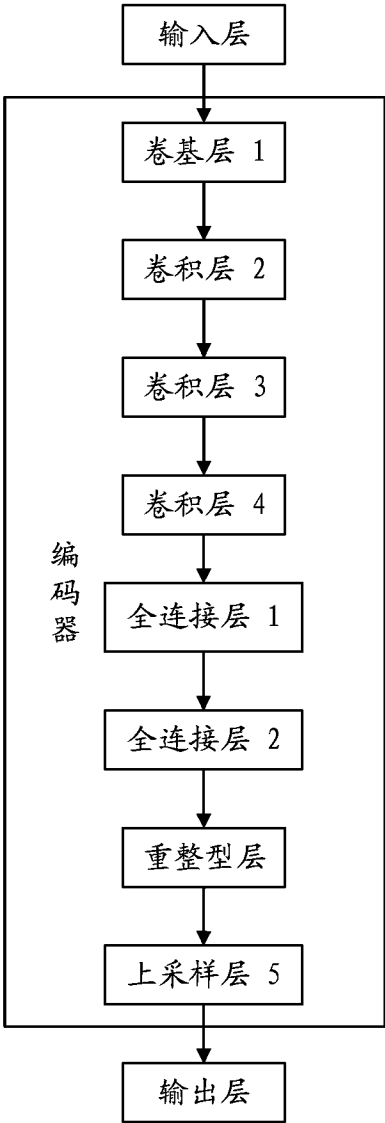


图 2

—2/11—

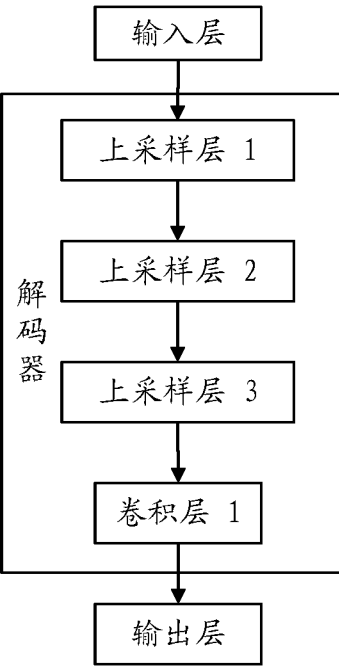


图 3

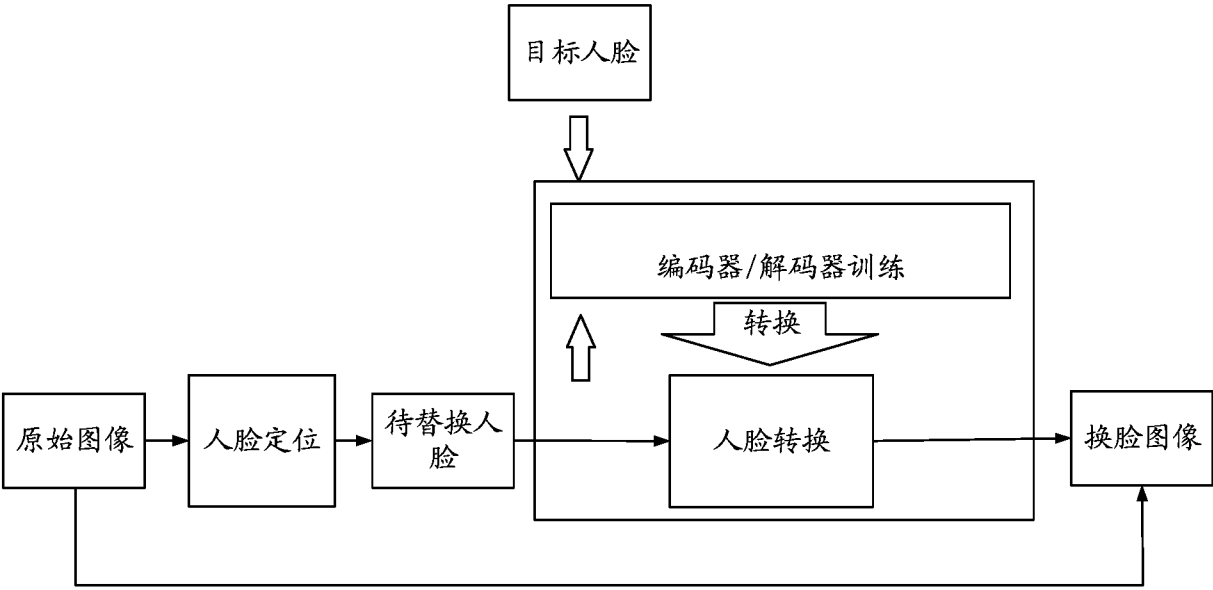


图 4

