

(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
2021 年 4 月 1 日 (01.04.2021)



(10) 国际公布号
WO 2021/057426 A1

(51) 国际专利分类号:
G06K 9/62 (2006.01)

(21) 国际申请号: PCT/CN2020/113305

(22) 国际申请日: 2020 年 9 月 3 日 (03.09.2020)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201910935739.4 2019 年 9 月 29 日 (29.09.2019) 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: 北京派特恩知识产权代理有限公司 (CHINA PAT INTELLECTUAL PROPERTY OFFICE); 中国北京市海淀区海淀南路 21 号中关村知识产权大厦 B 座 2 层, Beijing 100080 (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, IT, 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: METHOD AND APPARATUS FOR TRAINING IMAGE FUSION PROCESSING MODEL, DEVICE, AND STORAGE MEDIUM

(54) 发明名称: 融合图像处理模型的训练方法、装置、设备及存储介质

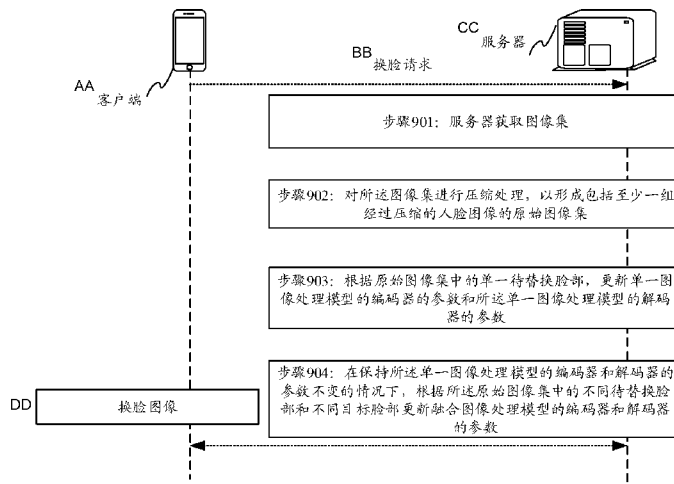


图 9

Step 901: A server acquires an image set
Step 902: Compress the image set, so as to form an original image set including at least one set of compressed facial images
Step 903: Update, according to a single face to be replaced in an original image set, parameters of an encoder of a single image processing model and parameters of a decoder of the single image processing model
Step 904: In cases where the parameters of the encoder and decoder of the single image processing model remain unchanged, update, according to different faces to be replaced and different target faces in the original image set, parameters of an encoder and a decoder of an image fusion processing model
AA Client
BB Face replacement request
CC Server
DD Image of replaced face

(57) Abstract: A method for training an image fusion processing model, comprising: acquiring an image set, and compressing the image set; updating, according to a single face to be replaced in an original image set, parameters of an encoder of a single image processing model and parameters of a decoder of the single image processing model; and in cases where the parameters of the encoder and decoder of the single image processing model are kept unchanged, updating, according to different faces to be replaced and different target faces in the original image set, the parameters of an encoder and a decoder of an image fusion processing model.

(57) 摘要: 一种融合图像处理模型的训练方法, 包括: 获取图像集, 并对所述图像集进行压缩处理, 根据原始图像集中的单一待替换脸部, 更新单一图像处理模型的编码器的参数和所述单一图像处理模型的解码器的参数, 在保持所述单一图像处理模型的编码器和解码器的参数不变的情况下, 根据所述原始图像集中的不同待替换脸部和不同目标脸部更新融合图像处理模型的编码器和解码器的参数。

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))。

融合图像处理模型的训练方法、装置、设备及存储介质

本申请要求申请号为 201910935739.4、申请日为 2019 年 09 月 29 日、申请名称为“融合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。

5 技术领域

本申请涉及图像处理技术，尤其涉及融合图像处理模型的训练方法、图像处理方法、装置、电子设备及存储介质。

背景技术

10 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。

图像处理模型是基于特定的待替换脸部和目标脸部进行训练的，当待替换脸部发生更新时，需要对图像处理模型中的参数进行全量的更新，训练图像处理模型的效率受到影响，15 进而影响了图像处理模型在生产环境中上线以实现上述应用的效率，同时，图像处理模型中的参数进行全量的更新的过程占用了大量的硬件资源，造成了硬件设备成本的提升，不利于图像处理模型的大范围部署。

发明内容

20 有鉴于此，本申请实施例提供融合图像处理模型的训练方法、图像处理方法、装置、电子设备及存储介质，能够保证图像处理模型的识别精度的情况下提升融合图像处理模型的训练效率，实现对图像的批量处理，训练完成的融合图像处理模型的复杂度低，节省存储资源，实现了图像处理模型在生产环境中上线以提升应用的效率，同时，图像处理模型中的参数进行全量的更新的过程降低了硬件资源的占用，减少了硬件设备成本的提升，实现了图像处理模型的大范围部署。

25 本申请实施例提供了一种融合图像处理模型的训练方法，包括：获取图像集，并对所述图像集进行压缩处理，以形成包括至少一组经过压缩的人脸图像的原始图像集；根据原始图像集中的单一待替换脸部，更新单一图像处理模型的编码器的参数和所述单一图像处理模型的解码器的参数，以使所述单一图像处理模型的解码器能够基于所述单一待替换脸部的风格特征进行解码，且解码得到与所述单一待替换脸部具有相同风格的单一目标脸部；30 在保持所述单一图像处理模型的编码器和解码器的参数不变的情况下，根据所述原始图像集中的不同待替换脸部和不同目标脸部更新融合图像处理模型的编码器和解码器的参数，以使所述融合图像处理模型中的解码器能够基于不同待替换脸部的风格特征进行解码，且解码得到与相应待替换脸部具有相同风格的不同目标脸部。

35 本申请实施例还提供了一种融合图像处理模型的图像处理方法，所述方法包括：通过融合图像处理模型的状态编码器，对原始图像集进行编码得到所述原始图像集中不同待替换脸部的风格特征；通过融合图像处理模型的五官编码器，对不同目标脸部进行编码得到所述不同目标脸部所对应的不同目标脸部的五官特征；将所述风格特征和所述五官特征输

入所述融合图像处理模型的融合解码器；在所述融合解码器中基于不同待替换脸部的风格特征和不同目标脸部的五官特征进行解码，得到与所述不同待替换脸部具有相同风格的不同目标脸部；其中，所述融合图像处理模型的编码器和解码器的参数是在保持所述单一图像处理模型的编码器和解码器的参数不变的情况下，根据所述单一目标图像进行更新所得到的。

本申请实施例还提供了一种融合图像处理模型的训练装置，所述训练装置包括：图像处理模块，配置为获取图像集，并对所述图像集进行压缩处理，以形成包括至少一组经过压缩的人脸图像的原始图像集；单一图像处理模型训练模块，配置为根据原始图像集中的单一待替换脸部，更新单一图像处理模型的编码器的参数和所述单一图像处理模型的解码器的参数，以使，所述单一图像处理模型的解码器能够基于所述单一待替换脸部的风格特征进行解码，且解码得到与所述单一待替换脸部具有相同风格的单一目标脸部；融合图像处理模型训练模块，配置为在保持所述单一图像处理模型的编码器和解码器的参数不变的情况下，根据所述原始图像集中的不同待替换脸部和不同目标脸部更新融合图像处理模型的编码器和解码器的参数，以使所述融合图像处理模型中的解码器能够基于不同待替换脸部的风格特征进行解码，且解码得到与相应待替换脸部具有相同风格的不同目标脸部。

本申请实施例还提供了一种图像处理装置，所述图像处理装置包括：编码器模块，配置为通过融合图像处理模型的状态编码器，对原始图像集进行编码得到所述原始图像集中不同待替换脸部的风格特征；所述编码器模块，配置为通过融合图像处理模型的五官编码器，对不同目标脸部进行编码得到所述不同目标脸部所对应的不同目标脸部的五官特征；解码器模块，配置为将所述风格特征和所述五官特征输入所述融合图像处理模型的融合解码器；所述解码器模块，配置为在所述融合解码器中基于不同待替换脸部的风格特征和不同目标脸部的五官特征进行解码，得到与所述不同待替换脸部具有相同风格的不同目标脸部；其中，所述融合图像处理模型的编码器和解码器的参数是在保持所述单一图像处理模型的编码器和解码器的参数不变的情况下，根据所述单一目标图像进行更新所得到的。

本申请实施例还提供了一种电子设备，所述电子设备包括：存储器，配置为存储可执行指令；处理器，配置为运行所述存储器存储的可执行指令时，实现前序的融合图像处理模型的训练方法。

本申请实施例还提供了一种电子设备，所述电子设备包括：存储器，配置为存储可执行指令；处理器，配置为运行所述存储器存储的可执行指令时，实现前序的图像处理方法。

本申请实施例还提供了一种计算机可读存储介质，存储有可执行指令，所述可执行指令被处理器执行时实现前序的融合图像处理模型的训练方法，或者实现前序的图像处理方法。

应用本申请实施例提供的融合图像处理模型的训练方法、图像处理方法、装置、电子设备及存储介质，至少具有以下有益技术效果：

通过获取图像集，并对所述图像集进行压缩处理，以形成包括至少一组经过压缩的人脸图像的原始图像集；根据原始图像集中的单一待替换脸部，更新单一图像处理模型的编码器的参数和所述单一图像处理模型的解码器的参数，以使所述单一图像处理模型的解码器能够基于所述单一待替换脸部的风格特征进行解码，且解码得到与所述单一待替换脸部具有相同风格的单一目标脸部；在保持所述单一图像处理模型的编码器和解码器的参数不变的情况下，根据所述原始图像集中的不同待替换脸部和不同目标脸部更新融合图像处理模型的编码器和解码器的参数，以使所述融合图像处理模型中的解码器能够基于不同待替换脸部的风格特征进行解码，且解码得到与相应待替换脸部具有相同风格的不同目标脸部，由此，实现了对融合图像处理模型的解码器的训练与原始图像集的解耦，减少融合图像处理模型的复杂度低，节省硬件存储资源的同时，经过训练的融合图像处理模型能够批量处理不同的原始图像，增强了融合图像处理模型的通用性；同时，图像处理模型中的参

数进行全量的更新的过程降低了硬件资源的占用，减少了硬件设备成本的提升，实现了图像处理模型的大范围部署，有利于图像处理模型部署在移动终端中，减轻用户的使用成本。

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；
- 5 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；
- 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；
- 10 图；
- 图 7 为本申请实施例提供的融合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融合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融合图像处理模型的训练方法一个可选的流程示意图；
- 15 图 10 为本申请实施例提供的用于交换人脸的单一图像处理模型的训练原理示意图；
- 图 11A 为本申请实施例提供的单一图像处理模型的训练方法的原理示意图；
- 图 11B 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；
- 图 12 为本申请实施例提供的初始化编码器的原理示意图；
- 图 13 为本申请实施例对单一图像处理模型的解码器的训练原理示意图；
- 20 图 14 为本申请实施例单一图像处理模型进行换脸的原理示意图；
- 图 15 为本申请实施例所示的融合图像处理模型结构示意图；
- 图 16 为本申请实施例提供的图像处理装置的结构示意图；
- 图 17 为本申请实施例提供的融合图像处理模型的训练方法一个可选的流程示意图；
- 图 18A 为用户通过终端所输入的原始图像集中的不同待替换脸部和不同目标脸部示意图；
- 25 图；
- 图 18B 为用户通过终端所输入的原始图像集中的不同待替换脸部和不同目标脸部示意图；
- 图 19 为对单一图像处理模型的训练和对融合图像处理模型训练的过程示意图；
- 图 20 为本申请实施例提供的编码器的示例性结构图；
- 30 图 21 为本申请实施例提供的解码器的示例性结构图；
- 图 22 为本申请实施例提供的融合图像处理方法一个可选的处理流程；
- 图 23 为本申请中数据集获取示意图；
- 图 24 为本申请中数据集压缩处理过程示意图；
- 图 25 为本申请实施例提供的融合图像处理方法一个可选的训练流程示意图；
- 35 图 26 为本申请实施例中单一换脸模型的网络结构示意图；
- 图 27 为本申请实施例中图像增广示意图；
- 图 28 为本申请实施例中随机扭曲的方法的处理过程示意图；
- 图 29A 为本申请实施例提供的融合图像处理方法一个可选的训练流程示意图；
- 图 29B 为本申请实施例提供的融合图像处理方法数据结构示意图；
- 40 图 30A 为本申请实施例提供的融合图像处理方法一个可选的使用流程示意图；
- 图 30B 为本申请实施例提供的融合图像处理方法数据结构示意图；
- 图 31 为本申请实施例提供的融合图像处理模型的图像处理 100 的架构示意图；
- 图 32 为本申请实施例提供的区块链网络 200 中区块链的结构示意图；

图 33 是本申请实施例提供的区块链网络 200 的功能架构示意图。

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。

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。

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

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。

3) 编码器，配置为从原始图像集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）从原始图像集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。

5) 解码器，与目标脸部一一对应的模型，配置为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，即将原始图像集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。

7) 单一图像处理模型，每次仅只能处理一张图像，如果需要处理不同的图像，则需要训练不同的单一图像处理模型。

8) 融合图像处理模型，每次能够处理至少两张图像，可以根据用户的指令或预设的程序对原始图像集中的图像进行批量处理。

9) 交易（Transaction），等同于计算机术语“事务”，交易包括了需要提交到区块链网络执行的操作，并非单指商业语境中的交易，鉴于在区块链技术中约定俗成地使用了“交易”这一术语，本申请实施例遵循了这一习惯。

10) 区块链（Blockchain），是由区块（Block）形成的加密的、链式的交易的存储结构。

11) 区块链网络（Blockchain Network），通过共识的方式将新区块纳入区块链的一系列节点的集合。

12) 账本（Ledger），是区块链（也称为账本数据）和与区块链同步的状态数据库的统称。

13) 智能合约（Smart Contracts），也称为链码（Chaincode）或应用代码，部署在区块链网络的节点中的程序，节点执行接收的交易中所调用的智能合约，来对状态数据库的键值对数据进行更新或查询的操作。

14) 共识（Consensus），是区块链网络中的一个过程，用于在涉及的多个节点之间对区块中的交易达成一致，达成一致的区块将被追加到区块链的尾部，实现共识的机制包括工作量证明（PoW, Proof of Work）、权益证明（PoS, Proof of Stake）、股份授权证明（DPoS, Delegated Proof-of-Stake）、消逝时间量证明（PoET, Proof of Elapsed Time）等。

图 1 为本申请实施例提供的单一图像处理模型的结构示意图，单一图像处理模型包括

编码器和解码器。解码器是与用于替换“单一待替换脸部”（其中，单一待替换脸部可以理解为：原始图像集中的待替换脸部 A，目标脸部 B，将原始图像集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，以实现待替换脸部 A 具有目标脸部 B 的风格特征）的单一目标脸部是一一对应的，因此单一图像处理模型中解码器的数量取决于单一图像处理模型需要处理的不同单一目标脸部（例如不同人脸）的数量。例如当单一图像处理模型需要将视频中的单一待替换脸部替换为 2 个不同的目标人脸时，单一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 个卷积层、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 2 和表 1 所示。

表 1

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

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。

表 2

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
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	-

图 4 为本申请实施例提供的单一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，编码器从原始图像集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单一图像处理模型进行换脸的原理示意图， x 为输入的单一待替换脸部，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单一待替换脸部。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，解码得到重构的单一待替换脸部 r 可以表示为 $g(x/h) * h$ 。

当单一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 的单一目标脸部是对象 A 的脸部，解码器 B 的单一目标脸部是对象 B 的脸部，对于编码器针对同一个原始图像集（对象 C）中的单一待替换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单一目标脸部 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单一目标脸部 B。

上述的对象可以是任意具有五官的生物（包括人和动物），以人脸为例，继续对本申请实施例提供的单一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单一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 A 进行编码得到人脸 A 的降维后的特征，解码器 A 配置为基于人脸 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 B 进行编码得到人脸 B 的降维后的特征，解码器 B 配置为基于人脸 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 示出的单一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单一图像处理模型训练存在以下问题：

1、单一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）的单一图像处理模型之后，如果还要使单一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，部署阶段占用存储资源较多

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，进而影响了单一图像处理模型换脸的效果，同时，由于图像处理模型仅能够对单一图像进行处理，用户需要同时处理多张不同的图像时，需要部署多个单一图像处理模型，其部署时存储复杂度较高，占用了大量的存储资源。

3、传统的换脸模型仅支持对单人换脸的处理，如果需要换到新的人脸就需要重新训练新的模型，耗时且麻烦，而且每个模型都需要占用自己的存储空间，过程繁琐，增加整个换脸处理流程的耗时，影响用户的使用体验。同时，训练一个单一换脸模型需要 100 张以上的图片（人脸图像）作为训练样本，而换脸模型的使用者（用户）往往出于隐私的原因，并不会提供自身人脸图像作为训练样本，影响了训练样本图像的离散度。

针对上述问题，本申请实施例提供了一种融合图像处理模型的训练方法和训练装置，以及融合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融合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，实现本申请实施例融合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 40 连接，其中网络 40 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 提交针对不同融合目标脸部的训练任务，由训练装置 30 训练相应的融合图像处理模型，并相应的服务器下发融合图像处理模型的参数到训练装置 30。

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 完成融合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融合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。

继续目标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融合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。

下面对本申请实施例的融合图像处理模型的训练装置的结构做详细说明，融合图像处理模型的训练装置可以各种形式来实施，如带有融合图像处理模型训练装置的终端，也可以为设置有融合图像处理模型训练功能的服务器，其中，带有融合图像处理模型训练装置的终端，可以为前序所示的实施例中带有融合图像处理模型训练装置的电子设备。例如前序图 7 中的服务器；图 8 为本申请实施例提供的融合图像处理模型的训练装置的组成结构示意图，可以理解，图 8 仅仅示出了融合图像处理模型的训练装置的示例性结构而非全部结构，根据需要可以实施图 8 示出的部分结构或全部结构。

本申请实施例提供的融合图像处理模型的训练装置包括：至少一个处理器 801、存储器 802、用户接口 803 和至少一个网络接口 804。融合图像处理模型训练装置中的各个组件通过总线系统 805 耦合在一起。可以理解，总线系统 805 配置为实现这些组件之间的连接通信。总线系统 805 除包括数据总线之外，还包括电源总线、控制总线和状态信号总线。但是为了清楚说明起见，在图 8 中将各种总线都标为总线系统 805。

其中，用户接口 803 可以包括显示器、键盘、鼠标、轨迹球、点击轮、按键、按钮、触感板或者触摸屏等。

可以理解，存储器 802 可以是易失性存储器或非易失性存储器，也可包括易失性和非易失性存储器两者。本申请实施例中的存储器 802 能够存储数据以支持终端（如 10-1）的操作。这些数据的示例包括：用于在终端（如 10-1）上操作的任何计算机程序，如操作系统和应用程序。其中，操作系统包含各种系统程序，例如框架层、核心库层、驱动层等，配置为实现各种基础业务以及处理基于硬件的任务。应用程序可以包含各种应用程序。

在一些实施例中，本申请实施例提供的融合图像处理模型的训练装置可以采用软硬件结合的方式实现，作为示例，本申请实施例提供的融合图像处理模型训练装置可以是采用

硬件译码处理器形式的处理器，其被编程以执行本申请实施例提供的融合图像处理模型的训练方法。例如，硬件译码处理器形式的处理器可以采用一个或多个应用专用集成电路（ASIC, Application Specific Integrated Circuit）、DSP、可编程逻辑器件（PLD, Programmable Logic Device）、复杂可编程逻辑器件（CPLD, Complex Programmable Logic Device）、现场可编程门阵列（FPGA, Field-Programmable Gate Array）或其他电子元件。

作为本申请实施例提供的融合图像处理模型的训练装置采用软硬件结合实施的示例，本申请实施例所提供的融合图像处理模型的训练装置可以直接体现为由处理器 801 执行的软件模块组合，软件模块可以位于存储介质中，存储介质位于存储器 802，处理器 801 读取存储器 802 中软件模块包括的可执行指令，结合必要的硬件（例如，包括处理器 801 以及连接到总线 805 的其他组件）完成本申请实施例提供的融合图像处理模型的训练方法。

作为示例，处理器 801 可以是一种集成电路芯片，具有信号的处理能力，例如通用处理器、数字信号处理器（DSP, Digital Signal Processor），或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件等，其中，通用处理器可以是微处理器或者任何常规的处理器等。

作为本申请实施例提供的融合图像处理模型的训练装置采用硬件实施的示例，本申请实施例所提供的装置可以直接采用硬件译码处理器形式的处理器 801 来执行完成，例如，被一个或多个应用专用集成电路（ASIC, Application Specific Integrated Circuit）、DSP、可编程逻辑器件（PLD, Programmable Logic Device）、复杂可编程逻辑器件（CPLD, Complex Programmable Logic Device）、现场可编程门阵列（FPGA, Field-Programmable Gate Array）或其他电子元件执行实现本申请实施例提供的融合图像处理模型的训练方法。

本申请实施例中的存储器 802 配置为存储各种类型的数据以支持融合图像处理模型训练装置的操作。这些数据的示例包括：用于在融合图像处理模型训练装置上操作的任何可执行指令，如可执行指令，实现本申请实施例的从融合图像处理模型的训练方法的程序可以包含在可执行指令中。