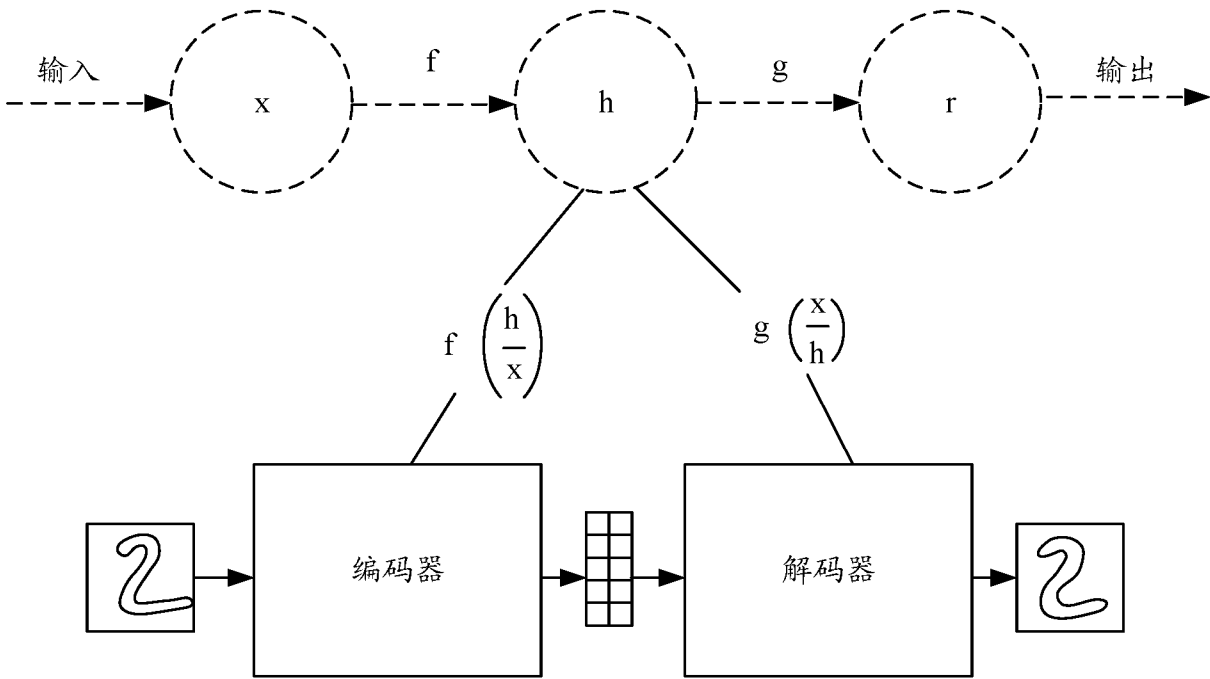


图 5

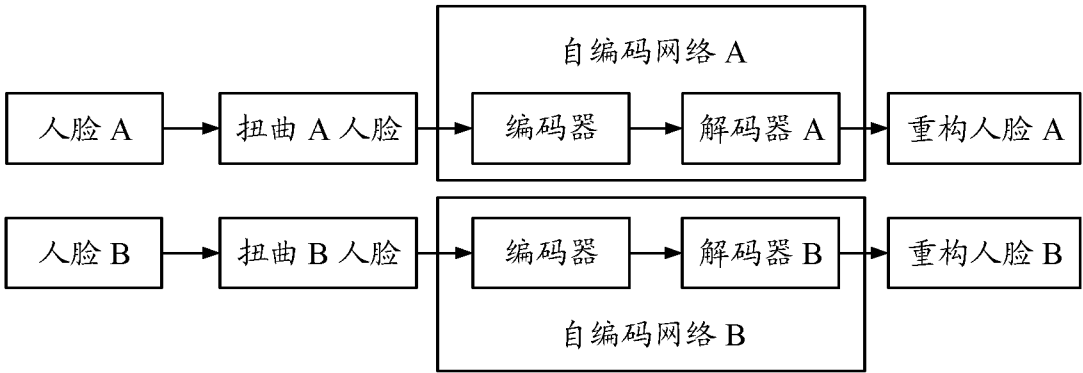


图 6

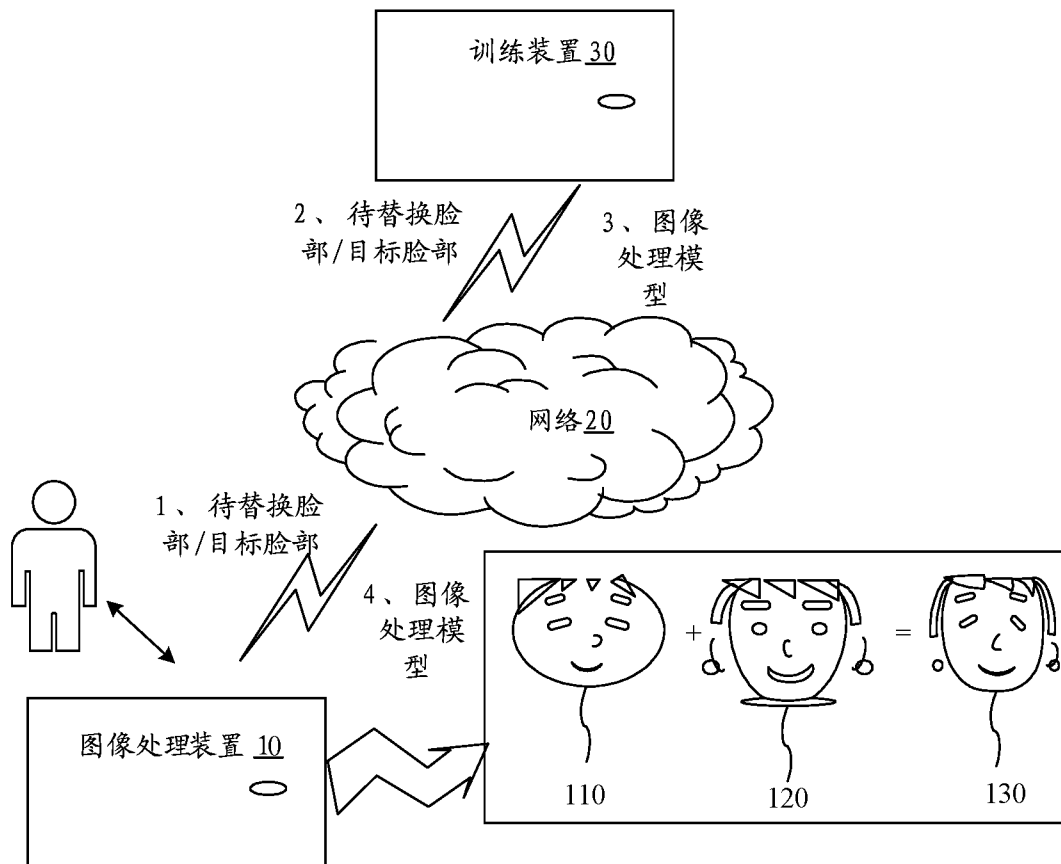


图 7

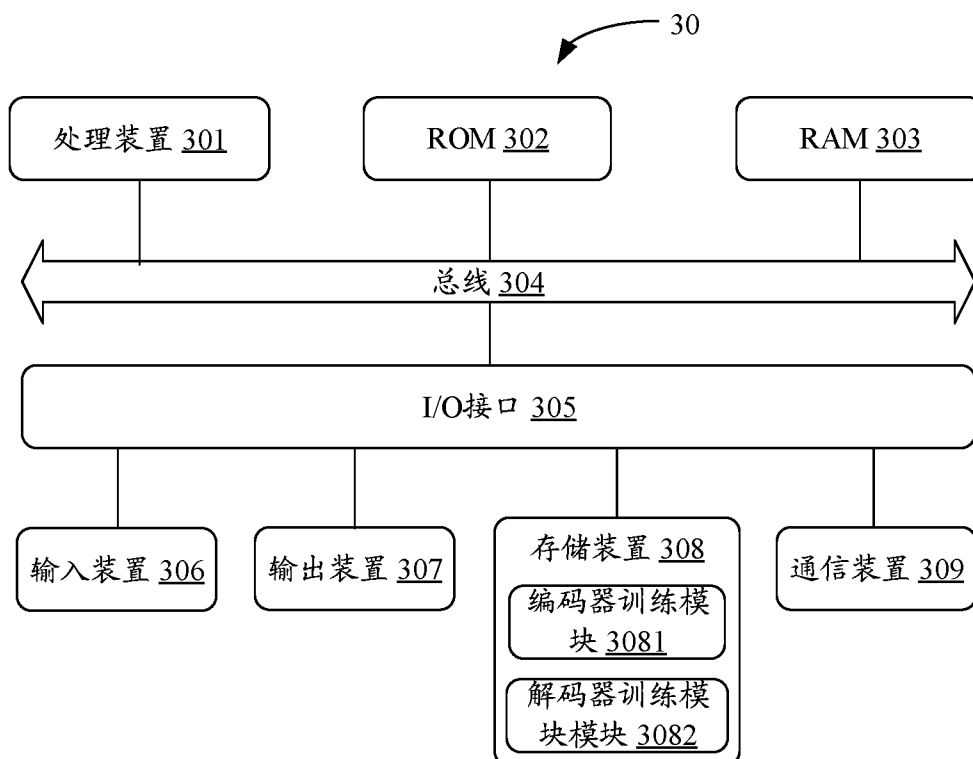


图 8

— 5/11 —

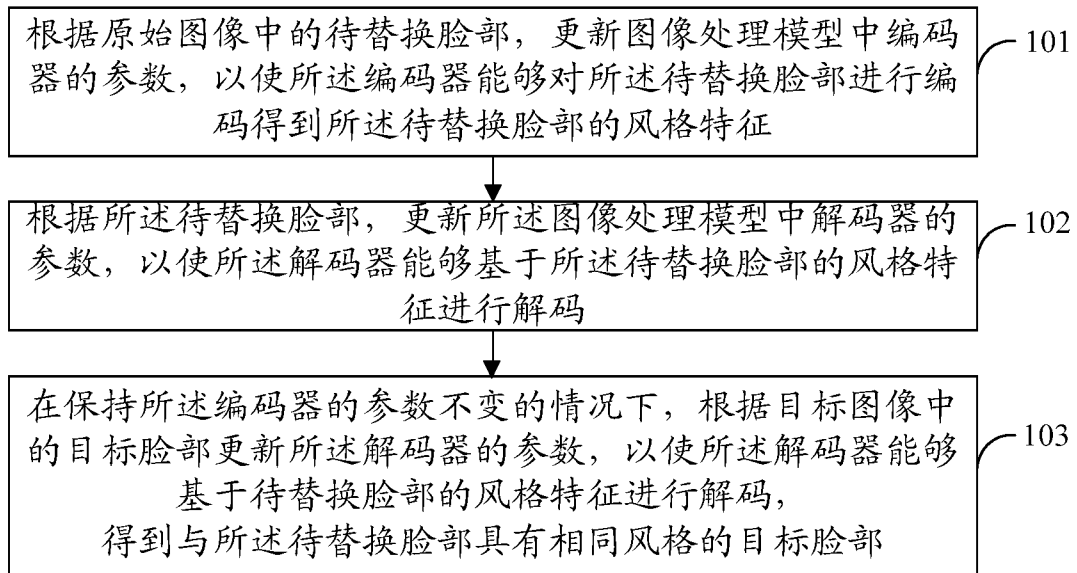


图 9

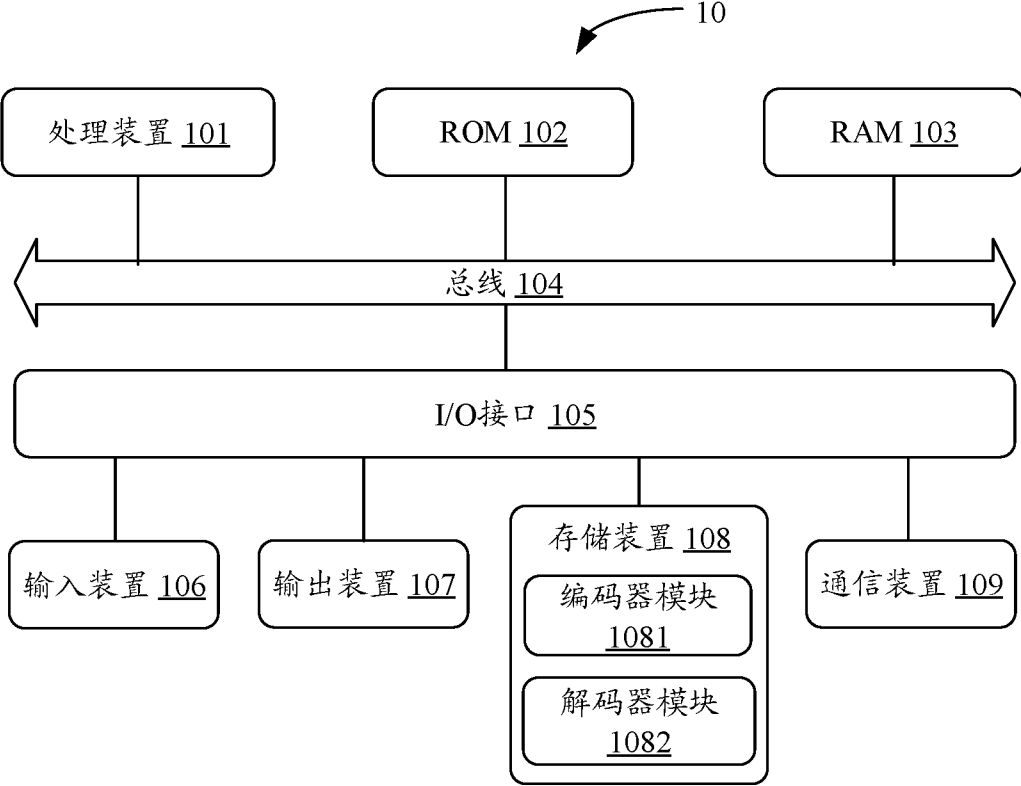


图 10