在另一些实施例中，本申请实施例提供的融合图像处理模型的训练装置可以采用软件方式实现，图 8 示出了存储在存储器 802 中的融合图像处理模型的训练装置，其可以是程序和插件等形式的软件，并包括一系列的模块，作为存储器 802 中存储的程序的示例，可以包括融合图像处理模型的训练装置，融合图像处理模型的训练装置中包括以下的软件模块：图像处理模块 8081、单一图像处理模型训练模块 8082、融合图像处理模型训练模块 8083。当融合图像处理模型的训练装置中的软件模块被处理器 801 读取到 RAM 中并执行时，将实现本申请实施例提供的融合图像处理模型的训练方法，其中，融合图像处理模型的训练装置中各个软件模块的功能，包括：

图像处理模块 8081，配置为获取图像集，并对所述图像集进行压缩处理，以形成包括至少一组经过压缩的人脸图像的原始图像集。

单一图像处理模型训练模块 8082，配置为根据原始图像集中的单一待替换脸部，更新单一图像处理模型的编码器的参数和所述单一图像处理模型的解码器的参数，以使，

所述单一图像处理模型的解码器能够基于所述单一待替换脸部的风格特征进行解码，且解码得到与所述单一待替换脸部具有相同风格的单一目标脸部；

融合图像处理模型训练模块 8083，配置为在保持所述单一图像处理模型的编码器和解码器的参数不变的情况下，根据所述原始图像集中的不同待替换脸部和不同目标脸部更新融合图像处理模型的编码器和解码器的参数，以使，

所述融合图像处理模型中的解码器能够基于不同待替换脸部的风格特征进行解码，且解码得到与相应待替换脸部具有相同风格的不同目标脸部。

结合图 8 示出的电子设备 80 说明本申请实施例提供的融合图像处理模型的训练方法，参见图 9，图 9 为本申请实施例提供的融合图像处理模型的训练方法一个可选的流程示意

图，可以理解地，图 9 所示的步骤可以由运行融合图像处理模型的训练装置的各种电子设备执行，例如可以是如带有融合图像处理模型的训练功能的终端、服务器或者服务器集群。下面针对图 9 示出的步骤进行说明，可以包括：

步骤 901：服务器获取图像集。

5 其中，所获取的原始图像集中的图像可以作为单一图像处理模型和融合图像处理模型的训练样本。

步骤 902：对所述图像集进行压缩处理，以形成包括至少一组经过压缩的人脸图像的原始图像集。

10 在本申请的一些实施例中，当根据原始图像集中的单一待替换脸部，更新所述单一图像处理模型中编码器的参数之前，所述方法还包括：

从样本图像集合的各个样本图像中截取包括所述单一待替换脸部的所述原始图像集；将所截取的各个所述原始图像集基于单一待替换脸部的特征点进行对齐。

在本申请的一些实施例中，获取图像集，并对所述图像集进行压缩处理，以形成包括至少一组经过压缩的人脸图像的原始图像集，可以通过以下方式实现：

15 爬取网络中的图像资源，并对所爬取的图像资源中的原始图像进行人脸特征定位与数据清洗；对所述原始图像集进行随机增广处理；将经过随机增广处理的原始图像集进行随机扭曲处理，得到扭曲的所述原始图像集；截取扭曲的所述原始图像集中单一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，以适配所述单一图像处理模型的编码器支持输入的分辨率；对经过缩放处理的所述五官图像进行压缩，触发形成与所述五官图像的格式相对应的未解压状态的人脸图像的二进制数据，以作为包括至少一组经过压缩的人脸图像的原始图像集。其中，形成与所述五官图像的格式相对应的未解压状态的人脸图像的二进制数据，使得单一换脸模型和融合换脸模型在训练过程中无需加载全部图像数据，仅在训练的初始化阶段加载与自身训练过程相匹配的图像的二进制数据即可完成相应训练，有效减少了由于训练图像众多所造成的硬件资源成本的上升。

25 在本申请的一些实施例中，对所述原始图像集进行随机增广处理，可以通过以下方式实现：

基于所述原始图像集的中心进行随机旋转；将随机旋转后的所述原始图像集进行随机缩放；将随机缩放后所述原始图像集进行随机平移。

30 在本申请的一些实施例中，将经过随机增广处理的原始图像集进行随机扭曲处理，可以通过以下方式实现：

在经过随机增广处理的原始图像集中插入噪声，所述噪声包括固定噪声和动态噪声中的至少一种。其中，通过在经过随机增广处理的原始图像集中插入噪声可以有效提升训练样本的离散度，避免由于训练样本的离散度过低影响单一换脸模型和融合换脸模型的参数失真。

35 在本申请的一些实施例中，在经过随机增广处理的原始图像集中插入噪声，可以通过以下方式实现：

将所述经过随机增广处理的原始图像集进行网格化，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。

40 步骤 903：根据原始图像集中的单一待替换脸部，更新单一图像处理模型的编码器的参数和所述单一图像处理模型的解码器的参数。

由此，可以使所述单一图像处理模型的解码器能够基于所述单一待替换脸部的风格特征进行解码，且解码得到与所述单一待替换脸部具有相同风格的单一目标脸部。

在本申请的一些实施例中，根据原始图像集中的单一待替换脸部，更新单一图像处理模型的编码器的参数和所述单一图像处理模型的解码器的参数，可以通过以下方式实现：

45 根据原始图像集中的单一待替换脸部，更新单一图像处理模型中编码器的参数，以使

单一图像处理模型的编码器能够对单一待替换脸部进行编码，且编码得到单一待替换脸部的风格特征；根据原始图像集的单一待替换脸部，更新单一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单一待替换脸部的风格特征进行解码；在保持所述单一图像处理模型的编码器的参数不变的情况下，根据单一目标图像中的单一目标脸部更新所述解

5 码器的参数，以使所述解码器能够基于所述单一待替换脸部的风格特征进行解码，且解码得到与单一待替换脸部具有相同风格的单一目标脸部。

在本申请的一些实施例中，根据原始图像集中的单一待替换脸部，更新单一图像处理模型中编码器的参数，可以通过以下方式实现：

将扭曲原始图像集中单一待替换脸部的五官图像，代入由所述单一图像处理模型的编

10 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单一图像处理模型的编码器的参数，其中，所述扭曲原始图像集是对所述原始图像集进行扭曲处理得到。

在本申请的一些实施例中，根据所述原始图像集的单一待替换脸部，更新所述单一图像处理模型中解码器的参数，可以通过以下方式实现：

15 将扭曲原始图像集中单一待替换脸部的五官图像，代入由所述单一图像处理模型的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集是对所述原始图像集进行扭曲处理得到。

在本申请的一些实施例中，在保持所述单一图像处理模型的编码器的参数不变的情况下，根据单一目标图像中单一目标脸部的风格特征更新所述解码器的参数，可以通过以下方式实现：

20 将扭曲单一目标图像中单一目标脸部的五官图像，代入由所述单一图像处理模型的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单一图像处理模型的编码器的参数不变的情况下，执行以下处理：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，所述扭曲单一目标图像是对所述单一目标图像进行扭曲处理得到。

下面对训练并应用交换人脸的单一图像处理模型（也可以称为换脸模型）的过程进行说明，其中，单一图像处理模型的解码器的输出结果可以作为融合图像处理模型训练的监督数据。

25 参见图 10，图 10 为本申请实施例提供的用于交换人脸的单一图像处理模型的训练原理示意图。单一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

30 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。

35 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。

40 参考图 11A，图 11A 为本申请实施例提供的单一图像处理模型的训练方法的原理示意

图, 结合图 11A 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, 拉正的倍数在 $[1-\beta, 1+\beta]$ 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 11B, 图 11B 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) $\times N$ 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 11B 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 11B 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 $x_{ij}(i=0, \dots, 4, j=0, \dots, 4)$ 、 $y_{ij}(i=0, \dots, 4, j=0, \dots, 4)$ 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 σ^2 , 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 (X_2, Y_2) , 其中矩阵 X_2 的元素 $x_{2ij}=i/W$, 矩阵 Y_2 中元素 $y_{2ij}=j/W$, 其中 $i=0, \dots, W, j=0, \dots, W$ 。

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 X_2 中比 x_{ab} 小的元素, 记为 x_{2c} , 查找矩阵 Y_2 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 d_A 、 d_B 、 d_C 和 d_D , 根据这四个距离加权得到 z_{ab} 的值:

$$z_{ab} = d_A / (d_A + d_B + d_C + d_D) \times I_{c,d} + d_B / (d_A + d_B + d_C + d_D) \times I_{c+1,d} + d_C / (d_A + d_B + d_C + d_D) \times I_{c,d+1} + d_D / (d_A + d_B + d_C + d_D) \times I_{c+1,d+1}.$$

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 11A 示出的自编码网络来初始化编码器, 在图 11A 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 $Encoder(x)$, x 为输入的扭曲的五官图像, 临时解码器为 $Decoder_tmp(f)$, f 为 x 经编码器编码后得到的输出 (即 $Encoder(x)$), 则损失函数 $Loss_init$ 可以表示为:

$loss_init = \sum (Decoder_tmp(Encoder(x)) - x)^2$ 。

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 12，图 12 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。

下面目标图 13，图 13 为本申请实施例对单一图像处理模型的解码器的训练原理示意图，结合图 12 说明单一图像处理模型解码器的训练过程。

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 $Encoder(x)$ ， x 为输入的扭曲的五官图像，解码器 i 为 $Decoder_i(f)$ ，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 $loss_i$ ，更新编码器 $Decoder_i$ 的参数： $loss_i = \sum (Decoder_i(Encoder(x)) - x)^2$ 。

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 14，图 14 为本申请实施例单一图像处理模型进行换脸的原理示意图，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 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。

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。

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单一图像处理模型（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。

至此，单一图像处理模型的训练完成，启动对融合图像处理模型的训练过程，其中，单一图像处理模型的解码器的输出结果可以作为融合图像处理模型训练的监督数据。

在本申请的一些实施例中，所述方法还包括：

根据所述编码器的参数和所述解码器的参数，迭代更新与所述单一图像处理模型对应的单一图像处理模型集合中的所有单一图像处理模型的编码器的参数和解码器的参数，直至所述单一图像处理模型集合中的所有单一图像处理模型的编码器的参数和解码器的参数的损失函数满足各自对应的收敛条件，其中，所述单一图像处理模型集合中的单一图像处理模型数量与所述单一图像处理模型的使用环境相适配。其中，由于融合换脸模型可以部署在不同的使用环境中，例如：融合换脸模型可以部署于短视频运营商的服务器中，或

者用户的终端中（包括联网环境和离线环境），其对于融合图像处理模型的训练能力并不相同，根据所述编码器的参数和所述解码器的参数，迭代更新与所述单一图像处理模型对应的单一图像处理模型集合中的所有单一图像处理模型的编码器的参数和解码器的参数，直至所述单一图像处理模型集合中的所有单一图像处理模型的编码器的参数和解码器的参数的损失函数满足各自对应的收敛条件，可以降低硬件资源的占用，减少了硬件设备成本

的提升，实现了融合图像处理模型的大范围部署，提升用户的使用体验。

步骤 904：在保持所述单一图像处理模型的编码器和解码器的参数不变的情况下，根据所述原始图像集中的不同待替换脸部和不同目标脸部更新融合图像处理模型的编码器和解码器的参数。

由此，可以使所述融合图像处理模型中的解码器能够基于不同待替换脸部的风格特征进行解码，且解码得到与相应待替换脸部具有相同风格的不同目标脸部。

在本申请的一些实施例中，当根据所述原始图像集中的不同待替换脸部，更新所述融合图像处理模型中状态编码器的参数之前，所述方法还包括：

根据所述融合图像处理模型的使用环境，抽取所述原始图像集中的第一数量的人脸图像作为第一训练集合；抽取所述原始图像集中的第二数量的人脸图像作为第二训练集合，其中，所述第二训练集合中的人脸图像为待替换脸部的正脸图像；响应于所述第一训练集合和所述第二训练集合，加载与所述第一训练集合和所述第二训练集合分别对应的未解压状态的人脸图像的二进制数据，以及相应数量的单一换脸模型，以完成所述融合换脸模型的训练初始化。其中，由于融合图像处理模型的部署使用环境并不相同，即可以不属于相应软件运营商的服务器中，也可以部署于用户的终端中，不同的使用环境对于融合图像处理模型的处理能力（训练和使用）并不相同，根据所述融合图像处理模型的使用环境，抽取所述原始图像集中的第一数量的人脸图像作为第一训练集合；抽取所述原始图像集中的第二数量的人脸图像作为第二训练集合，可以实现融合图像处理模型的大范围部署，提升用户的使用体验。

在本申请的一些实施例中，在保持所述单一图像处理模型的编码器和解码器的参数不变的情况下，根据所述原始图像集中的不同待替换脸部和不同目标脸部更新融合图像处理模型的编码器和解码器的参数，可以通过以下方式实现：

以所述单一图像处理模型的解码器的解码结果作为所述融合图像处理模型中的训练监督数据；根据所述原始图像集中的不同待替换脸部，更新所述融合图像处理模型中状态编码器的参数，以使所述融合图像处理模型中的状态编码器能够对所述不同待替换脸部分别进行编码，且编码得到分别对应所述不同待替换脸部的状态特征；根据所述不同目标脸部，更新所述融合图像处理模型中五官编码器的参数，以使所述融合图像处理模型中的五官编码器能够对所述不同目标脸部分别进行编码，且编码得到分别对应所述不同目标脸部的五官特征；根据所述融合图像处理模型中的状态编码器的编码结果和五官编码器编码结果，更新所述融合图像处理模型中融合解码器的参数，以使所述融合图像处理模型中的融合解码器能够基于不同待替换脸部的风格特征进行解码，且解码得到与相应待替换脸部具有相同风格的不同目标脸部。

在本申请的一些实施例中，根据所述原始图像集中的不同待替换脸部，更新所述融合图像处理模型中状态编码器的参数，可以通过以下方式实现：

将原始图像集中不同待替换脸部的状态特征，代入由所述融合图像处理模型的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融合图像处理模型中状态编码器的参数。

在本申请的一些实施例中，根据所述不同目标脸部，更新所述融合图像处理模型中五官编码器的参数，可以通过以下方式实现：

将不同目标脸部的五官特征，代入由所述融合图像处理模型的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融合图像处理模型中五官编码器的参数。

在本申请的一些实施例中，根据所述融合图像处理模型中的状态编码器的编码结果和五官编码器编码结果，更新所述融合图像处理模型中融合解码器的参数，可以通过以下方式实现：

将所述原始图像集中不同待替换脸部的状态特征和不同目标脸部的五官特征，代入由所述融合图像处理模型的状态编码器、五官编码器和所述融合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融合解码器的参数。

下面对训练并应用交换人脸的融合图像处理模型（也可以称为融合换脸模型）的具体过程进行说明，其中，图 15 为本申请实施例所示的融合图像处理模型结构示意图，融合图像处理模型的结构如图 15 所示，包括：

状态编码器、五官编码器和融合解码器，其中，状态编码器用于提取待换人脸的状态特征——比如表情、姿态、光照情况等；而五官编码器的编码结果仅只保留目标人脸的五官特征，而去除状态特征。最后融合编码器根据状态特征与五官特征，解码得到与相应待替换脸部具有相同风格的不同目标脸部；新的目标脸部有着待换人脸的状态，和目标脸部的五官特征，实现批量的对不同目标脸部的换脸功能。

通过输入不同的目标脸部，就能解码得到与相应待替换脸部具有相同风格的不同目标脸部。其中，图 15 所示的融合图像处理模型的状态编码器、五官编码器和融合解码器的参数如表 3 所示，

表 3

层类型	通道数	卷积核大小	步长	填充	激活函数	
待换人脸(64*64*3)	3	-	-	-	-	
卷积层 a1	128	(5, 5)	(2, 2)	same	pReLU	状态编码器
卷积层 a2	256	(5, 5)	(2, 2)	same	pReLU	
卷积层 a3	512	(5, 5)	(2, 2)	same	pReLU	
卷积层 a4	1024	(5, 5)	(2, 2)	same	pReLU	
全连接层 a1	1024	(5, 5)	-	-	pReLU	
全连接层 a2	16384	(5, 5)	-	-	pReLU	
Reshape 层 a	-	(4, 4, 1024)	-	-	-	
上采样卷积层 a1	512	(3, 3)	(2, 2)	same	pReLU	
目标人脸(64*64*3)	3	-	-	-	-	
卷积层 a1	128	(5, 5)	(2, 2)	same	pReLU	五官编码器
卷积层 a2	256	(5, 5)	(2, 2)	same	pReLU	
卷积层 a3	512	(5, 5)	(2, 2)	same	pReLU	
卷积层 a4	1024	(5, 5)	(2, 2)	same	pReLU	
全连接层 a1	1024	(5, 5)	-	-	pReLU	
全连接层 a2	16384	(5, 5)	-	-	pReLU	
Reshape 层 a	-	(4, 4, 1024)	-	-	-	
拼接层	在通道的维度上拼接					融合解码器 i
上采样卷积层 a2	512	(3, 3)	(2, 2)	same	pReLU	
上采样卷积层 a3	256	(3, 3)	(2, 2)	same	pReLU	
上采样卷积层 a4	128	(3, 3)	(2, 2)	same	pReLU	
上采样卷积层 5	64	(3, 3)	(2, 2)	same	pReLU	
卷积层 4	3	(3, 3)	(1, 1)	same	pReLU	

输出(128*128*3)	3	-	-	-	-	-
---------------	---	---	---	---	---	---

至此，完成了对融合图像处理模型的训练，训练完成的融合图像处理模型可以部署于相应的服务器中，当然也可以部署于用户终端中（实现不联网状态下融合图像处理模型对不同图像的处理），融合图像处理模型的部署位置本申请不做限制。

5 当通过前序实施例对融合图像处理模型训练完成后，需要将融合图像处理模型部署于相应的终端或服务器中，继续目标附图描述实现本申请实施例的图像处理装置的硬件结构，实现本申请实施例的图像处理装置可以为服务器或各种类型的终端。参考图 16，图 16 为本申请实施例提供的图像处理装置的结构示意图，结合图 16 示出的结构进行说明。

下面对本申请实施例的融合图像处理模型的图像处理装置的结构做详细说明，融合图像处理模型的图像处理装置可以各种形式来实施，如带有融合图像处理模型执行功能的终端，也可以为设置有融合图像处理模型的图像处理功能的服务器，例如前序图 7 中的相应服务器。图 16 为本申请实施例提供的融合图像处理模型的图像处理装置的组成结构示意图，可以理解，图 16 仅仅示出了融合图像处理模型的图像处理装置的示例性结构而非全部结构，根据需要可以实施图 16 示出的部分结构或全部结构。

15 本申请实施例提供的融合图像处理模型的图像处理装置包括：至少一个处理器 1601、存储器 1602、用户接口 1603 和至少一个网络接口 1604。融合图像处理模型的图像处理装置各个组件通过总线系统 805 耦合在一起。可以理解，总线系统 1605 配置为实现这些组件之间的连接通信。总线系统 805 除包括数据总线之外，还包括电源总线、控制总线和状态信号总线。但是为了清楚说明起见，在图 8 中将各种总线都标为总线系统 1605。其中，图像处理装置中各个软件模块的功能，包括：

20 编码器模块 16081，配置为通过融合图像处理模型的状态编码器，对原始图像集进行编码得到所述原始图像集中不同待替换脸部的风格特征；所述编码器模块 16081，配置为通过融合图像处理模型的五官编码器，对不同目标脸部进行编码得到所述不同目标脸部所对应的不同目标脸部的五官特征；解码器模块 16082，配置为将所述风格特征和所述五官特征输入所述融合图像处理模型的融合解码器；所述解码器模块 16082，配置为在所述融合解码器中基于不同待替换脸部的风格特征和不同目标脸部的五官特征进行解码，得到与
25 所述不同待替换脸部具有相同风格的不同目标脸部；其中，所述融合图像处理模型的编码器和解码器的参数是在保持所述单一图像处理模型的编码器和解码器的参数不变的情况下，根据所述单一目标图像进行更新所得到的。

结合图 16 示出的电子设备 160 说明本申请实施例提供的融合图像处理模型的训练方法，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图，可以理解地，图 17 所示的步骤可以由运行融合图像处理模型的图像处理装置的各种电子设备执行，例如可以是如带有融合图像处理模型的图像处理功能的终端、服务器或者服务器集群。下面针对图 17 示出的步骤进行说明。

35 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 1701：视频解码。

其中，通过步骤 1701 视频会被解码得到一系列的帧图像，下面以选取 n (n 为大于或等于 1 的整数) 帧图像为例说明。

步骤 1702：提取 n 帧图像。

40 其中，所提取的 n 帧图像为待替换人脸。

步骤 1703：对 n 帧图像进行人脸检测，如果有人脸，则执行步骤 1704；如果没有人脸，则返回提取 $n+1$ 帧图像继续进行人脸检测。

步骤 1704：对 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 1705；否则跳转到步骤 1708。

步骤 1705: 截取待替换人脸。

步骤 1706: 调用融合换脸模型对 n 帧图像进行目标人脸重构。

融合换脸模型中状态编码器对待替换人脸进行编码得到待替换人脸的状态特征, 五官编码器对目标脸部的进行编码得到目标脸部的五官特征, 融合解码器利用待替换人脸的状态特征和目标脸部的五官特征进行解码, 得到具有风格特征的目标人脸。

步骤 1707: 将重构的目标人脸回贴 n 帧中的待替换人脸。

步骤 1708: 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。

下面对本申请实施例所提供的融合图像处理模型进行说明, 其中, 包括对单一图像处理模型的训练和对融合图像处理模型训练两部分, 图 18A 为用户通过终端所输入的原始图像集中的不同待替换脸部和不同目标脸部示意图。其中, 原始图像集中的不同待替换脸部的数量和不同目标脸部的数量均可以通过用户终端进行调整。图 18B 为用户通过终端所输入的原始图像集中的不同待替换脸部和不同目标脸部示意图。其中, 原始图像集中的不同待替换脸部的数量和不同目标脸部的数量均可以通过用户终端进行调整, 以实现 $M \times N$ 组的待替换脸部和不同目标脸部的处理, 其中, M 和 N 的取值可以根据不同的需求进行调整。

下面介绍对图 18A 和 18B 所示的对不同待替换脸部和不同目标脸部的处理过程。

首先对相关技术中的换脸模型的工作原理进行说明, 其中, 参考图 19, 图 19 为对单一图像处理模型的训练和对融合图像处理模型训练的过程示意图, 其中, 单一图像处理模型训练方法包括每个自编码器分别尝试恢复输入人脸中的扭曲, 把各种扭曲的人脸、还原到对应人的正常的人脸。两个自编码器共用一个编码器, 使得两个人脸经过编码器编码后, 能够映射到同一函数空间中——使得解码器 B 能够解码自编码器 A 中编码器编码出的 A 脸的特征 (例如光照、表情、视线方向等), 以实现相应的换脸的处理效果。

其中, 在使用换脸模型的时候, 如果需要把 B 的脸换到 A 的脸, 把 B 的脸输入到自编码器 A 中, 即可得到换脸后的脸; 同理如果要把 A 的脸换到 B, 则把 A 的脸输入到自编码器 B 中。

在本申请的一些实施例中, 模型的结构中编码器如图 20 所示, 解码器的结构如图 21 所示, 图 20 为本申请实施例提供的编码器的示例性结构图, 如图 20 所示, 包括 4 个卷积层、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 4 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 20 和表 4 所示。

表 4

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

图 21 为本申请实施例提供的解码器的示例性结构图，如图 21 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 5 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 21 和表 5 所示。

表 5

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
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	-

5 前序所示的换脸模型的训练方法如下：

1、记编码器为 Encoder，解码器 A 为 decoder_A，编码器 B 为 decoder_B，扭曲人脸的函数为 warp，使用 A 人脸的数据 x_1 与如下损失函数优化一次编码器与解码器 A 的值：

$$\text{loss_A} = \sum (\text{decoder_A}(\text{encoder}(\text{warp}(x_1))) - x_1)^2$$

2、使用 B 人脸的数据 x_2 与如下损失函数优化一次编码器与解码器 B 的值：

10 $\text{loss_B} = \sum (\text{decoder_B}(\text{encoder}(\text{warp}(x_2))) - x_2)^2$

3、重复步骤 1、2 直到 loss_A 、 loss_B 的和的下降值小于设定阈值 ε 。

前序传统换脸模型的共工作过程中，所产生的缺陷包括：

15 1) 传统的换脸模型仅支持对单人换脸的处理，如果需要换到新的人脸就需要重新训练新的模型，耗时且麻烦，而且每个模型都需要占用自己的存储空间，过程繁琐，增加整个换脸处理流程的耗时，影响用户的使用体验。

2) 训练一个单一换脸模型需要 100 张以上的图片（人脸图像）作为训练样本，而换脸模型的使用者（用户）往往出于隐私的原因，并不会提供自身人脸图像作为训练样本，影响了训练样本图像的离散度。

20 为克服上述缺陷，本申请实施例提供了一种融合图像处理方法，其中，图 22 为本申请实施例提供的融合图像处理方法一个可选的处理流程，可以理解地，图 22 所示的步骤可以由运行融合图像处理模型的图像处理装置的各种电子设备执行，例如可以是如带有融合图像处理模型的图像处理功能的终端、服务器或者服务器集群。下面针对图 22 示出的步骤进行说明，包括以下步骤：

步骤 2201：获取 N 组数据集。

25 其中，N 为大于等于 1 的正整数，N 的取值可以根据不同的需求进行调整，以适配相应的运行融合图像处理模型的图像处理装置的各种电子设备。

步骤 2202：压缩 N 组数据集。

步骤 2203：低内存占用的单一换脸模型的训练。

步骤 2204: 低内存占用的融合换脸模型训练。

下面, 对图 22 所示的步骤进行分别介绍, 其中, 图 23 为本申请中数据集获取示意图, 可以包括以下步骤:

1) 通过网络爬取、下载明星库图片等途径, 收集包含要训练的单一换脸模型的人的图片文件, 每人至少 100 张。

2) 调用人脸检测算法 (这里采用的 MTCNN 人脸检测算法, 其他的方法也可行), 找到人脸的大致位置。

3) 调用人脸对齐算法, 以人眼为中心, 修正人脸框选的位置, 让框选的人脸位置更加精准, 保证在框出的图像上人眼都处于固定位置。

4) 按照自动框选的结果, 截取人脸图片, 放缩到固定尺寸 (优选为 256×256) 并保存, 作为人脸原始图像集合。

5) 人工筛查步骤 4) 得到的集合, 筛除错误识别的人脸数据。

6) 对于每个人, 分别执行前序的步骤 1) - 5), 生成每个目标人物所对应的人脸图像数据集。

图 24 为本申请中数据集压缩处理过程示意图, 可以包括以下步骤:

1) 使用有损 jpg 格式保存所有人脸图像数据集 i。

2) 直接将数据集中所有 jpg 文件载入到内存中, 并不解码图像得到 RGB 矩阵。

3) 前序内存中, 第 i 个人的未解码人脸图片数据, 构成了压缩后人脸数据集 i。

由此, 在后续训练的过程中, 实时解码出随机抽中的人脸图片, 使得不需要把所有的
人脸图片都解码出来, 以保证低内存占用; 而通过巧妙的利用 jpg 格式的压缩率, 实现数
据集的压缩, 且这种压缩的方式, 能够保证每次进行数据增广的时候, 无需从硬盘载入数
据; 对于需要使用的人脸数据, 在内存中解码得到 RGB 矩阵是非常快的, 这总压缩方法
导致的解压耗时是用户可以接受的。

图 25 中本申请实施例提供的融合图像处理方法一个可选的训练流程示意图, 可以理
解地, 图 25 所示的步骤可以由运行融合图像处理模型的图像处理装置的各种电子设备执
行, 例如可以是如带有融合图像处理模型的图像处理功能的终端、服务器或者服务器集群。
下面针对图 25 示出的步骤进行说明, 包括以下步骤:

步骤 2501: 初始化 M 个临时单一换脸模型。

其中, M 为大于等于 1 的正整数, M 的取值可以根据不同的需求进行调整, 以适配相
应的运行融合图像处理模型的图像处理装置的各种电子设备。

步骤 2502: 随机出 M 组下标。

步骤 2503: 对于每个下标, 分别执行如下操作: 记第 i 个下标为 X_i , 取出第 X_i 个单
一换脸模型的网络参数、以及训练时需要用到的参数, 配置到第 i 个临时换脸模型中。

步骤 2504: 按照定义的 loss, 优化所有的 M 个临时单一换脸模型。

对于每个临时单一换脸模型, 执行如下操作:

步骤 2505: 取出第 i 个临时单一换脸模型参数、训练时用到的参数, 分别保存到第
 X_i 个临时换脸模型、第 X_i 个临时换脸模型训练时所用的参数中。

步骤 2506: 重复步骤 2502 至 2505 (优选次数为五万次), 得到训练出的 N 组单一换
脸模型的参数。

其中, 硬件设备能够同时处理的最大单一换脸模型数量为 M, 需要同时训练的单一换
脸模型数量为 N。

下面对图 25 所示的步骤依次进行说明,

其中, 图 26 为本申请实施例中单一换脸模型的网络结构示意图, 其中, 所有的单一
换脸模型, 共用同一个编码器, 图 26 为第 i 个人的换脸模型的网络结构示意图; 其中每个
人的换脸模型类似前述过程中类似, 由编码器后串联一个解码器来实现, 其中, 所有的换

脸模型共用一个编码器，而每个人的换脸模型有自己独立的编码器 i 。记共用的编码器为 encoder，第 i 个人的编码器为 decoder $_i$ ，则第 i 个人的换脸模型的网络结构如图 26 所示，其中，网络各层的参数如下表 6 所示：

表 6

层类型	通道数	卷积核大小	步长	填充	激活函数	
输入(64*64*3)	3	-	-	-	-	
卷积层 1	128	(5, 5)	(2, 2)	same	pReLU	编 码 器
卷积层 2	256	(5, 5)	(2, 2)	same	pReLU	
卷积层 3	512	(5, 5)	(2, 2)	same	pReLU	
卷积层 4	1024	(5, 5)	(2, 2)	same	pReLU	
全连接层 1	1024	(5, 5)	-	-	pReLU	
全连接层 2	16384	(5, 5)	-	-	pReLU	
Reshape 层	-	(4, 4, 1024)	-	-	-	
上采样卷积层 1	512	(3, 3)	(2, 2)	same	pReLU	解 码 器 i
上采样卷积层 2	256	(3, 3)	(2, 2)	same	pReLU	
上采样卷积层 3	256	(3, 3)	(2, 2)	same	pReLU	
上采样卷积层 4	128	(3, 3)	(2, 2)	same	pReLU	
上采样卷积层 5	64	(3, 3)	(2, 2)	same	pReLU	
卷积层 4	3	(3, 3)	(1, 1)	same	pReLU	
输出(128*128*3)	3	-	-	-	-	-

下面对图 25 所示的带解压的数据增广算法进行说明，参考图 27，图 27 为本申请实施例中图像增广示意图。

其中，每次增广处理时，增广出 Y 组数据，记录每次随机增广的时候，增广出 Y 组扭曲人脸、无扭曲人脸用于训练，带解压的数据增广算法的流程包括以下步骤：

步骤 2701：从压缩后人脸数据集 i 中，随机抽取 Y 个未解码人脸图片。

步骤 2702：随机变形，生成无扭曲人脸。

步骤 2703：在随机变形的基础上，进行随机扭曲，得到扭曲人脸。

其中，随机变形的在本申请的一些实施例中的步骤如下：

1) 将输入的检测好的人脸以图像中心为中心，做随机旋转，旋转的角度范围为 α ， α 的优选值为 30° 。

2) 将旋转后的人脸的宽高进行随机拉伸，拉伸的倍数在 $[1-\beta, 1+\beta]$ 之间， β 的优选值为 0.01。

3) 随机平移，将图像在数值和水平方向随机平移 $+\theta_1, +\theta_2$ ，优选 θ_1, θ_2 的值为宽高的 0.01。

4) 五官图像截取：以截取图像中间固定位置的图像，优选截取的百分比为 62.5%。之后将图像缩放到编码器输入的分辨率。

以 512×512 的人脸图像为例，截取宽高行数在 (48, 208) 内的图像，后将得到的 161×161 图像缩放到 128×128 图像。

其中，随机扭曲的方法的处理过程参考图 28，图 28 为本申请实施例中随机扭曲的方法的处理过程示意图，包括：

绘制网格：记截取的人脸图像为正方形图像，为 I ，宽高均为 W ，如下图所示在输入图片上打上 $N \times N$ 的均匀网格（优选的网格的密度为 5×5 （宽 5 个节点，高 5 个节点））。

添加坐标噪声：将每个节点分配坐标 x, y ，记第 i 行第 j 列的坐标分别为 x_{ij}, y_{ij} ， $x_{ij}=i/N$ ， $y_{ij}=j/N$ ，则所有的 $x_{ij}(i=0, \dots, 4, j=0, \dots, 4)$ 、 $y_{ij}(i=0, \dots, 4, j=0, \dots, 4)$ ，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 σ^2 , 方差的优选值为 0.02。

将矩阵 X, Y 视为 图像, 使用图像插值的方法 (优选使用线性插值), 放大矩阵 X, Y, 宽高到 W, 得到新的图像 (即矩阵) X', Y'

5 构造宽高均为 W 的矩阵 X2、Y2, 其中矩阵 X2 的中元素 $x2_{ij}=i/W$, 矩阵 Y2 中元素 $y2_{ij}=j/W$, 其中 $i=0, \dots, W, j=0, \dots, W$ 。

图像重映射: 使用图像重映射的方法, 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 中恰好比 x_{ab} 小的元素记为 $x2_c$, 查找矩阵 Y2 中恰好比 y_{ij} 小的元素记为 $y2_d$, 则可以构造如下四个点:

10 点 A ($x2_c, d, y2_c, d$), B ($x2_{c+1}, d, y2_{c+1}, d$), C ($x2_c, d+1, y2_c, d+1$), D ($x2_{c+1}, d+1, y2_{c+1}, d+1$)。

分别计算点 (x_{ab}, y_{ab}) 到 A、B、C、D 的距离 dA 、 dB 、 dC 、 dD , 根据这四个距离, 加权得到 z_{ab} 的值;

15
$$z_{ab} = dA / (dA + dB + dC + dD) \times I_{c, d} + dB / (dA + dB + dC + dD) \times I_{c+1, d} + dC / (dA + dB + dC + dD) \times I_{c, d+1} + dD / (dA + dB + dC + dD) \times I_{c+1, d+1}。$$

采用与随机变形中相同的方法, 对扭曲得到的人脸进行裁切, 最后, 裁切的扭曲人脸与变形人脸形成 <扭曲人脸> <无扭曲人脸> 数据对, 用于训练单换脸模型。

下面对本申请所涉及的低内存占用的融合换脸模型训练进行介绍, 其中, 利用单一换脸模型, 生成用于训练融合换脸模型的数据, 处理策略包括:

20 1) 不载入所有的人脸数据集, 仅使用部分人脸数据集与正脸数据集, 使得内存占用减少: 选用的人脸数据集, 由从全部人脸数据集中随机取出, 取出的比例为 20%。而正脸数据集由人工, 从随机抽取的人脸数据集中筛选得到。需要提及的是, 这种不使用所有数据的方法, 也大大减少了人工标注的量, 提高构建融合换脸模型的整个过程。

2) 同时只处理固定数量的临时换脸模型用于构造训练融合换脸模型的数据: 记 M 为机器能够同时处理的单一换脸模型的数量, N 为所有单一换脸模型的数量, 构造 M 个临时单一换脸模型与步骤 1 中的人脸、正脸, 进行实时解压&增广出训练用数据。

其处理过程参考图 29A 和图 29B, 图 29A 为本申请实施例提供的融合图像处理方法一个可选的训练流程示意图; 图 29B 为本申请实施例提供的融合图像处理方法数据结构示意图; 可以理解地, 图 29A 所示的步骤可以由运行融合图像处理模型的图像处理装置的各种电子设备执行, 例如可以是如带有融合图像处理模型的图像处理功能的终端、服务器或者服务器集群。下面针对图 29A 示出的步骤进行说明, 包括以下步骤:

步骤 2901: 从所有的人脸数据集中, 随机抽取至少 20% 的数据构造 人脸数据集。

步骤 2902: 在所有人脸数据集中, 随机抽取并人工标记, 标出至少 5% 数据集总量的正脸图片, 组成人脸数据集。

35 步骤 2903: 将人脸数据集、正脸数据集图片转为 jpg 格式, 并载入到计算机内存中且不解码。

步骤 2904: 初始化 M 个临时单一换脸模型, 其中 M 为机器能够处理的单一换脸模型的数量。

其中, M 为大于等于 1 的正整数, M 的取值可以根据不同的需求进行调整, 以适配相应的运行融合图像处理模型的图像处理装置的各种电子设备。

步骤 2905: 载入所有 N 个单一换脸模型的参数到内存中。

其中, N 为大于等于 1 的正整数, N 的取值可以根据不同的需求进行调整, 以适配相应的运行融合图像处理模型的图像处理装置的各种电子设备。

步骤 2906: 初始化融合换脸模型。

45 步骤 2907: 重复循环步骤 10 万次, 以完成训练融合换脸模型。

其中，循环步骤包括：

a) 从 N 个单一换脸模型中，随机抽取 M 个下标，对于每个下标进行如下操作：

i) 对于第 i 下标 X_i ，将第 X_i 个单一换脸模型参数输入到临时单一换脸模型 i 中。

ii) 从人脸数据集中随机抽取 Y 个，进行与前述相同数据增广，得到扭曲人脸图像。

5 iii) 从正脸数据集中随机抽取 Y 个，进行与前述相同数据增广，得到无扭曲正脸图像；
将扭曲人脸图像，输入到临时换脸模型 i 中，得到临时无扭曲人脸 i。

iv) 将无扭曲正脸输入到临时换脸模型 i 中，得到临时无扭曲正脸 i。

v) 将扭曲人脸图像、临时无扭曲正脸 X，带入到融合换脸模型中，得到换脸输出。

vi) 计算换脸输出、与临时无扭曲人脸 i 的平方误差，得到 $Loss_fuse_i$ 。

10 b) 对所有 $Loss_fuse_i$ 求和，其中 $i=1, \dots, M$ ，得到 $Loss_fuse$ 。

c) 根据 $Loss_fuse$ ，优化一次融合换脸模型的参数。

下面，融合图像处理方法可以继续参考图 30A 和 30B，其中，图 30A 为本申请实施例提供的融合图像处理方法一个可选的使用流程示意图，图 30B 为本申请实施例提供的融合图像处理方法数据结构示意图；可以理解地，图 30A 所示的步骤可以由运行融合图像处理的图像处理装置的各种电子设备执行，例如可以是如带有融合图像处理模型的图像处理的终端、服务器或者服务器集群。下面针对图 30A 示出的步骤进行说明。

对本申请融合换脸模型的使用方法进行说明，包括以下步骤：

步骤 3001：解码视频。

步骤 3002：对每一帧进行如下 a) 到 d) 的处理，记未处理的帧为待处理帧 j。

20 a) 进行人脸检测，若没有检测到待换人脸，则直接对 处理帧 j 编码，结束对该帧的处理；否则进行步骤 b)。

b) 将检测到的待处理人脸、参考正脸 i 输入到融合换脸模型中，得到已换人脸。

c) 使用人脸融合的方法将已换人脸替换到待处理帧 j 中，得到已处理帧 j。

d) 对已处理帧 j 进行编码。

25 步骤 3003：得到已处理视频。

参见图 31，图 31 是本申请实施例提供的融合图像处理模型的图像处理 100 的架构示意图，包括区块链网络 200（示例性示出了共识节点 210-1 至共识节点 210-3）、认证中心 300、业务主体 400 和业务主体 500，下面分别进行说明。

30 区块链网络 200 的类型是灵活多样的，例如可以为公有链、私有链或联盟链中的任意一种。以公有链为例，任何业务主体的电子设备例如用户终端和服务器，都可以在不需要授权的情况下接入区块链网络 200；以联盟链为例，业务主体在获得授权后其下辖的电子设备（例如终端/服务器）可以接入区块链网络 200，此时，成为区块链网络 200 中的客户端节点。

35 在一些实施例中，客户端节点可以只作为区块链网络 200 的观察者，即提供支持业务主体发起交易（例如，用于上链存储数据或查询链上数据）功能，对于区块链网络 200 的共识节点 210 的功能，例如排序功能、共识服务和账本功能等，客户端节点可以缺省或者有选择性（例如，取决于业务主体的具体业务需求）地实施。从而，可以将业务主体的数据和业务处理逻辑最大程度迁移到区块链网络 200 中，通过区块链网络 200 实现数据和业务处理过程的可信和可追溯。

40 区块链网络 200 中的共识节点接收来自不同业务主体（例如如图 1 中示出的业务主体 400 和业务主体 500）的客户端节点（例如，图 1 中示出的归属于业务主体 400 的客户端节点 410、以及归属于数据库运营商系统的客户端节点 510）提交的交易，执行交易以更新账本或者查询账本，执行交易的各种中间结果或最终结果可以返回业务主体的客户端节点中显示。

45 例如，客户端节点 410/510 可以订阅区块链网络 200 中感兴趣的事件，例如区块链网

络 200 中特定的组织/通道中发生的交易,由共识节点 210 推送相应的交易通知到客户端节点 410/510,从而触发客户端节点 410/510 中相应的业务逻辑。

下面以多个业务主体接入区块链网络以实现同目标脸部和相应不同待替换脸部的管理为例,说明区块链网络的示例性应用。

参见图 31,管理环节涉及的多个业务主体,如业务主体 400 可以是基于人工智能的融合图像处理模型的图像处理,业务主体 500 可以是带有不同待替换脸部处理功能的显示系统,从认证中心 300 进行登记注册获得各自的数字证书,数字证书中包括业务主体的公钥、以及认证中心 300 对业务主体的公钥和身份信息签署的数字签名,用来与业务主体针对交易的数字签名一起附加到交易中,并被发送到区块链网络,以供区块链网络从交易中取出数字证书和签名,验证消息的可靠性(即是否未经篡改)和发送消息的业务主体的身份信息,区块链网络会根据身份进行验证,例如是否具有发起交易的权限。业务主体下辖的电子设备(例如终端或者服务器)运行的客户端都可以向区块链网络 200 请求接入而成为客户端节点。

业务主体 400 的客户端节点 410 配置为通过融合图像处理模型的状态编码器,对原始图像集进行编码得到所述原始图像集中不同待替换脸部的风格特征;

通过融合图像处理模型的五官编码器,对不同目标脸部进行编码得到所述不同目标脸部所对应的不同目标脸部的五官特征;将所述风格特征和所述五官特征输入所述融合图像处理模型的融合解码器;在所述融合解码器中基于不同待替换脸部的风格特征和不同目标脸部的五官特征进行解码,得到与所述不同待替换脸部具有相同风格的不同目标脸部;其中,所述融合图像处理模型的编码器和解码器的参数是在保持所述单一图像处理模型的编码器和解码器的参数不变的情况下,根据所述单一目标图像进行更新所得到的;并将融合图像处理模型所对应的与不同待替换脸部具有相同风格的不同目标脸部以及相应的目标脸部发送至区块链网络 200。