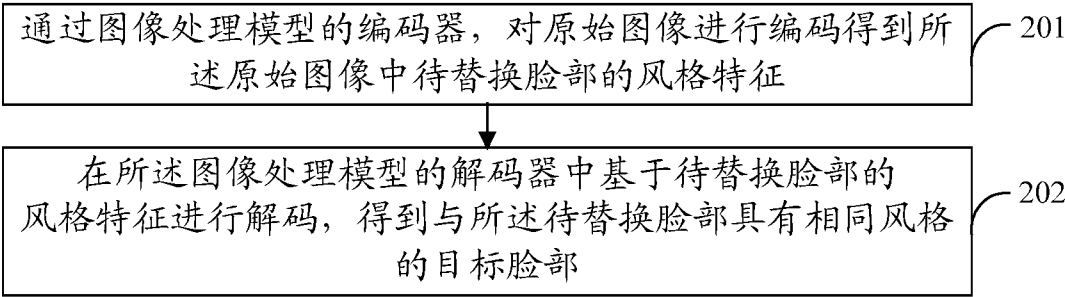


图 11

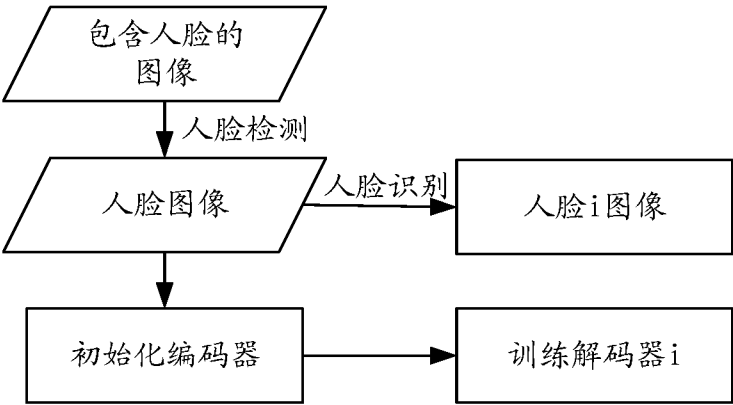


图 12

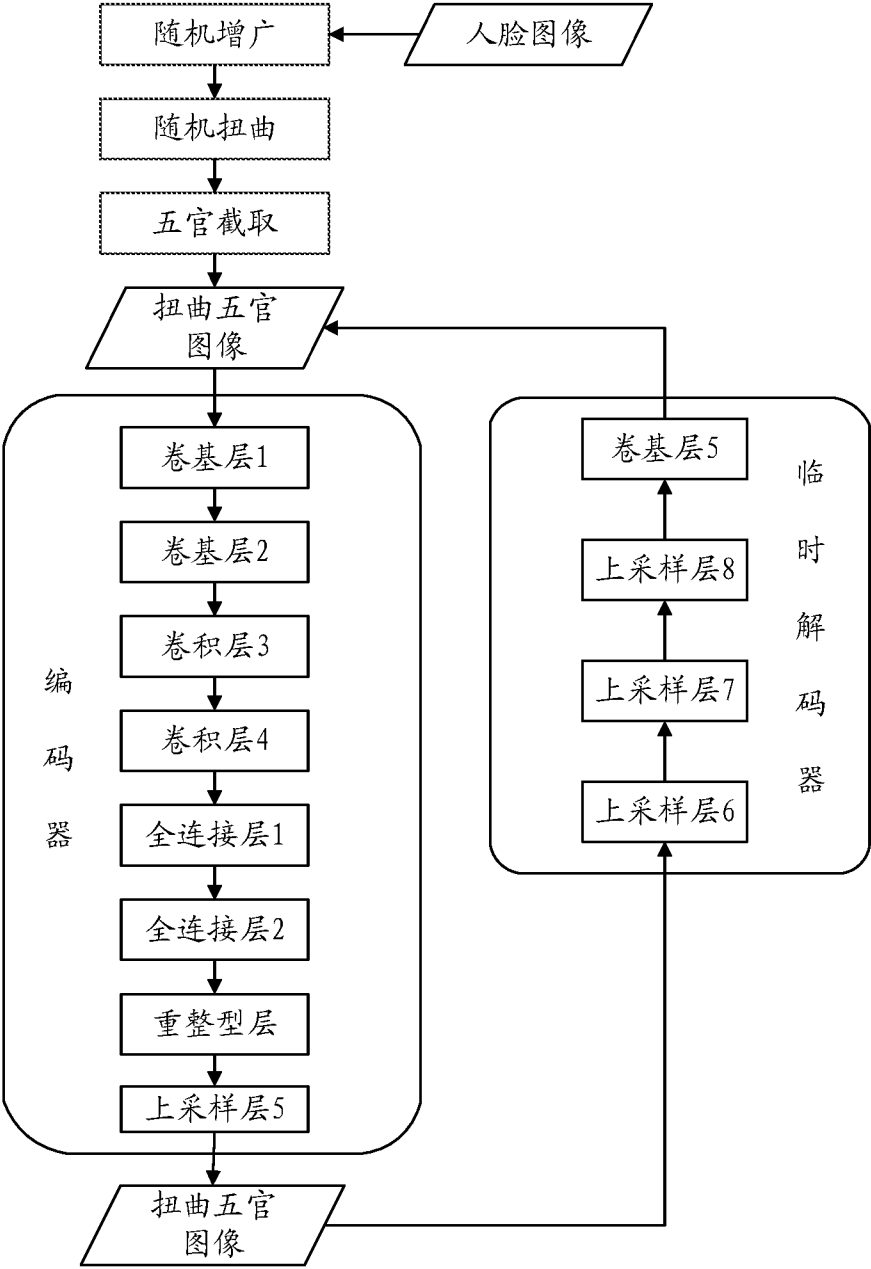


图 13A