其中,将融合图像处理模型所对应的与不同待替换脸部具有相同风格的不同目标脸部以及相应的目标脸部发送至区块链网络 200,可以预先在客户端节点 410 设置业务逻辑,当形成与不同待替换脸部具有相同风格的不同目标脸部时,客户端节点 410 将融合图像处理模型所对应的与不同待替换脸部具有相同风格的不同目标脸部以及相应的目标脸部发送至区块链网络 200,也可以由业务主体 400 的业务人员在客户端节点 410 中登录,手动打包将融合图像处理模型所对应的与不同待替换脸部具有相同风格的不同目标脸部以及相应的目标脸部发送至区块链网络 200。在发送时,客户端节点 410 根据待处理的目标文本及对应的与所触发的与所述目标文本的类型相匹配的业务进程信息生成对应更新操作的交易,在交易中指定了实现更新操作需要调用的智能合约、以及向智能合约传递的参数,交易还携带了客户端节点 410 的数字证书、签署的数字签名(例如,使用客户端节点 410 的数字证书中的私钥,对交易的摘要进行加密得到),并将交易广播到区块链网络 200 中的共识节点 210。

区块链网络 200 中的共识节点 210 中接收到交易时,对交易携带的数字证书和数字签名进行验证,验证成功后,根据交易中携带的业务主体 400 的身份,确认业务主体 400 是否是具有交易权限,数字签名和权限验证中的任何一个验证判断都将导致交易失败。验证成功后签署节点 210 自己的数字签名(例如,使用节点 210-1 的私钥对交易的摘要进行加密得到),并继续在区块链网络 200 中广播。

区块链网络 200 中的共识节点 210 接收到验证成功的交易后,将交易填充到新的区块中,并进行广播。区块链网络 200 中的共识节点 210 广播的新区块时,会对新区块进行共识过程,如果共识成功,则将新区块追加到自身所存储的区块链的尾部,并根据交易的结果更新状态数据库,执行新区块中的交易:对于提交更新待处理的目标文本及对应的与所触发的与所述目标文本的类型相匹配的业务进程信息的交易,在状态数据库中添加包括待

处理的目标文本及对应的与所触发的与所述目标文本的类型相匹配的业务进程信息的键值对。

业务主体 500 的业务人员在客户端节点 510 中登录, 输入同目标脸部或者不同待替换脸部查询请求, 客户端节点 510 根据同目标脸部或者不同待替换脸部查询请求生成对应更新操作/查询操作的交易, 在交易中指定了实现更新操作/查询操作需要调用的智能合约、以及向智能合约传递的参数, 交易还携带了客户端节点 510 的数字证书、签署的数字签名 (例如, 使用客户端节点 510 的数字证书中的私钥, 对交易的摘要进行加密得到), 并将交易广播到区块链网络 200 中的共识节点 210。

区块链网络 200 中的共识节点 210 中接收到交易, 对交易进行验证、区块填充及共识一致后, 将填充的新区块追加到自身所存储的区块链的尾部, 并根据交易的结果更新状态数据库, 执行新区块中的交易: 对于提交的更新某一目标脸部对应的人工识别结果的交易, 根据人工识别结果更新状态数据库中该目标脸部对应的键值对; 对于提交的查询某个与不同待替换脸部具有相同风格的不同目标脸部以及相应的目标脸部, 从状态数据库中查询同目标脸部对应的键值对, 并返回交易结果。

值得说明的是, 在图 31 中示例性地示出了将融合图像处理模型所对应的与不同待替换脸部具有相同风格的不同目标脸部以及相应的目标脸部直接上链的过程, 但在另一些实施例中, 对于不同目标脸部的数据量较大的情况, 客户端节点 410 可同目标脸部同目标脸部的哈希以及相应的不同待替换脸部的哈希成对上链, 将原始的同目标脸部以及相应的不同待替换脸部存储于分布式文件系统或数据库。客户端节点 510 从分布式文件系统或数据库获取到同目标脸部以及相应的不同待替换脸部后, 可结合区块链网络 200 中对应的哈希进行校验, 从而减少上链操作的工作量。

作为区块链的示例, 参见图 32, 图 32 是本申请实施例提供的区块链网络 200 中区块链的结构示意图, 每个区块的头部既可以包括区块中所有交易的哈希值, 同时也包含前一个区块中所有交易的哈希值, 新产生的交易的记录被填充到区块并经过区块链网络中节点的共识后, 会被追加到区块链的尾部从而形成链式的增长, 区块之间基于哈希值的链式结构保证了区块中交易的防篡改和防伪造。

下面说明本申请实施例提供的区块链网络的示例性的功能架构, 参见图 33, 图 33 是本申请实施例提供的区块链网络 200 的功能架构示意图, 包括应用层 201、共识层 202、网络层 203、数据层 204 和资源层 205, 下面分别进行说明。

资源层 205 封装了实现区块链网络 200 中的各个节点 210 的计算资源、存储资源和通信资源。

数据层 204 封装了实现账本的各种数据结构, 包括以文件系统中的文件实现的区块链、键值型的状态数据库和存在性证明 (例如区块中交易的哈希树)。

网络层 203 封装了点对点 (P2P, Point to Point) 网络协议、数据传播机制和数据验证机制、接入认证机制和业务主体身份管理的功能。

其中, P2P 网络协议实现区块链网络 200 中节点 210 之间的通信, 数据传播机制保证了交易在区块链网络 200 中的传播, 数据验证机制用于基于加密学方法 (例如数字证书、数字签名、公/私钥对) 实现节点 210 之间传输数据的可靠性; 接入认证机制用于根据实际的业务场景对加入区块链网络 200 的业务主体的身份进行认证, 并在认证通过时赋予业务主体接入区块链网络 200 的权限; 业务主体身份管理用于存储允许接入区块链网络 200 的业务主体的身份、以及权限 (例如能够发起的交易类型)。

共识层 202 封装了区块链网络 200 中的节点 210 对区块达成一致性的机制 (即共识机制)、交易管理和账本管理的功能。共识机制包括 POS、POW 和 DPOS 等共识算法, 支持共识算法的可插拔。

交易管理用于验证节点 210 接收到的交易中携带的数字签名, 验证业务主体的身份信

息，并根据身份信息判断确认其是否具有权限进行交易（从业务主体身份管理读取相关信息）；对于获得接入区块链网络 200 的授权的业务主体而言，均拥有认证中心颁发的数字证书，业务主体利用自己的数字证书中的私钥对提交的交易进行签名，从而声明自己的合法身份。

5 账本管理用于维护区块链和状态数据库。对于取得共识的区块，追加到区块链的尾部；执行取得共识的区块中的交易，当交易包括更新操作时更新状态数据库中的键值对，当交易包括查询操作时查询状态数据库中的键值对并向业务主体的客户端节点返回查询结果。支持对状态数据库的多种维度的查询操作，包括：根据区块向量号（例如交易的哈希值）查询区块；根据区块哈希值查询区块；根据交易向量号查询区块；根据交易向量号查询交易；根据业务主体的账号（向量号）查询业务主体的账号数据；根据通道名称查询通道中的区块链。

应用层 201 封装了区块链网络能够实现的各种业务，包括交易的溯源、存证和验证等。本申请具有以下有益技术效果：

1、融合图像处理模型的压缩效果明显，部署于相应的服务器中时，融合图像处理模型的体积减小到之前传统技术的 37.92%，有效降低了其在终端上部署、调用的复杂度。

2、实现技术条件下（部署于相应的服务器中），523 个单一换脸模型的融合压缩，由压缩前 1 个编码器、523 解码器，模型共占用存储空间 31GB；而经过本申请所提供的融合图像处理模型训练方法所生成的融合换脸模型，仅占用硬件资源 738MB，压缩率达到了 2.32%，并且随着硬件设备的能力提升可以实现更多数量的单一图像处理模型的融合。

3、通过该本实施例训练得到的融合图像处理模型通用性更好，同一个模型能够对多个目标脸部图像进行交换。

4、使用更为简单，只需要输入目标人脸的正脸图片，即可把视频图像帧中的人脸换成目标人脸的人脸，无需提前训练，便有在用户终端（处理能力弱的电子设备）中部署，实现了图像处理模型的大范围部署。

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。

工业实用性

本申请实施例中通过在训练模型时，获取图像集，并对图像集进行压缩处理，以形成包括至少一组经过压缩的人脸图像的原始图像集；根据原始图像集中的单一待替换脸部，更新单一图像处理模型的编码器的参数和单一图像处理模型的解码器的参数，以使单一图像处理模型的解码器能够基于单一待替换脸部的风格特征进行解码，且解码得到与单一待替换脸部具有相同风格的单一目标脸部；在保持单一图像处理模型的编码器和解码器的参数不变的情况下，根据原始图像集中的不同待替换脸部和不同目标脸部更新融合图像处理模型的编码器和解码器的参数，以使融合图像处理模型中的解码器能够基于不同待替换脸部的风格特征进行解码，且解码得到与相应待替换脸部具有相同风格的不同目标脸部，由此，实现了对融合图像处理模型的解码器的训练与原始图像集的解耦，减少融合图像处理模型的复杂度低，节省硬件存储资源的同时，经过训练的融合图像处理模型能够批量处理不同的原始图像，增强了融合图像处理模型的通用性；同时，图像处理模型中的参数进行全量的更新的过程降低了硬件资源的占用，减少了硬件设备成本的提升，实现了图像处理模型的大范围部署，有利于图像处理模型部署在移动终端中，减轻用户的使用成本。

权利要求书

1、一种融合图像处理模型的训练方法，所述方法由电子设备执行，所述融合图像处理模型的训练方法包括：

获取图像集，并对所述图像集进行压缩处理，以形成包括至少一组经过压缩的人脸图像的原始图像集；

根据原始图像集中的单一待替换脸部，更新单一图像处理模型的编码器的参数和所述单一图像处理模型的解码器的参数，以使所述单一图像处理模型的解码器能够基于所述单一待替换脸部的风格特征进行解码，且解码得到与所述单一待替换脸部具有相同风格的单一目标脸部；

在保持所述单一图像处理模型的编码器和解码器的参数不变的情况下，根据所述原始图像集中的不同待替换脸部和不同目标脸部更新融合图像处理模型的编码器和解码器的参数，以使所述融合图像处理模型中的解码器能够基于不同待替换脸部的风格特征进行解码，且解码得到与相应待替换脸部具有相同风格的不同目标脸部。

2、根据权利要求 1 所述的方法，其中，所述根据原始图像集中的单一待替换脸部，更新单一图像处理模型的编码器的参数和所述单一图像处理模型的解码器的参数，包括：

根据原始图像集中的单一待替换脸部，更新单一图像处理模型中编码器的参数，以使所述单一图像处理模型的编码器能够对所述单一待替换脸部进行编码，且编码得到所述单一待替换脸部的风格特征；

根据所述原始图像集的单一待替换脸部，更新所述单一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单一待替换脸部的风格特征进行解码；

在保持所述单一图像处理模型的编码器的参数不变的情况下，根据单一目标图像中的单一目标脸部更新所述解码器的参数。

3、根据权利要求 2 所述的方法，其中，所述根据原始图像集中的单一待替换脸部，更新单一图像处理模型中编码器的参数，包括：

将扭曲原始图像集中单一待替换脸部的五官图像，代入由所述单一图像处理模型的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单一图像处理模型的编码器的参数，其中，所述扭曲原始图像集是对所述原始图像集进行扭曲处理得到。

4、根据权利要求 2 所述的方法，其中，所述根据所述原始图像集的单一待替换脸部，更新所述单一图像处理模型中解码器的参数，包括：

将扭曲原始图像集中单一待替换脸部的五官图像，代入由所述单一图像处理模型的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集是对所述原始图像集进行扭曲处理得到。

5、根据权利要求 2 所述的方法，其中，所述在保持所述单一图像处理模型的编码器的参数不变的情况下，根据单一目标图像中单一目标脸部的风格特征更新所述解码器的参数，包括：

将扭曲单一目标图像中单一目标脸部的五官图像，代入由所述单一图像处理模型的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单一图像处理模型的编码器的参数不变的情况下，执行以下处理：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，所述扭曲单一目标图像是对所述单一目标图像进行扭曲处理得到。

6、根据权利要求 2-5 任一项所述的方法，其中，所述方法还包括：

根据所述编码器的参数和所述解码器的参数，迭代更新与所述单一图像处理模型对应的单一图像处理模型集合中的所有单一图像处理模型的编码器的参数和解码器的参数，直至所述单一图像处理模型集合中的所有单一图像处理模型的编码器的参数和解码器的参数的损失函数满足各自对应的收敛条件，其中，所述单一图像处理模型集合中的单一图像处理模型数量与所述单一图像处理模型的使用环境相适配。

7、根据权利要求1所述的方法，其中，

当根据原始图像集中的单一待替换脸部，更新所述单一图像处理模型中编码器的参数之前，所述方法还包括：

从样本图像集合的各个样本图像中截取包括所述单一待替换脸部的所述原始图像集；

将所截取的各个所述原始图像集基于单一待替换脸部的特征点进行对齐。

8、根据权利要求7所述的方法，其中，所述获取图像集，并对所述图像集进行压缩处理，以形成包括至少一组经过压缩的人脸图像的原始图像集，包括：

爬取网络中的图像资源，并对所爬取的图像资源中的原始图像进行人脸特征定位与数据清洗；

对所述原始图像集进行随机增广处理；

将经过随机增广处理的原始图像集进行随机扭曲处理，得到扭曲的所述原始图像集；

截取扭曲的所述原始图像集中单一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，以适配所述单一图像处理模型的编码器支持输入的分辨率；

对经过缩放处理的所述五官图像进行压缩，触发形成与所述五官图像的格式相对应的未解压状态的人脸图像的二进制数据，以作为包括至少一组经过压缩的人脸图像的原始图像集。

9、根据权利要求8所述的方法，其中，所述对所述原始图像集进行随机增广处理，包括：

基于所述原始图像集的中心进行随机旋转；

将随机旋转后的所述原始图像集进行随机缩放；

将随机缩放后所述原始图像集进行随机平移。

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

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

将所述经过随机增广处理的原始图像集进行网格化，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。

11、根据权利要求1所述的方法，其中，所述在保持所述单一图像处理模型的编码器和解码器的参数不变的情况下，根据所述原始图像集中的不同待替换脸部和不同目标脸部更新融合图像处理模型的编码器和解码器的参数，包括：

以所述单一图像处理模型的解码器的解码结果作为所述融合图像处理模型中的训练监督数据；

根据所述原始图像集中的不同待替换脸部，更新所述融合图像处理模型中状态编码器的参数，以使所述融合图像处理模型中的状态编码器能够对所述不同待替换脸部分别进行编码，且编码得到分别对应所述不同待替换脸部的状态特征；

根据所述不同目标脸部，更新所述融合图像处理模型中五官编码器的参数，以使所述融合图像处理模型中的五官编码器能够对所述不同目标脸部分别进行编码，且编码得到分别对应所述不同目标脸部的五官特征；

根据所述融合图像处理模型中的状态编码器的编码结果和五官编码器编码结果，更新

所述融合图像处理模型中融合解码器的参数。

12、根据权利要求 11 所述的方法，其中，所述根据所述不同目标脸部，更新所述融合图像处理模型中五官编码器的参数，包括：

5 将不同目标脸部的五官特征，代入由所述融合图像处理模型的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融合图像处理模型中五官编码器的参数。

13、根据权利要求 11 所述的方法，其中，所述方法还包括：

根据所述融合图像处理模型的使用环境，抽取所述原始图像集中的第一数量的人脸图像作为第一训练集合；

10 抽取所述原始图像集中的第二数量的人脸图像作为第二训练集合，其中，所述第二训练集合中的人脸图像为待替换脸部的正脸图像；

响应于所述第一训练集合和所述第二训练集合，加载与所述第一训练集合和所述第二训练集合分别对应的未解压状态的人脸图像的二进制数据，以及相应数量的单一换脸模型，以完成所述融合换脸模型的训练初始化。

15 14、一种融合图像处理模型的图像处理方法，所述方法由电子设备执行，所述融合图像处理模型的图像处理方法包括：

通过融合图像处理模型的状态编码器，对原始图像集进行编码得到所述原始图像集中不同待替换脸部的风格特征；

20 通过融合图像处理模型的五官编码器，对不同目标脸部进行编码得到所述不同目标脸部所对应的不同目标脸部的五官特征；

将所述风格特征和所述五官特征输入所述融合图像处理模型的融合解码器；

在所述融合解码器中基于不同待替换脸部的风格特征和不同目标脸部的五官特征进行解码，得到与所述不同待替换脸部具有相同风格的不同目标脸部；

25 其中，所述融合图像处理模型的编码器和解码器的参数是在保持所述单一图像处理模型的编码器和解码器的参数不变的情况下，根据所述单一目标图像进行更新所得到的。

15、一种融合图像处理模型的训练装置，所述训练装置包括：

图像处理模块，配置为获取图像集，并对所述图像集进行压缩处理，以形成包括至少一组经过压缩的人脸图像的原始图像集；

30 单一图像处理模型训练模块，配置为根据原始图像集中的单一待替换脸部，更新单一图像处理模型的编码器的参数和所述单一图像处理模型的解码器的参数，以使，

所述单一图像处理模型的解码器能够基于所述单一待替换脸部的风格特征进行解码，且解码得到与所述单一待替换脸部具有相同风格的单一目标脸部；

35 融合图像处理模型训练模块，配置为在保持所述单一图像处理模型的编码器和解码器的参数不变的情况下，根据所述原始图像集中的不同待替换脸部和不同目标脸部更新融合图像处理模型的编码器和解码器的参数，以使，