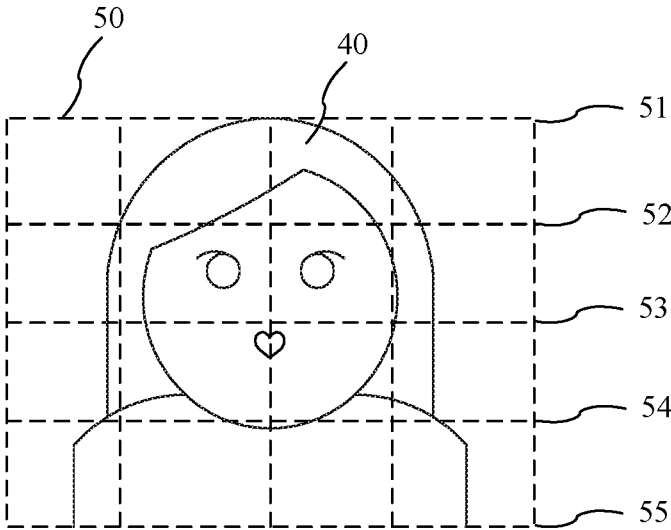


图 13B

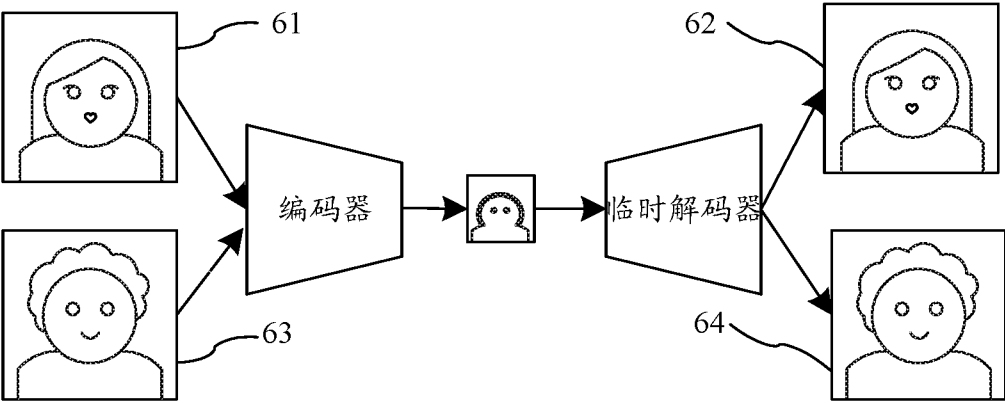


图 14

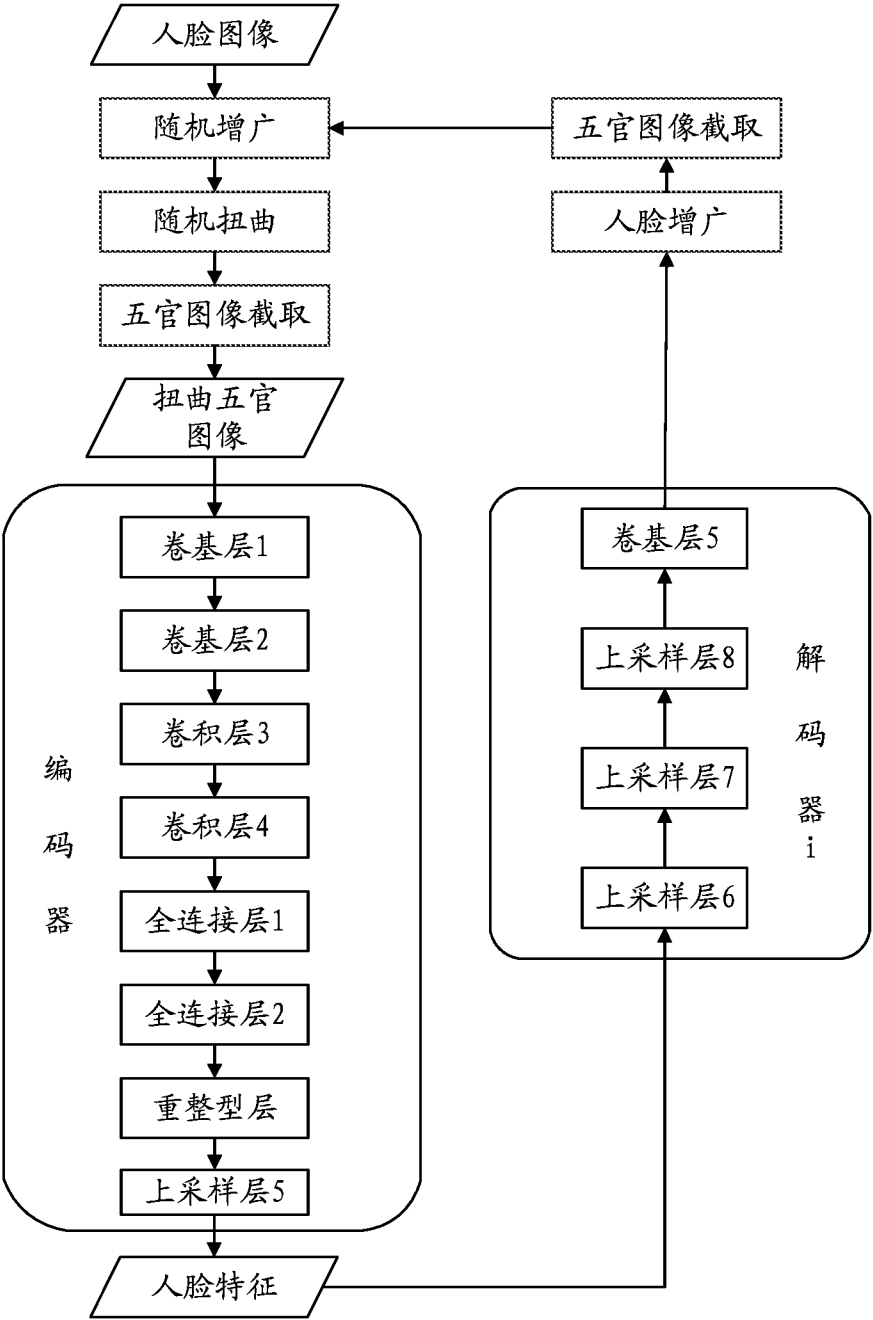


图 15