所述融合图像处理模型中的解码器能够基于不同待替换脸部的风格特征进行解码，且解码得到与相应待替换脸部具有相同风格的不同目标脸部。

16、一种图像处理装置，所述图像处理装置包括：

40 编码器模块，配置为通过融合图像处理模型的状态编码器，对原始图像集进行编码得到所述原始图像集中不同待替换脸部的风格特征；

所述编码器模块，配置为通过融合图像处理模型的五官编码器，对不同目标脸部进行编码得到所述不同目标脸部所对应的不同目标脸部的五官特征；

解码器模块，配置为将所述风格特征和所述五官特征输入所述融合图像处理模型的融合解码器；

45 所述解码器模块，配置为在所述融合解码器中基于不同待替换脸部的风格特征和不同

目标脸部的五官特征进行解码，得到与所述不同待替换脸部具有相同风格的不同目标脸部；

其中，所述融合图像处理模型的编码器和解码器的参数是在保持所述单一图像处理模型的编码器和解码器的参数不变的情况下，根据所述单一目标图像进行更新所得到的。

5 17、一种电子设备，所述电子设备包括：

存储器，配置为存储可执行指令；

处理器，配置为运行所述存储器存储的可执行指令时，实现权利要求 1 至 13 任一项所述的融合图像处理模型的训练方法。

18、一种电子设备，所述电子设备包括：

10 存储器，配置为存储可执行指令；

处理器，配置为运行所述存储器存储的可执行指令时，实现权利要求 14 所述的融合图像处理模型的图像处理方法。

19、一种计算机可读存储介质，存储有可执行指令，所述可执行指令被处理器执行时实现权利要求 1 至 13 任一项所述的融合图像处理模型的训练方法，或者实现权利要求

15 14 所述的融合图像处理模型的图像处理方法。

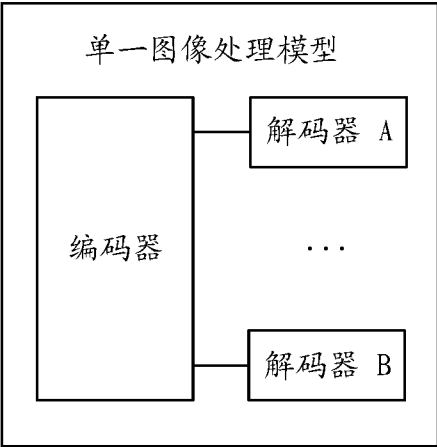


图 1

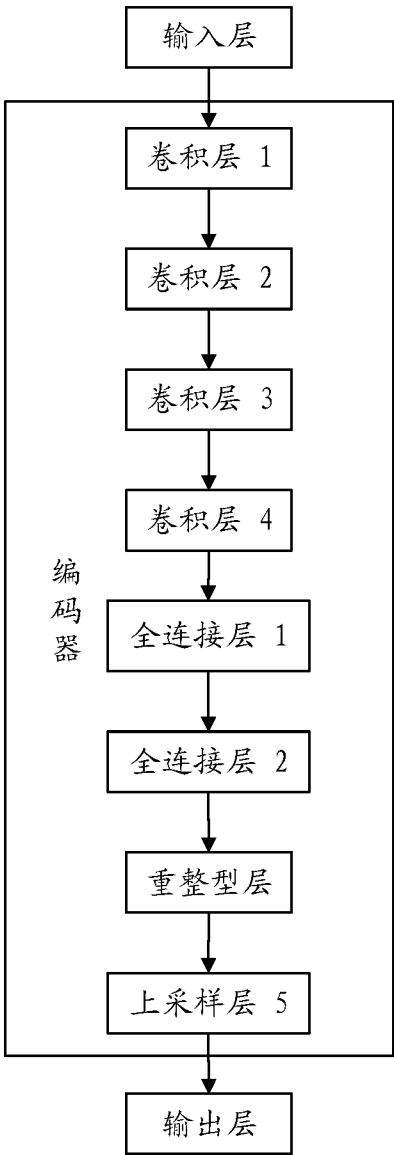


图 2

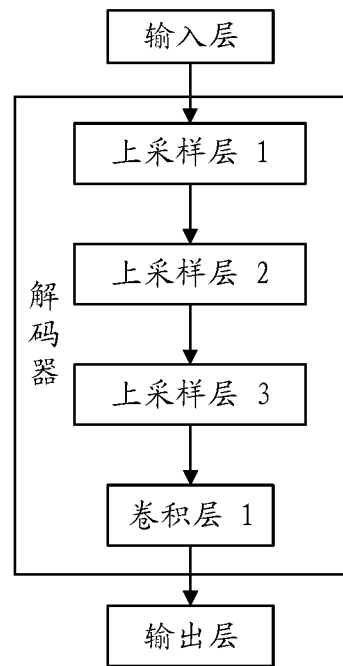


图 3

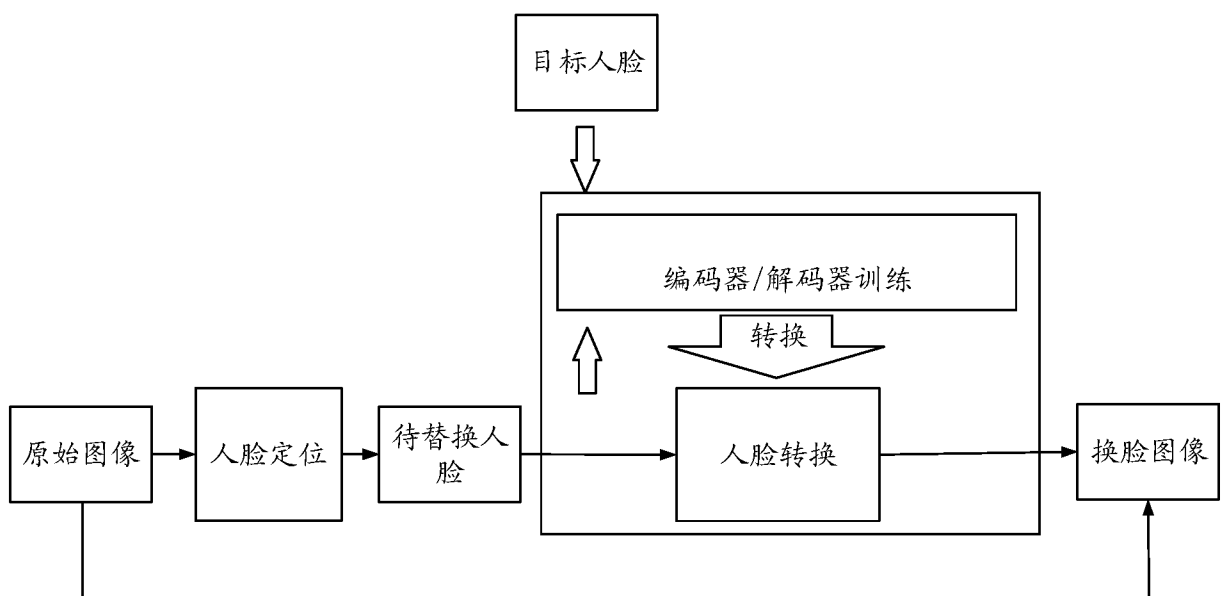


图 4

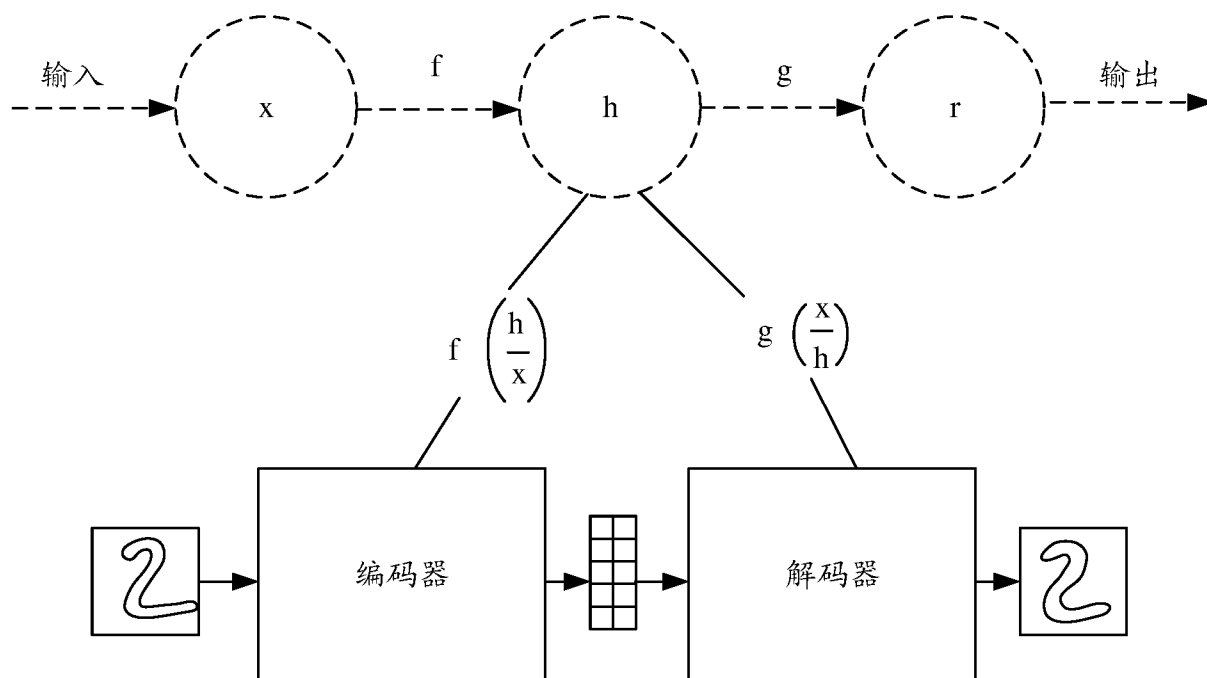


图 5

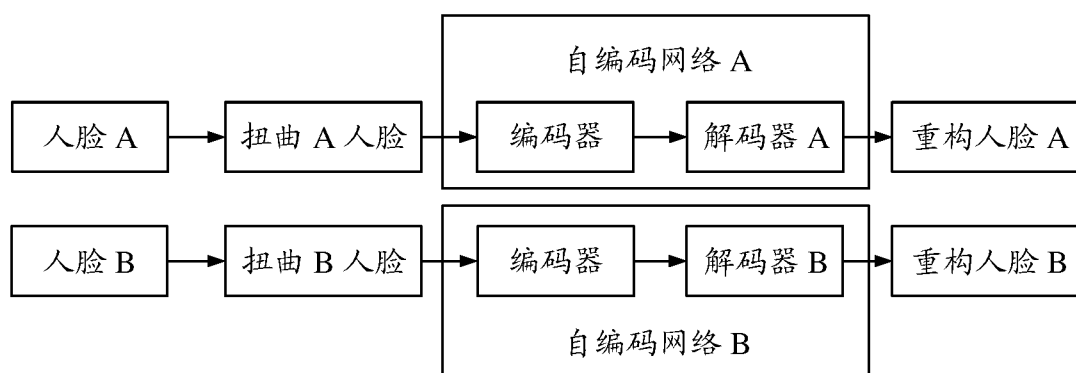


图 6

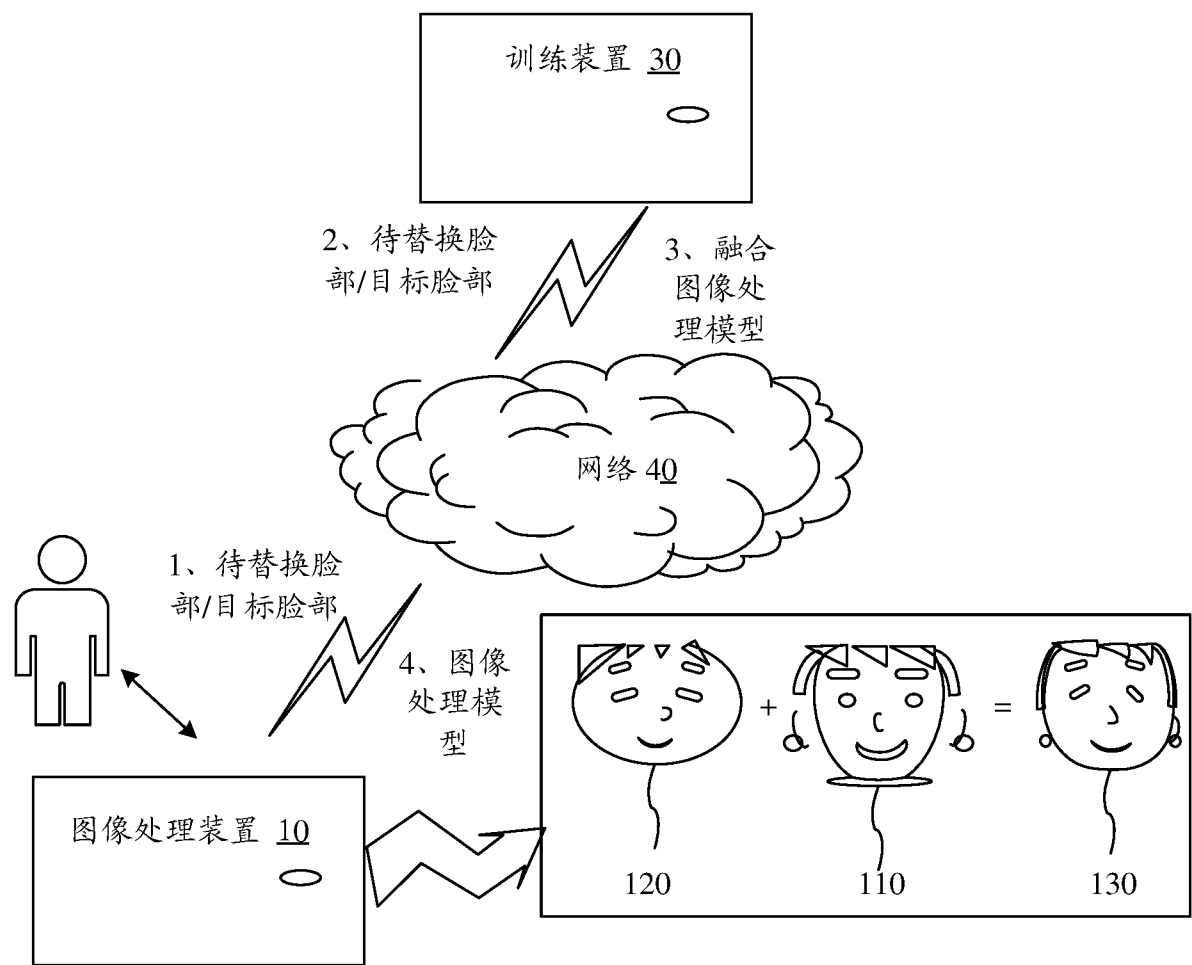


图 7

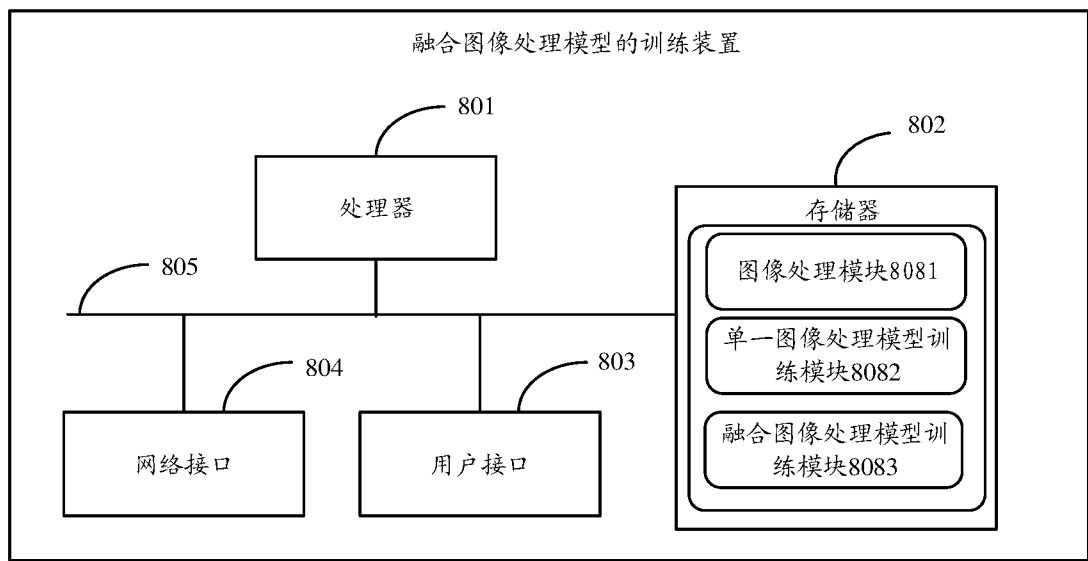


图 8

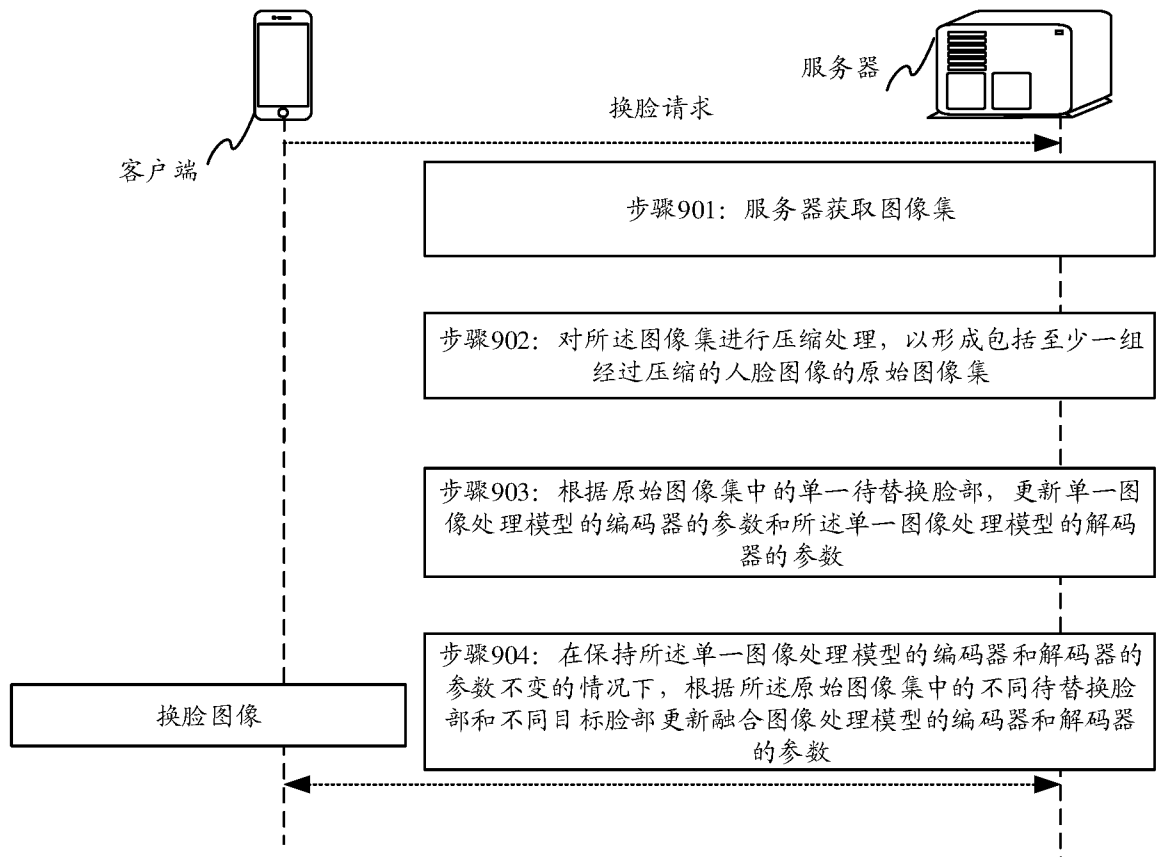


图 9

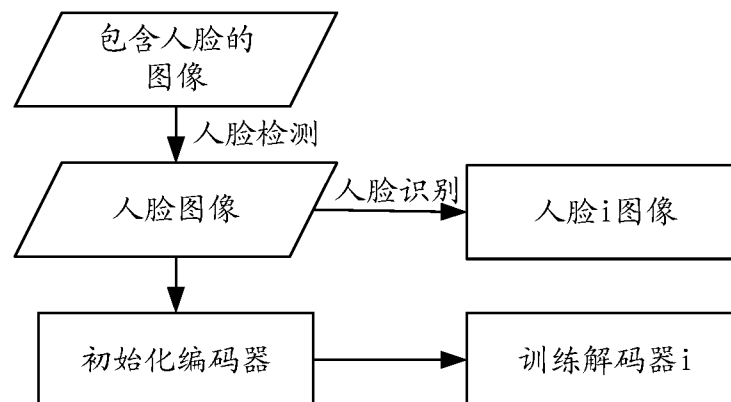


图 10

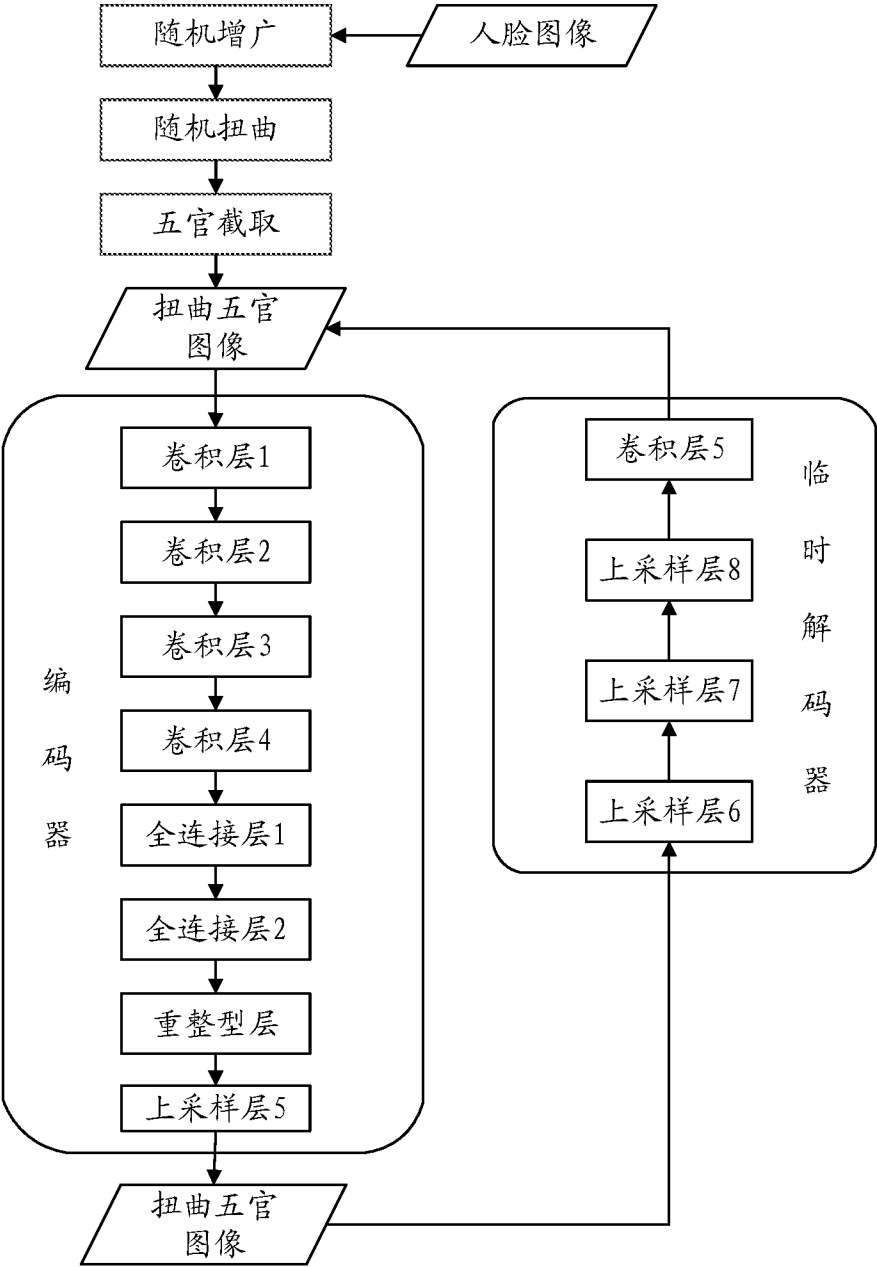


图 11A

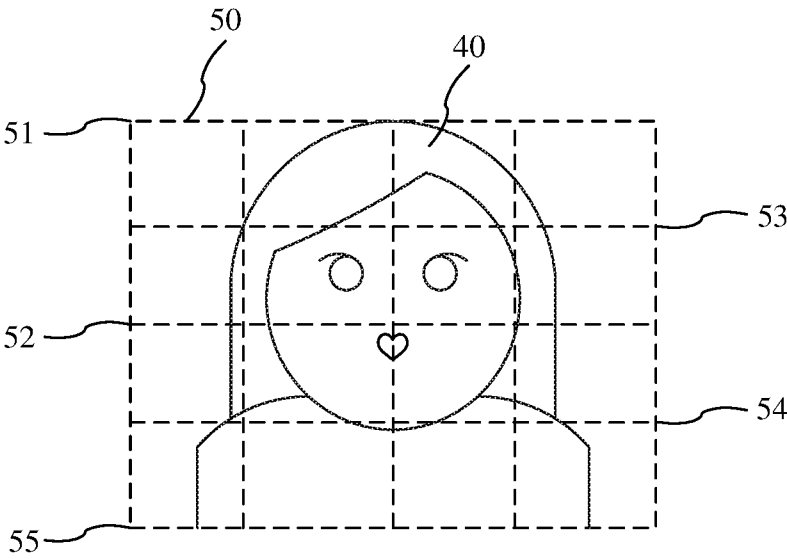


图 11B

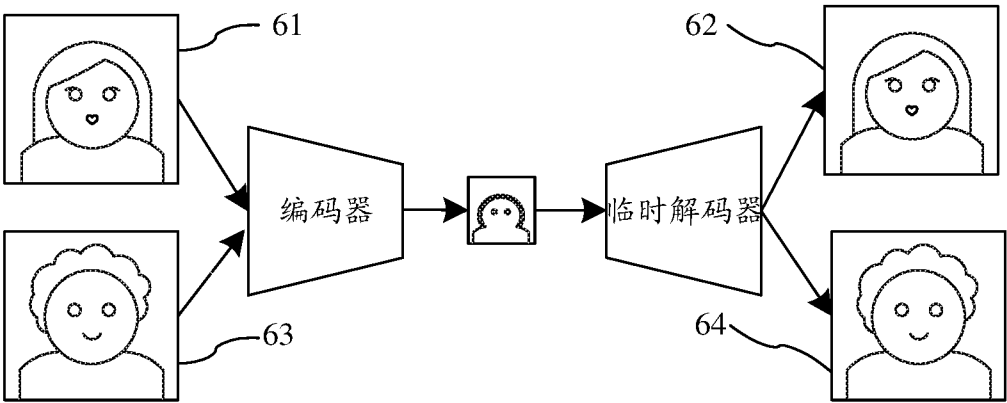


图 12

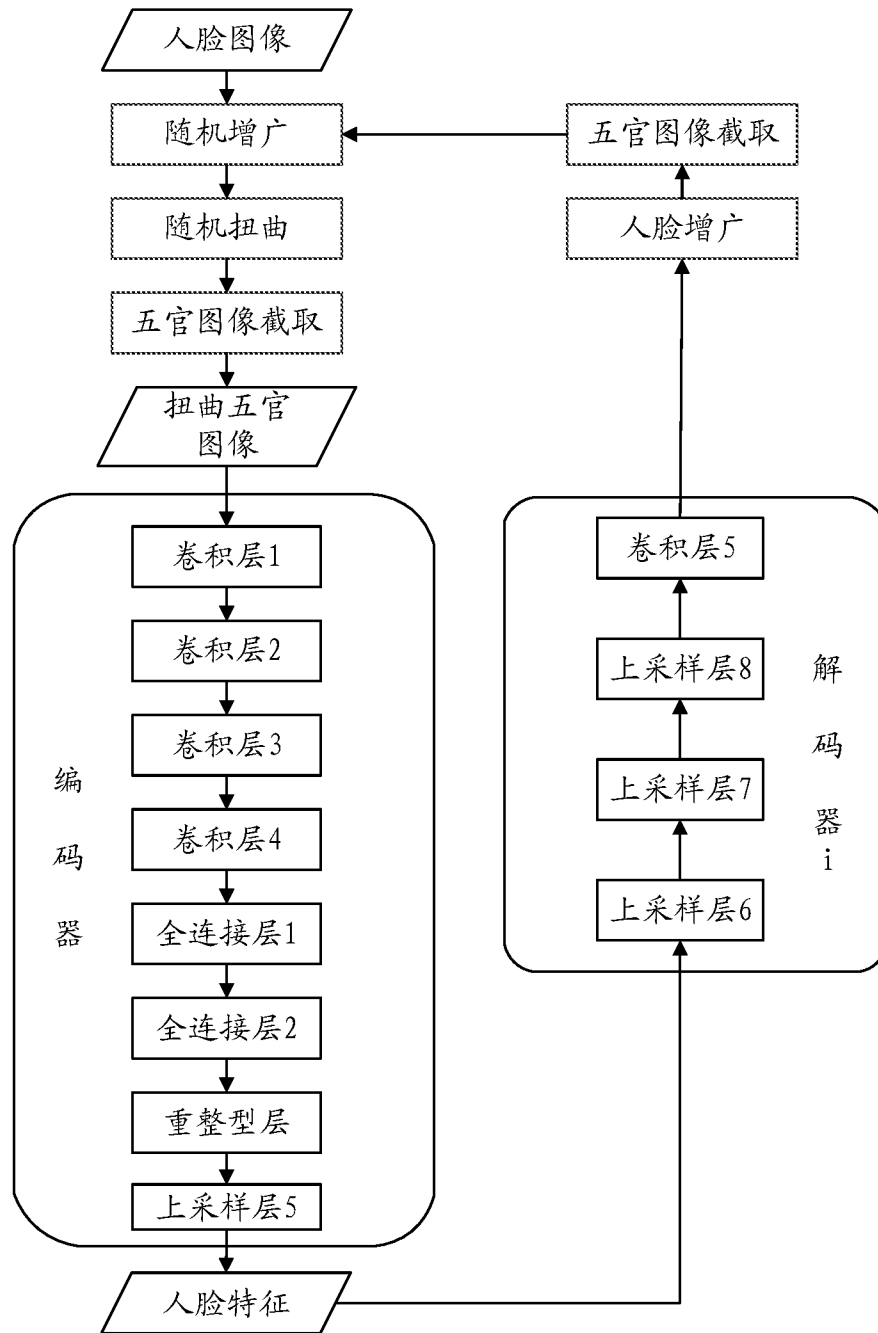


图 13

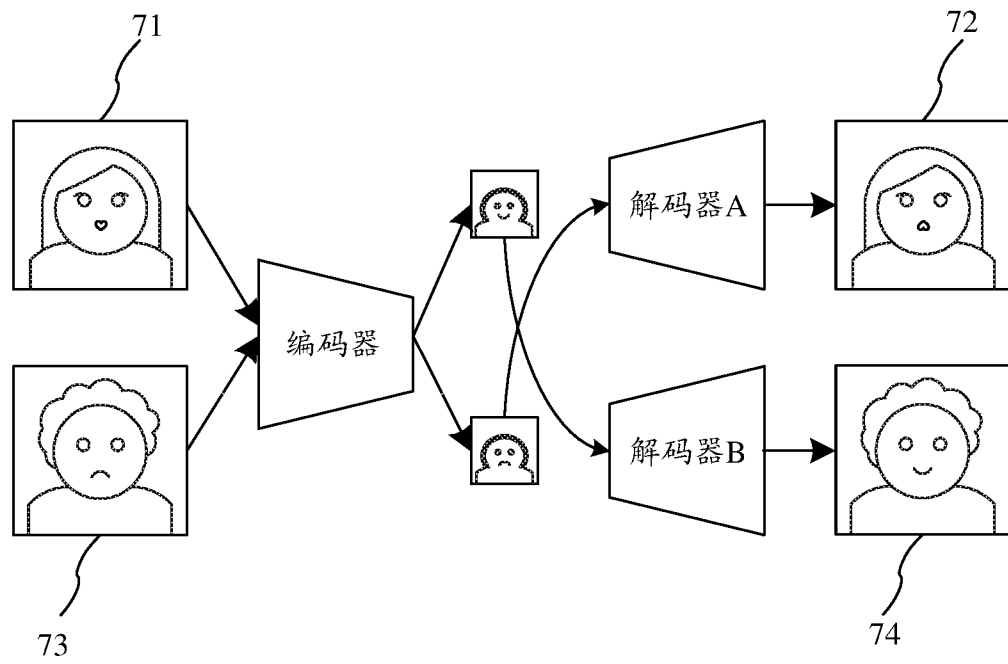


图 14

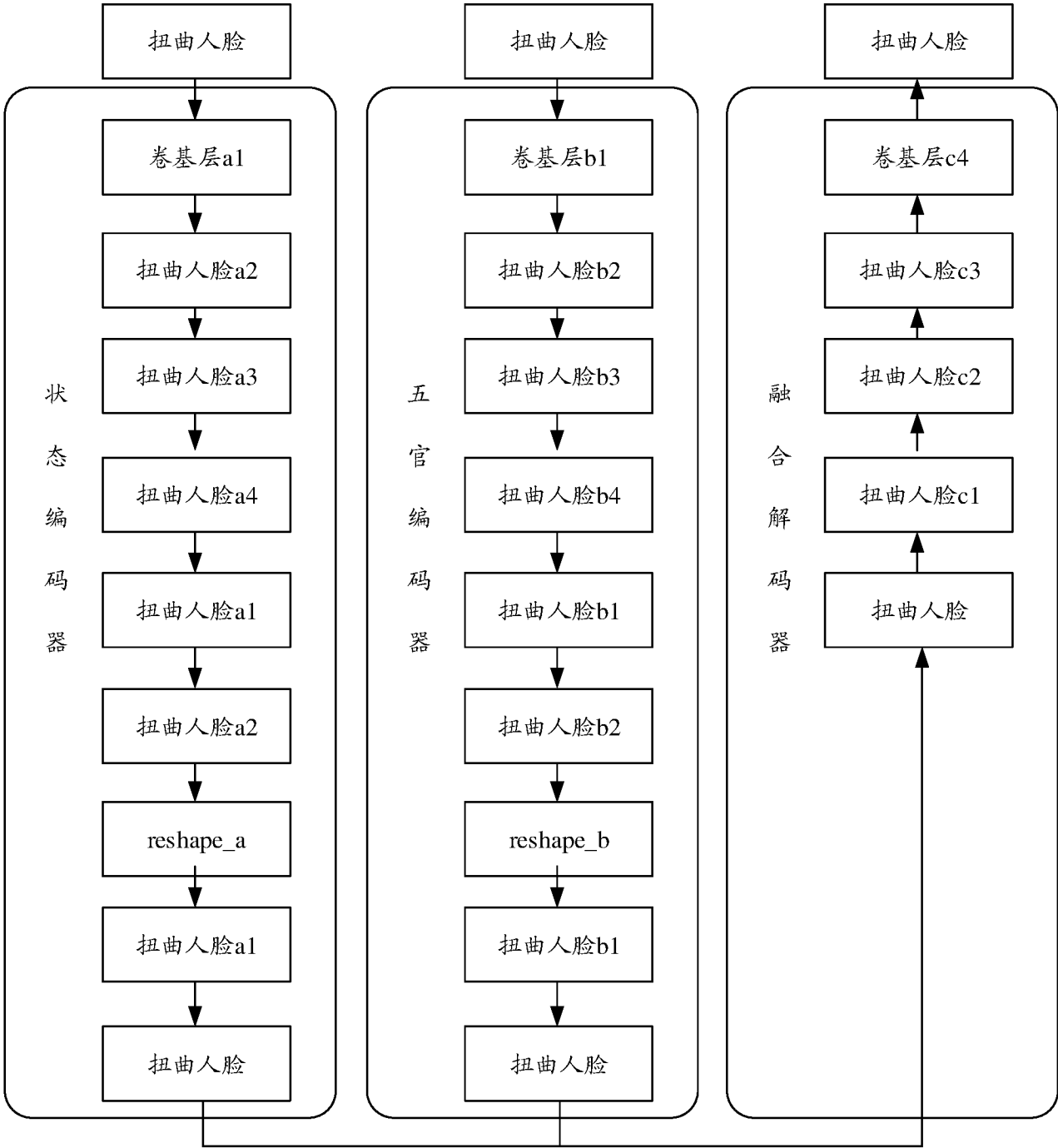


图 15

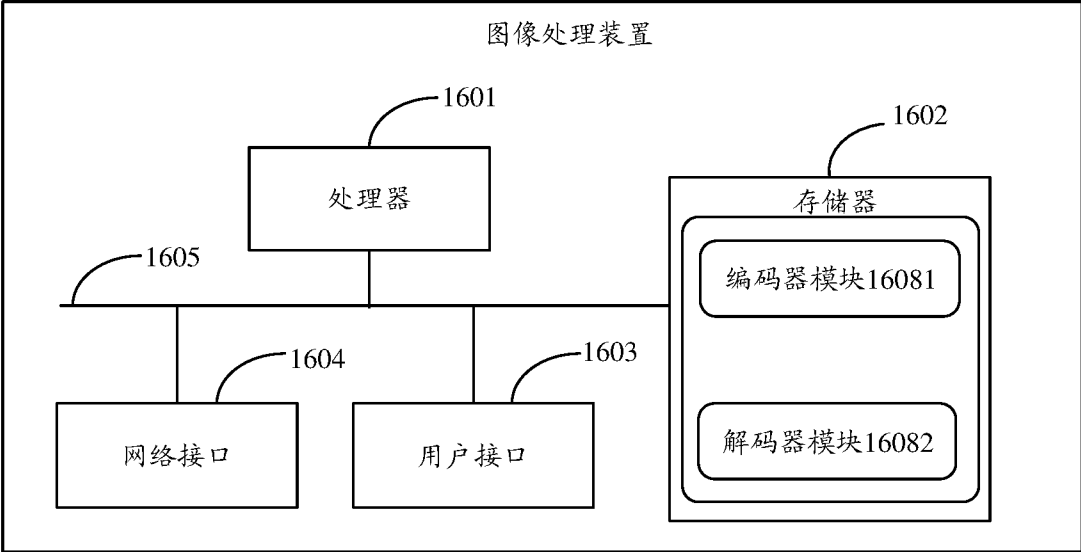