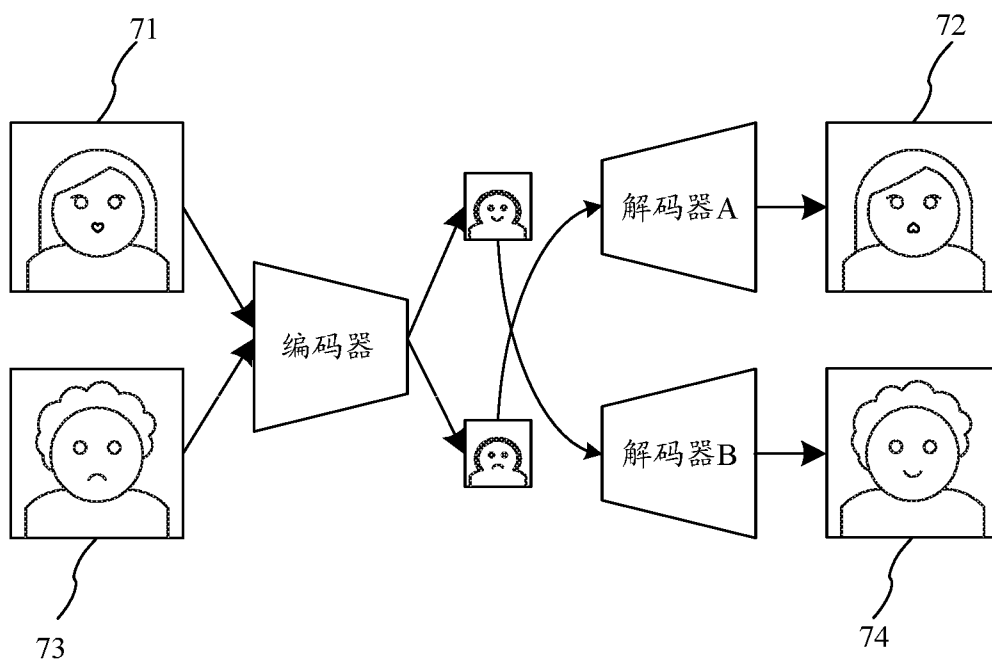


图 16

— 11/11 —

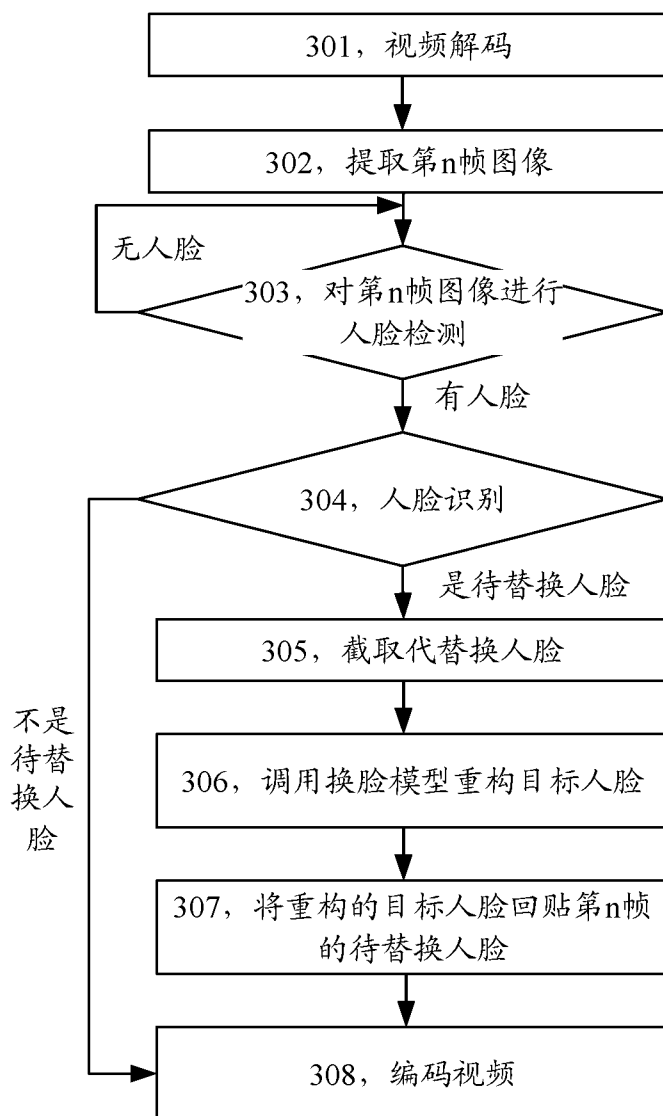


图 17

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2020/081208

A. CLASSIFICATION OF SUBJECT MATTER

G06K 9/00(2006.01)i; G06T 3/00(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; G06T