图 16

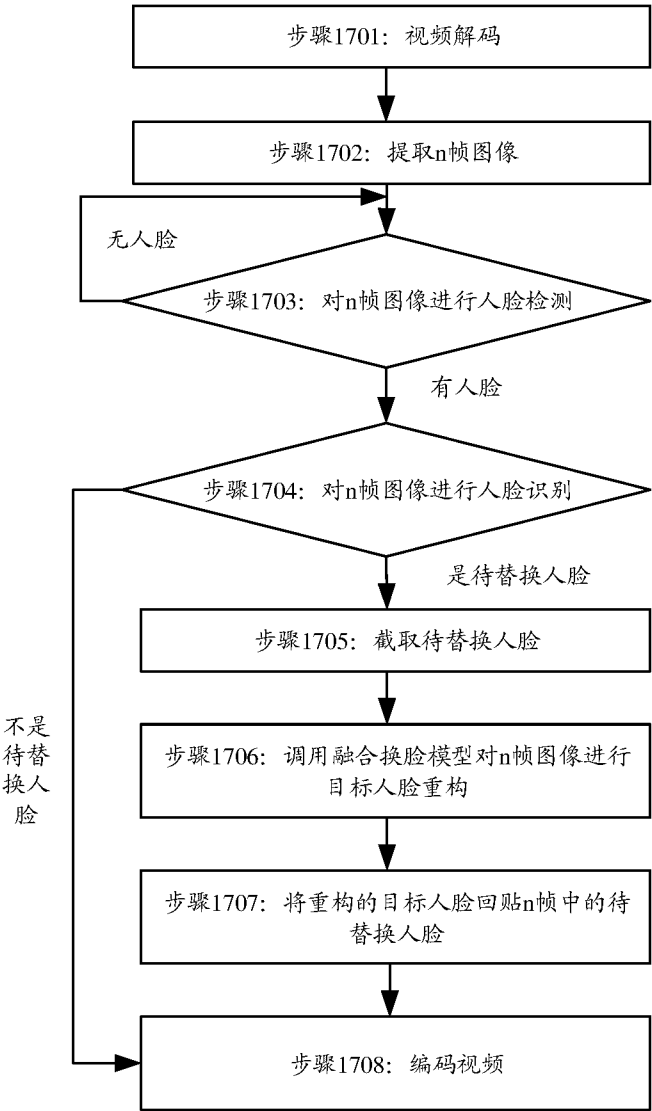


图 17

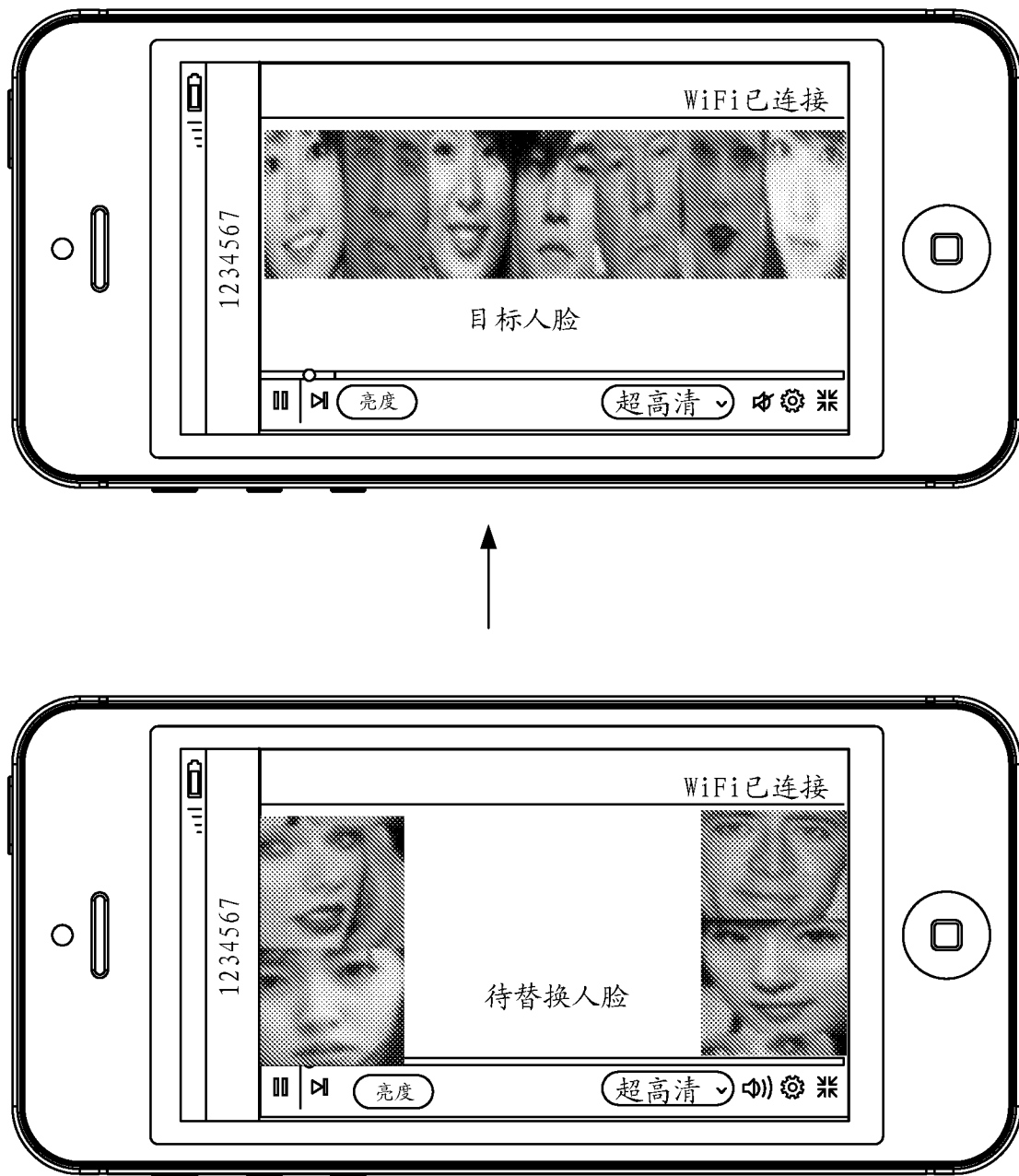


图 18A

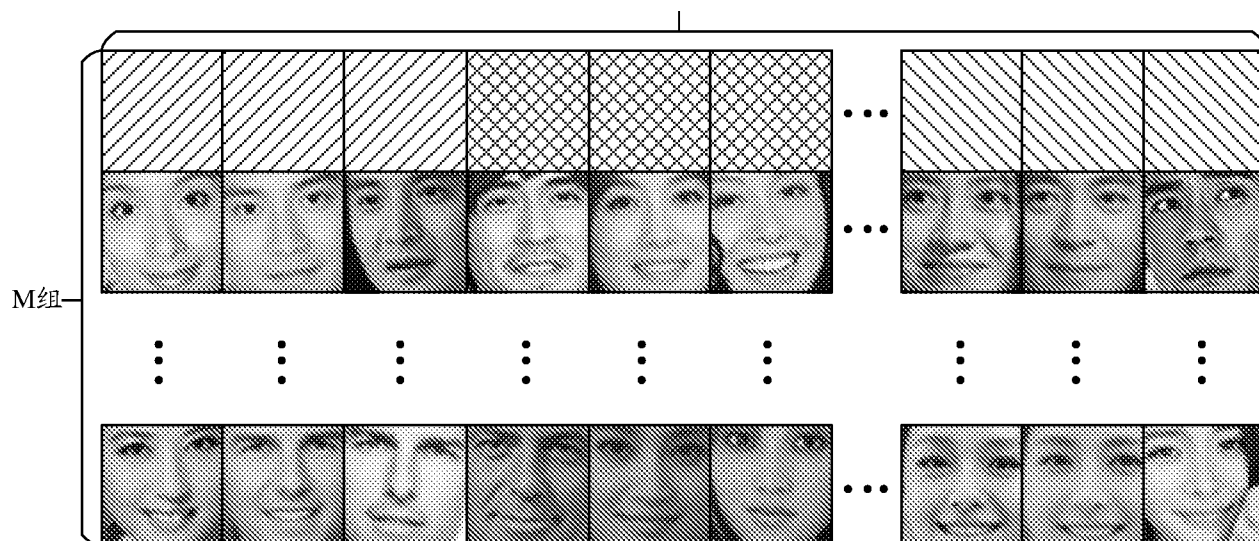


图 18B

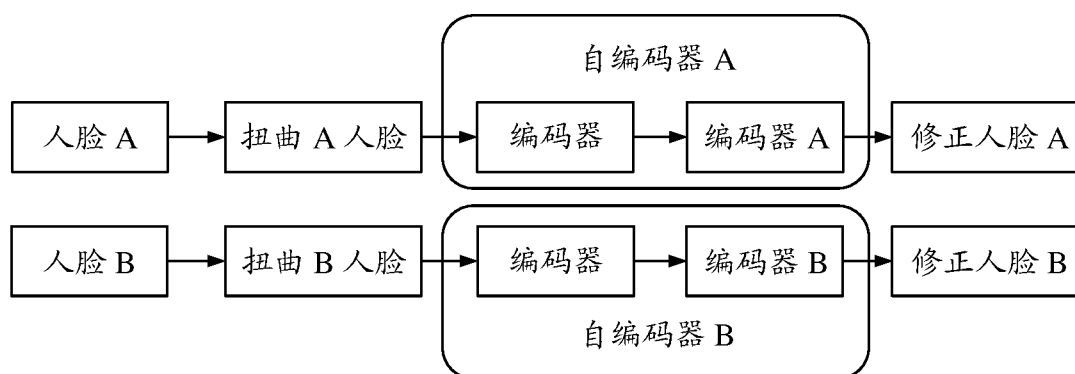


图 19

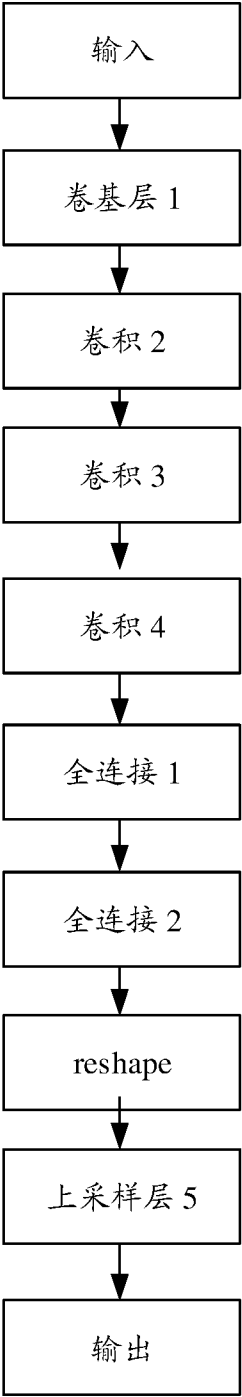


图 20

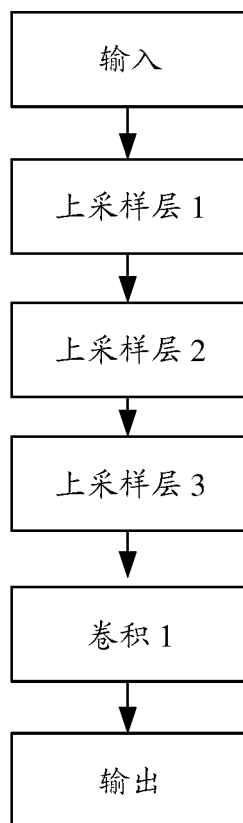


图 21

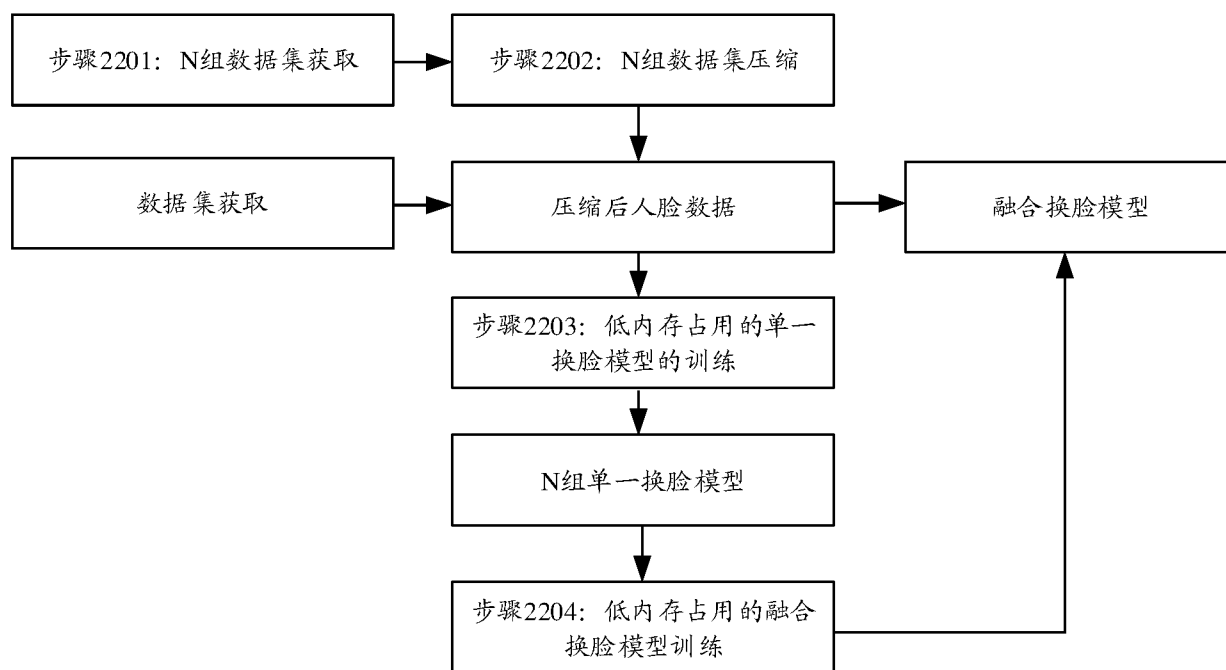


图 22

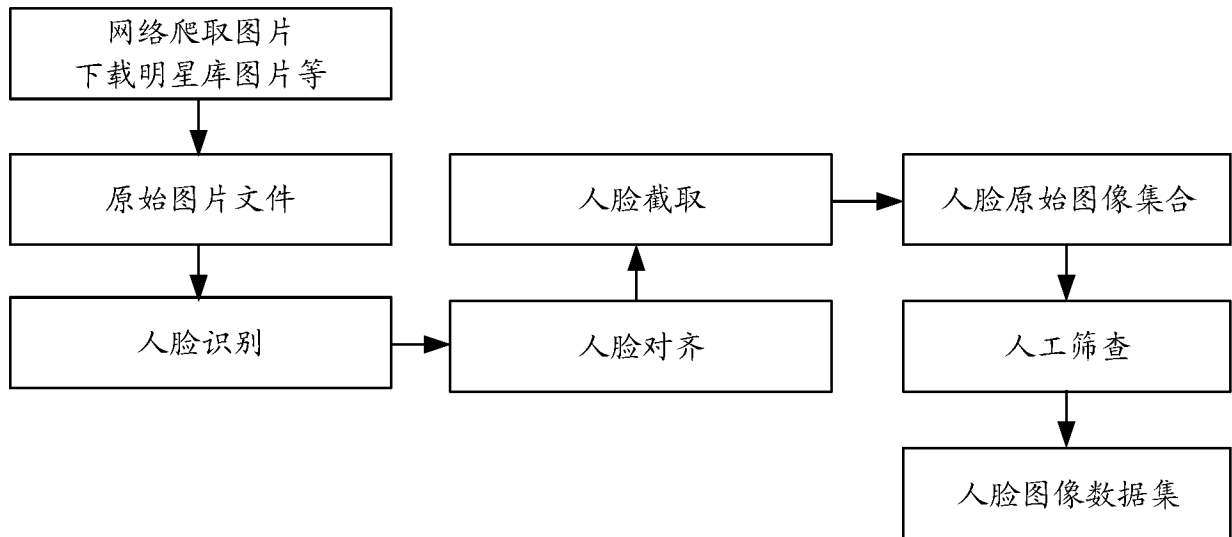


图 23

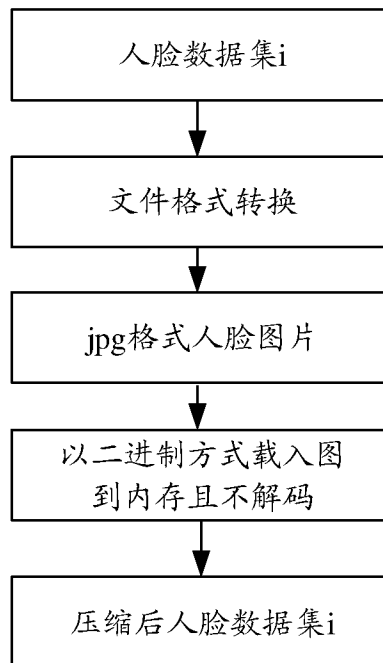


图 24

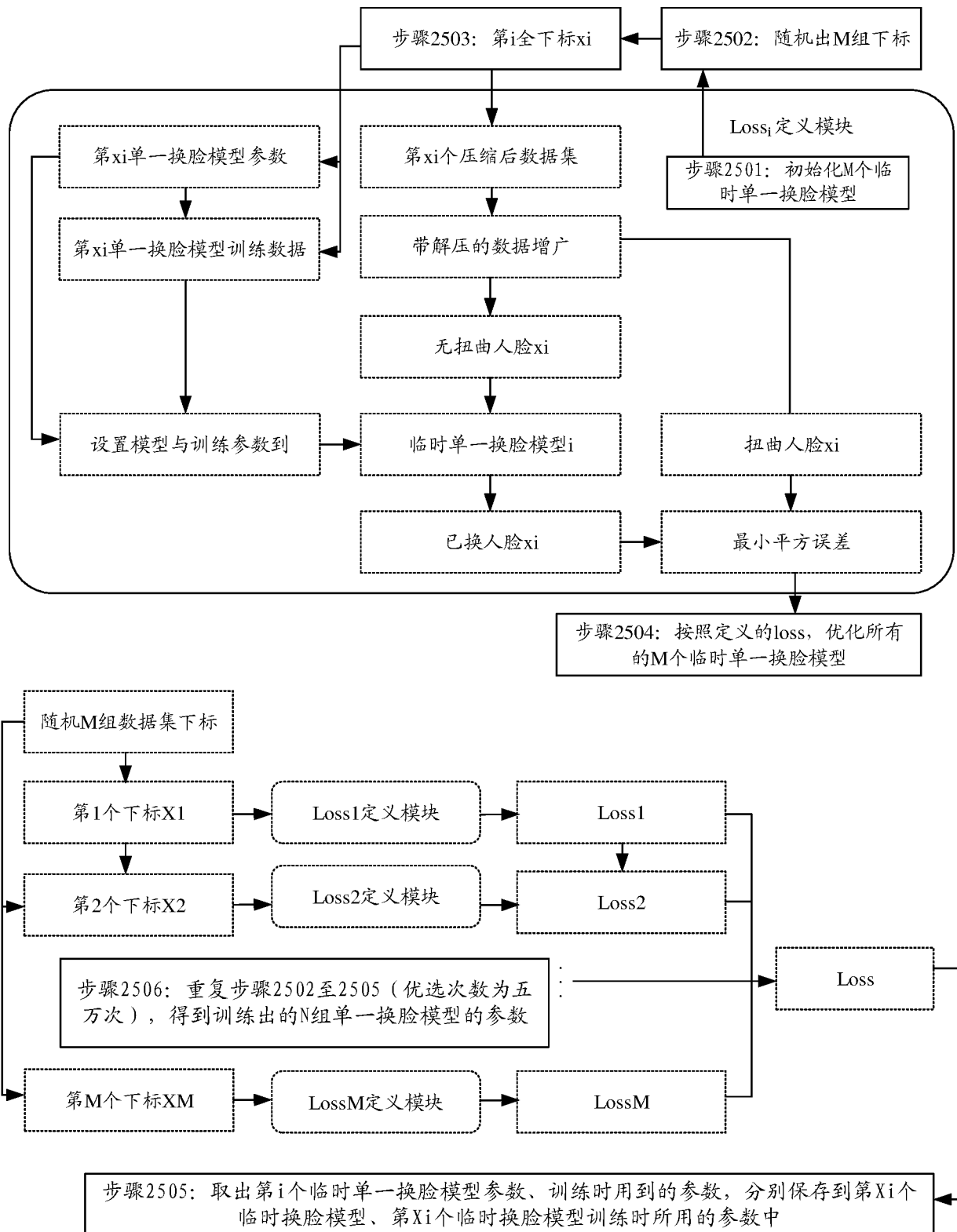


图 25

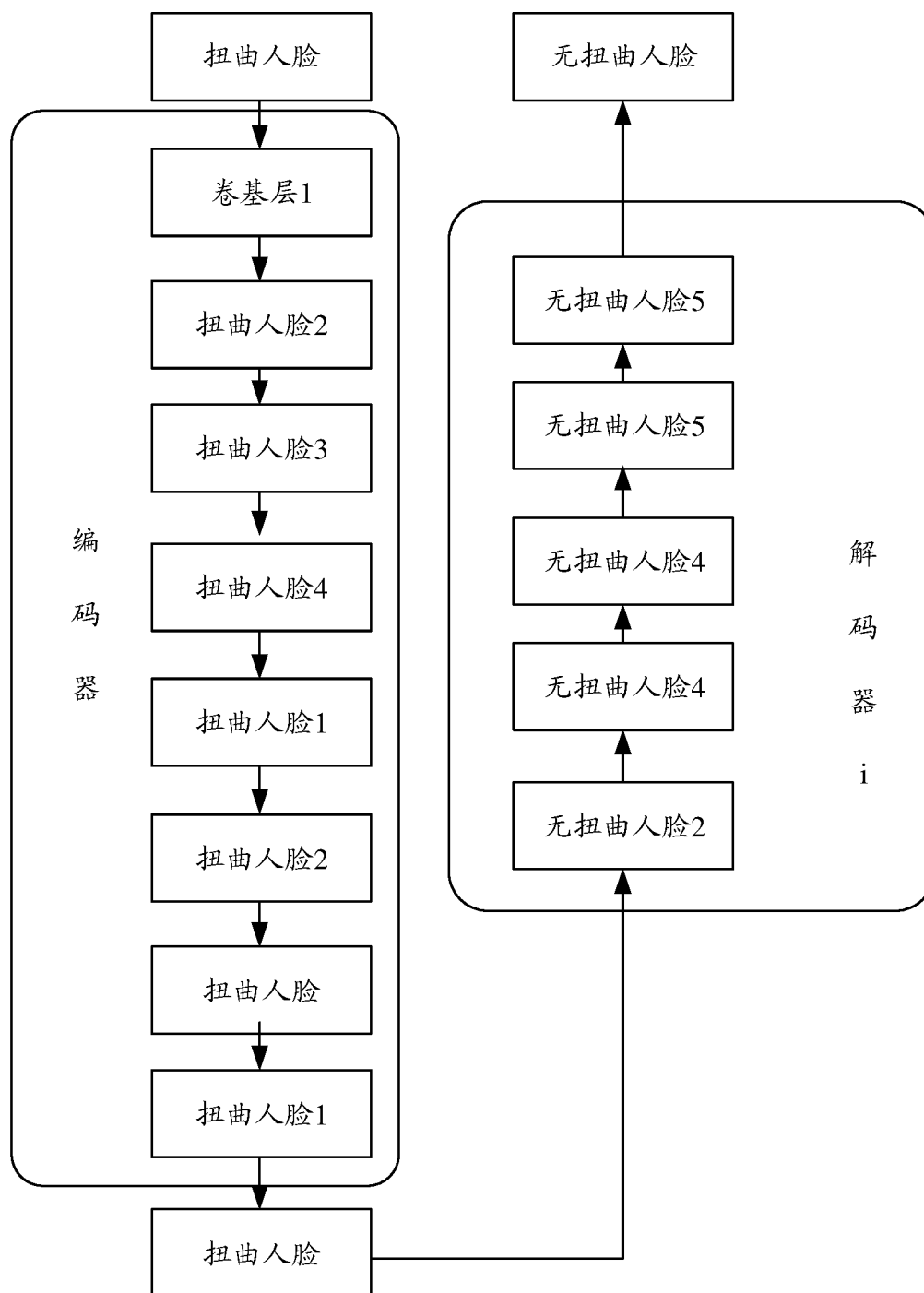


图 26

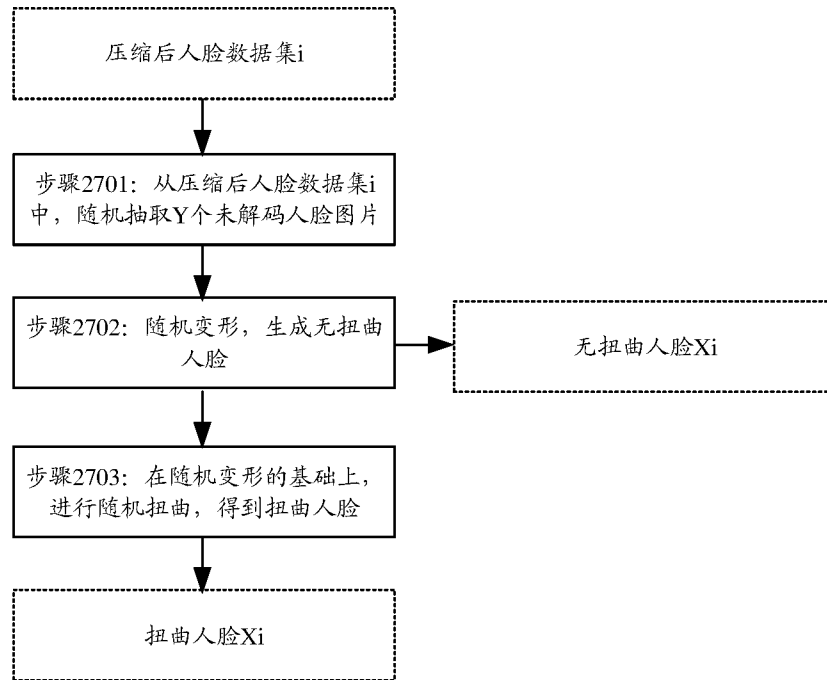


图 27

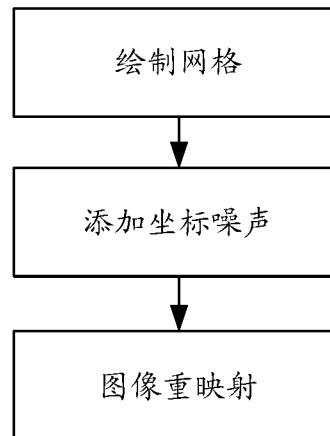


图 28

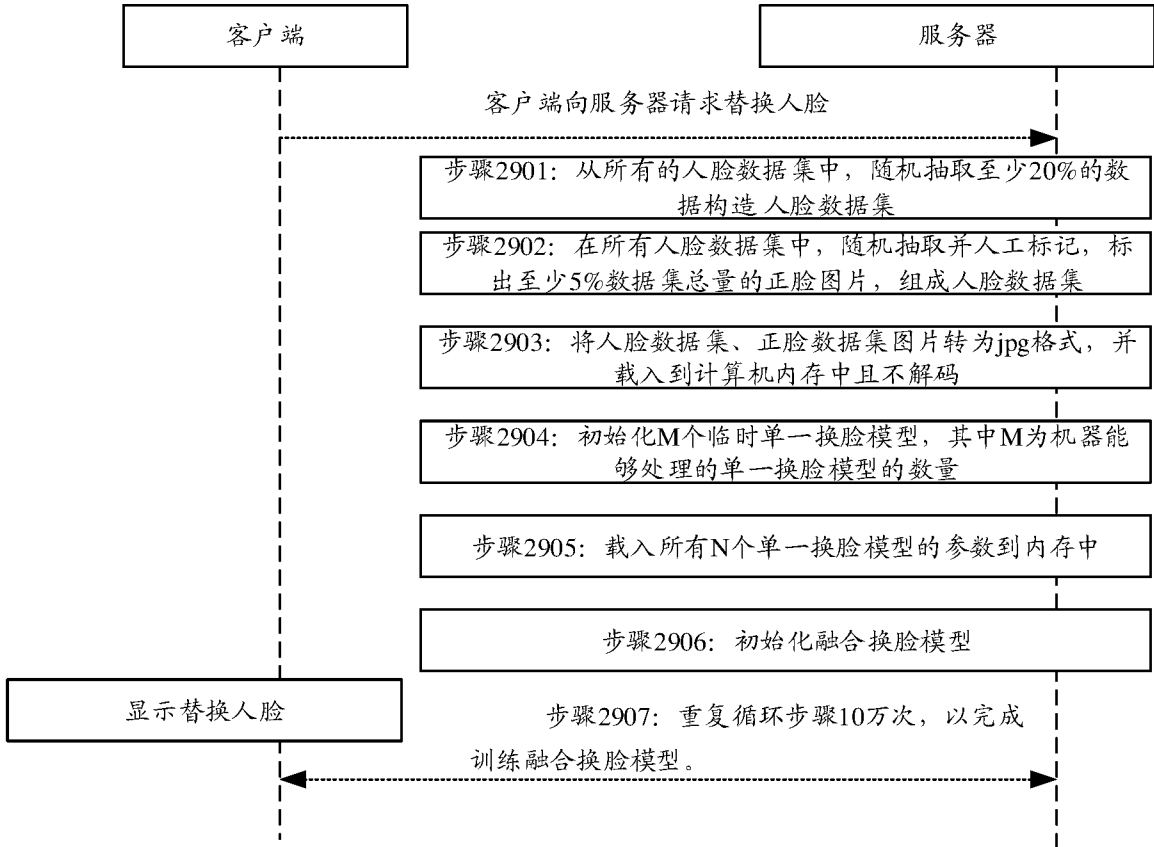


图 29A

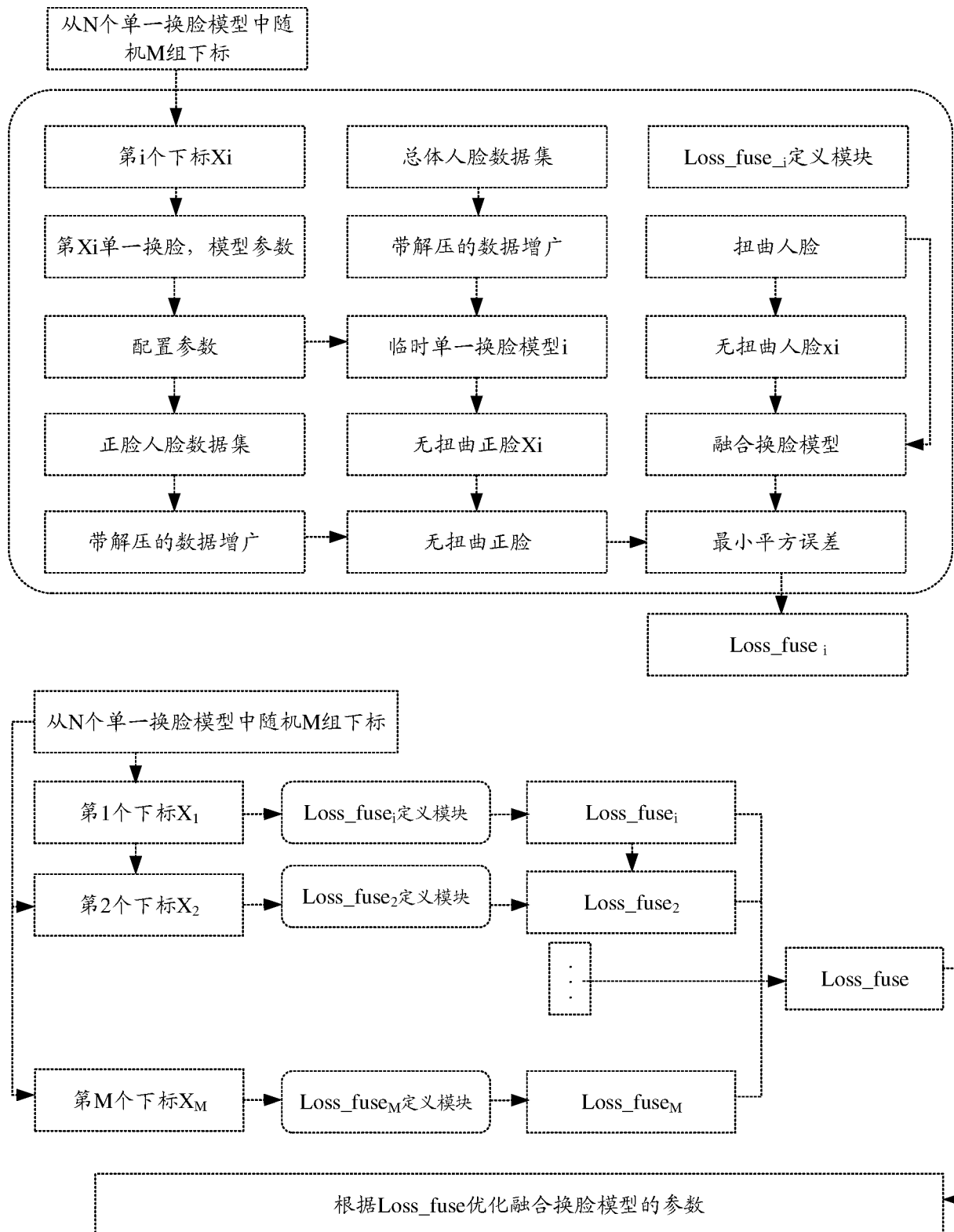


图 29B

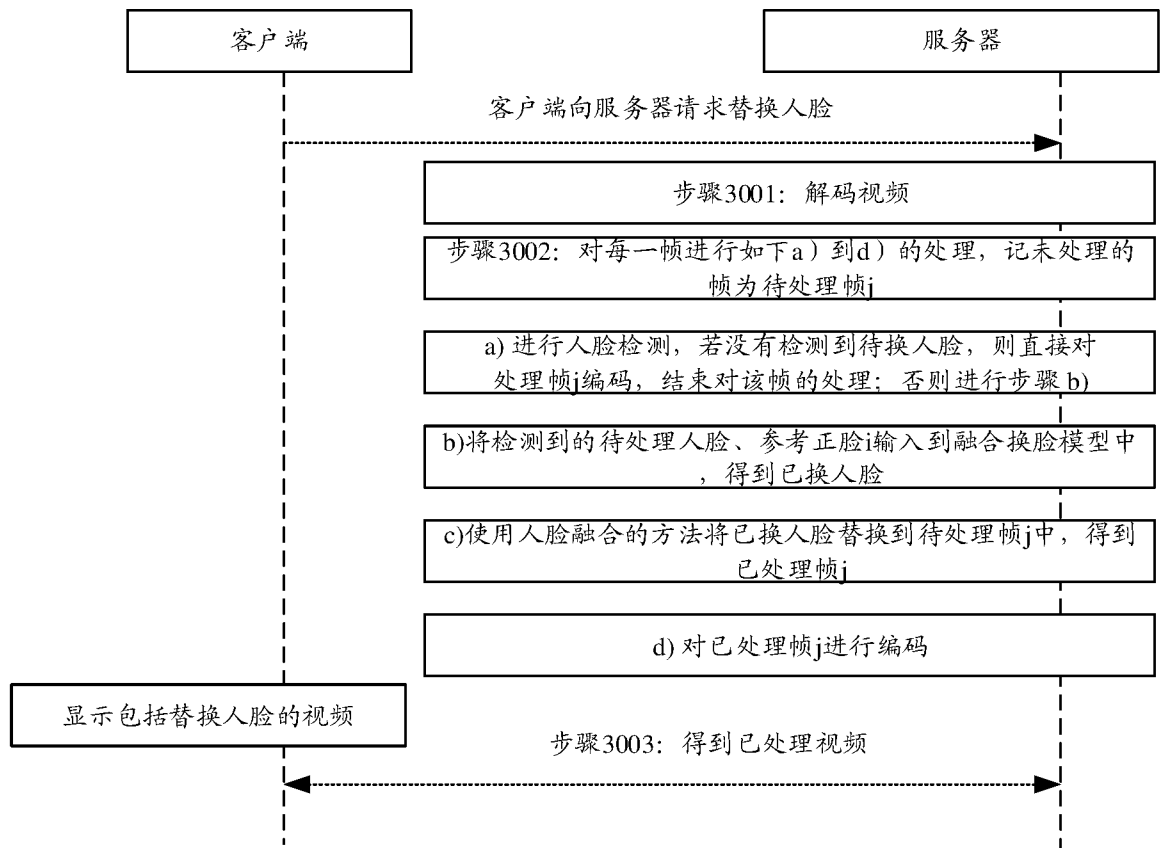


图 30A

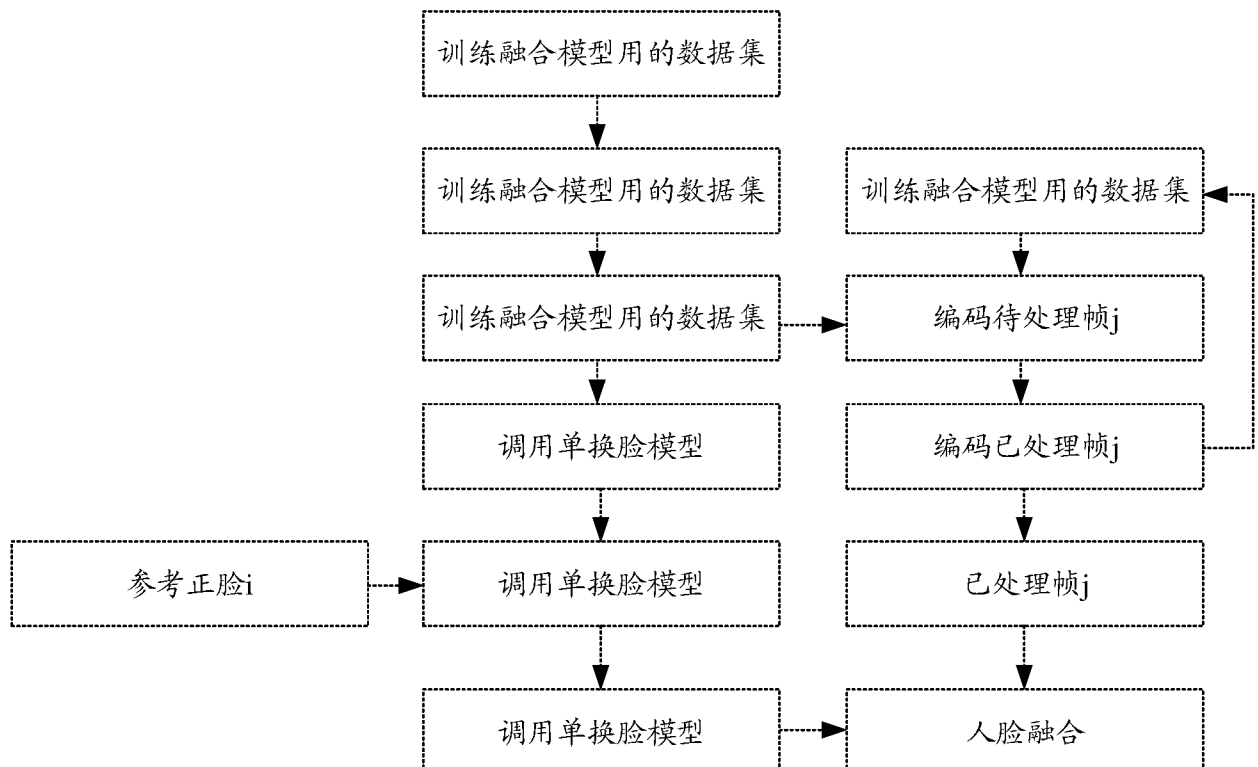


图 30B

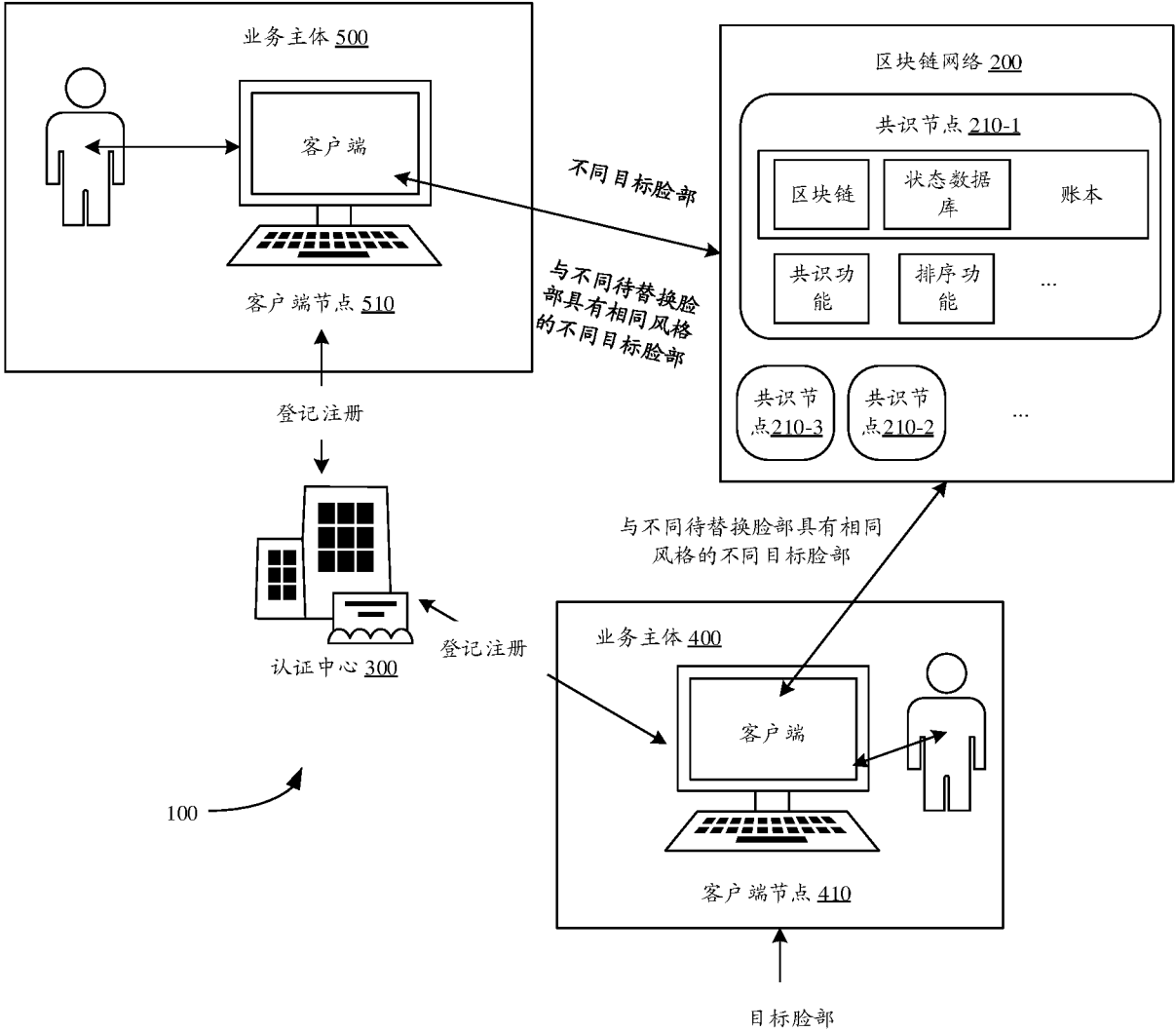


图 31

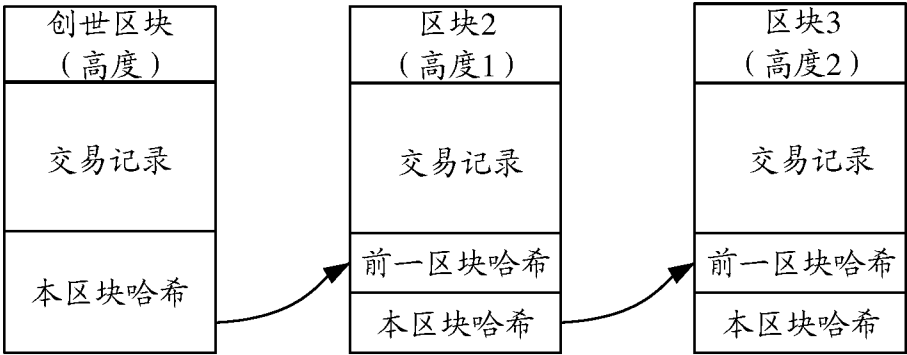


图 32

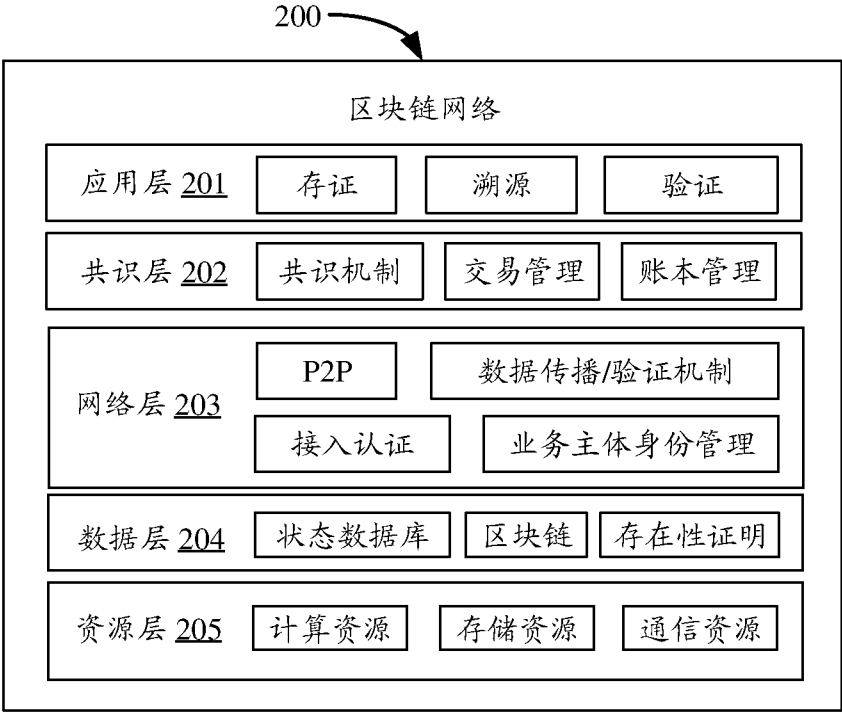


图 33

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2020/113305

A. CLASSIFICATION OF SUBJECT MATTER

G06K 9/62(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)

CNPAT, CNKI, WPI, EPODOC, IEEE: 编码器, 解码器, 融合, 自编码网络, 自动编码器, 图像, 风格, 人脸, 脸部, 特征, 替换, 状态, 五官, encoder, decoder, face, fusion, auto, image, style, feature, state, change

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 110826593 A (TENCENT TECHNOLOGY SHENZHEN CO., LTD.) 21 February 2020 (2020-02-21) description, paragraphs [0006]-[0131]	1-19
A	CN 110148081 A (TENCENT TECHNOLOGY SHENZHEN CO., LTD.) 20 August 2019 (2019-08-20) description, paragraphs [0006]-[0042]	1-19
A	CN 108171223 A (BEIJING XINDATEK CO., LTD. et al.) 15 June 2018 (2018-06-15) entire document	1-19
A	CN 108520503 A (XIANGTAN UNIVERSITY) 11 September 2018 (2018-09-11) entire document	1-19
A	CN 109886072 A (INSTITUTE OF AUTOMATION, CHINESE ACADEMY OF SCIENCES) 14 June 2019 (2019-06-14) entire document	1-19
A	US 2018176578 A1 (WAVEONE INC.) 21 June 2018 (2018-06-21) entire document	1-19



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

13 November 2020

Date of mailing of the international search report

07 December 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/CN2020/113305

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	110826593	A	21 February 2020	None			
CN	110148081	A	20 August 2019	WO	2020192704	A1	01 October 2020
CN	108171223	A	15 June 2018	None			
CN	108520503	A	11 September 2018	None			
CN	109886072	A	14 June 2019	None			
US	2018176578	A1	21 June 2018	US	2018174047	A1	21 June 2018
				US	2018173994	A1	21 June 2018
				US	2018174275	A1	21 June 2018
				US	2018176570	A1	21 June 2018
				US	2019266490	A1	29 August 2019
				US	2018176576	A1	21 June 2018
				US	2018174052	A1	21 June 2018

国际检索报告

国际申请号

PCT/CN2020/113305

A. 主题的分类

G06K 9/62 (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))

CNPAT, CNKI, WPI, EPDOC, IEEE: 编码器, 解码器, 融合, 自编码网络, 自动编码器, 图像, 风格, 人脸, 脸部, 特征, 替换, 状态, 五官, encoder, decoder, face, fusion, auto, image, style, feature, state, change