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, SIPOABS, CNTXT, CNKI, DWPI, USTXT, WOTXT, EPTXT, 百度, BAIDU: 人脸, 脸, 替换, 交换, 换脸, 自编码器, 编码器, 解码器, 训练, 参数不变; face, facial, replace, substitute, exchange, swap, deepfake, autocoder, coder, decoder, training, parameter, unchanged.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 110148081 A (TENCENT TECHNOLOGY SHENZHEN CO., LTD.) 20 August 2019 (2019-08-20) description, paragraphs [0063]-[0190]	1-16
PX	CN 110415166 A (TENCENT TECHNOLOGY SHENZHEN CO., LTD.) 05 November 2019 (2019-11-05) description, paragraphs [0132]-[0238]	1-16
PX	CN 110826593 A (TENCENT TECHNOLOGY SHENZHEN CO., LTD.) 21 February 2020 (2020-02-21) description, paragraphs [0171]-[0305]	1-16
Y	"AI换脸技术—DeepFakes概述 (一) (Non-official translation: AI Face Changing Technology—Overview of DeepFakes (1))" https://cloud.tencent.com/developer/article/1107307 , 24 April 2018 (2018-04-24), entire document	1-16
Y	CN 109447259 A (BEIJING ZIJIE TIAODONG NETWORK TECHNOLOGY CO., LTD.) 08 March 2019 (2019-03-08) description, paragraphs [0048]-[0086], and figures 2-4	1-16



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

05 June 2020

Date of mailing of the international search report

15 June 2020

Name and mailing address of the ISA/CN

China National Intellectual Property Administration
No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing
100088
China

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2020/081208**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	CN 108710831 A (SOUTH CHINA UNIVERSITY OF TECHNOLOGY) 26 October 2018 (2018-10-26) claim 4	5-9, 13, 15, 16
A	CN 108647560 A (SUN YAT-SEN UNIVERSITY) 12 October 2018 (2018-10-12) entire document	1-16
A	US 2015324633 A1 (EMOTIENT, INC.) 12 November 2015 (2015-11-12) entire document	1-16

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2020/081208

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN	110148081	A	20 August 2019	None	
CN	110415166	A	05 November 2019	None	
CN	110826593	A	21 February 2020	None	
CN	109447259	A	08 March 2019	None	
CN	108710831	A	26 October 2018	None	
CN	108647560	A	12 October 2018	None	
US	2015324633	A1	12 November 2015	US 10319130 B2	11 June 2019
				US 9639743 B2	02 May 2017
				US 2017301121 A1	19 October 2017

国际检索报告

国际申请号

PCT/CN2020/081208

A. 主题的分类

G06K 9/00(2006.01)i; G06T 3/00(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; G06T

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, SIPOABS, CNTXT, CNKI, DWPI, USTXT, WOTXT, EPTXT, 百度: 人脸, 脸, 替换, 交换, 换脸, 自编码器, 编码器, 解码器, 训练, 参数不变; face, facial, replace, substitute, exchange, swap, deepfake, autocoder, coder, decoder, training, parameter, unchanged.

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 110148081 A (腾讯科技深圳有限公司) 2019年 8月 20日 (2019 - 08 - 20) 说明书[0063]-[0190]段	1-16
PX	CN 110415166 A (腾讯科技深圳有限公司) 2019年 11月 5日 (2019 - 11 - 05) 说明书[0132]-[0238]段	1-16
PX	CN 110826593 A (腾讯科技深圳有限公司) 2020年 2月 21日 (2020 - 02 - 21) 说明书[0171]-[0305]段	1-16
Y	"AI 换脸技术——DeepFakes 概述 (一)" https://cloud.tencent.com/developer/article/1107307 , 2018年 4月 24日 (2018 - 04 - 24), 全文	1-16
Y	CN 109447259 A (北京字节跳动网络技术有限公司) 2019年 3月 8日 (2019 - 03 - 08) 说明书[0048]-[0086]段, 图2-4	1-16
Y	CN 108710831 A (华南理工大学) 2018年 10月 26日 (2018 - 10 - 26) 权利要求4	5-9, 13, 15, 16
A	CN 108647560 A (中山大学) 2018年 10月 12日 (2018 - 10 - 12) 全文	1-16

☒ 其余文件在C栏的续页中列出。☒ 见同族专利附件。

* 引用文件的具体类型:

“A” 认为不特别相关的表示了现有技术一般状态的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的)

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期

2020年 6月 5日

国际检索报告邮寄日期

2020年 6月 15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

张会

电话号码 62412328

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
A	US 2015324633 A1 (EMOTIENT, INC.) 2015年 11月 12日 (2015 - 11 - 12) 全文	1-16

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2020/081208

检索报告引用的专利文件			公布日 (年/月/日)	同族专利	公布日 (年/月/日)
CN	110148081	A	2019年 8月 20日	无	
CN	110415166	A	2019年 11月 5日	无	
CN	110826593	A	2020年 2月 21日	无	
CN	109447259	A	2019年 3月 8日	无	
CN	108710831	A	2018年 10月 26日	无	
CN	108647560	A	2018年 10月 12日	无	
US	2015324633	A1	2015年 11月 12日	US 10319130 B2	2019年 6月 11日
				US 9639743 B2	2017年 5月 2日
				US 2017301121 A1	2017年 10月 19日