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 110826593 A (腾讯科技深圳有限公司) 2020年 2月 21日 (2020 - 02 - 21) 说明书第[0006]-[0131]段	1-19
A	CN 110148081 A (腾讯科技深圳有限公司) 2019年 8月 20日 (2019 - 08 - 20) 说明书第[0006]-[0042]段	1-19
A	CN 108171223 A (北京中晟信达科技有限公司 等) 2018年 6月 15日 (2018 - 06 - 15) 全文	1-19
A	CN 108520503 A (湘潭大学) 2018年 9月 11日 (2018 - 09 - 11) 全文	1-19
A	CN 109886072 A (中国科学院自动化研究所) 2019年 6月 14日 (2019 - 06 - 14) 全文	1-19
A	US 2018176578 A1 (WAVEONE INC.) 2018年 6月 21日 (2018 - 06 - 21) 全文	1-19

☐ 其余文件在C栏的续页中列出。☒ 见同族专利附件。

* 引用文件的具体类型:

“A” 认为不特别相关的表示了现有技术一般状态的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

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

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

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

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

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

“&” 同族专利的文件

国际检索实际完成的日期

2020年 11月 13日

国际检索报告邮寄日期

2020年 12月 7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

刘剑

电话号码 86-(10)-53961304

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2020/113305

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	110826593	A	2020年 2月 21日	无			
CN	110148081	A	2019年 8月 20日	W0	2020192704	A1	2020年 10月 1日
CN	108171223	A	2018年 6月 15日	无			
CN	108520503	A	2018年 9月 11日	无			
CN	109886072	A	2019年 6月 14日	无			
US	2018176578	A1	2018年 6月 21日	US	2018174047	A1	2018年 6月 21日
				US	2018173994	A1	2018年 6月 21日
				US	2018174275	A1	2018年 6月 21日
				US	2018176570	A1	2018年 6月 21日
				US	2019266490	A1	2019年 8月 29日
				US	2018176576	A1	2018年 6月 21日
				US	2018174052	A1	2018年 6月 21日