

(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
2023 年 4 月 6 日 (06.04.2023)



(10) 国际公布号
WO 2023/050992 A1

(51) 国际专利分类号:

G06T 13/40 (2011.01) *G06N 3/08* (2006.01)
G06T 7/11 (2017.01) *G06K 9/62* (2022.01)
G06T 7/194 (2017.01) *G06V 10/774* (2022.01)
G06N 3/04 (2006.01)

(21) 国际申请号: PCT/CN2022/108304

(22) 国际申请日: 2022 年 7 月 27 日 (27.07.2022)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:

202111161955.1 2021 年 9 月 30 日 (30.09.2021) CN
202111389685.X 2021 年 11 月 22 日 (22.11.2021) CN

(71) 申请人: 广州视源电子科技有限公司 (GUANGZHOU SHIYUAN ELECTRONICS CO., LTD.) [CN/CN]; 中国广东省广州市黄

埔区云埔四路 6 号, Guangdong 510530 (CN)。广州视源人工智能创新研究院有限公司 (GUANGZHOU SHIYUAN ARTIFICIAL INTELLIGENCE RESEARCH INSTITUTE LIMITED COMPANY) [CN/CN]; 中国广东省广州市黄埔区云埔四路 6 号 1 栋 1001 房, Guangdong 510530 (CN)。

(72) 发明人: 张玉兵 (ZHANG, Yubing); 中国广东省广州市黄埔区云埔四路 6 号, Guangdong 510530 (CN)。

(74) 代理人: 北京泽方誉航专利代理事务所 (普通合伙) (BEIJING ZFANG PATENT AGENCY); 中国北京市丰台区南四环西路 186 号四区 2 号楼 7 层 23 室陈照辉, Beijing 100071 (CN)。

(81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU,

(54) Title: NETWORK TRAINING METHOD AND APPARATUS FOR FACIAL RECONSTRUCTION, AND DEVICE AND STORAGE MEDIUM

(54) 发明名称: 用于人脸重建的网络训练方法、装置、设备及存储介质

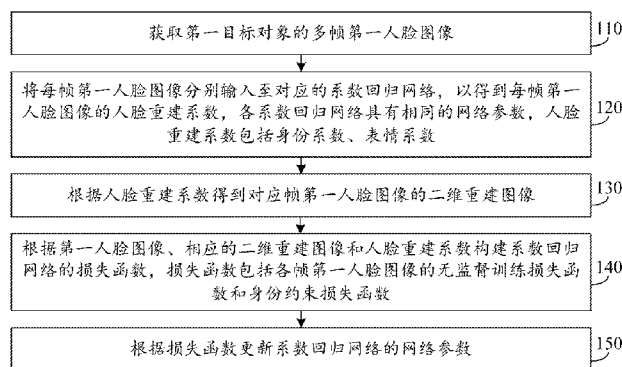


图 1

110 Acquire a plurality of frames of first facial images of a first target object
120 Respectively input each frame of first facial image into a corresponding coefficient regression network, so as to obtain a facial reconstruction coefficient of each frame of first facial image, wherein each coefficient regression network has the same network parameter, and the facial reconstruction coefficients comprise: an identity coefficient and an expression coefficient
130 Obtain a two-dimensional reconstructed image of the corresponding frame of first facial image according to the facial reconstruction coefficient
140 Construct loss functions of the coefficient regression network according to the first facial image, the corresponding two-dimensional reconstructed image and the facial reconstruction coefficient, wherein the loss functions comprise an unsupervised training loss function and an identity constraint loss function of each frame of first facial image
150 Update the network parameter of the coefficient regression network according to the loss functions

(57) Abstract: Disclosed in the embodiments of the present application are a network training method and apparatus for facial reconstruction, and a device and a storage medium. The method comprises: acquiring a plurality of frames of first facial images of a first target object; inputting each frame of first facial image into a corresponding coefficient regression network, so as to obtain a facial reconstruction coefficient of each frame of first facial image, wherein each coefficient regression network has the same network parameter, and the facial reconstruction coefficients comprise: an identity coefficient, an expression coefficient, a texture coefficient, a pose coefficient and an illumination coefficient; obtaining a two-dimensional reconstructed image of the corresponding frame of first facial image according to the facial reconstruction coefficient; constructing loss functions of the coefficient regression network according to the first facial image, the corresponding two-dimensional reconstructed image and the facial reconstruction coefficient, wherein the loss functions comprise an unsupervised training loss function and an identity constraint loss function of each frame of first facial image; and updating the network parameter of the coefficient regression network according to the loss functions. By using the method, the technical problem of low expression accuracy during the reconstruction of a three-dimensional facial image can be solved.

(57) 摘要: 本申请实施例公开了一种用于人脸重建的网络训练方法、装置、设备及存储介质, 其包括: 获

CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW。

- (84) 指定国 (除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

根据细则4.17的声明:

- 关于不损害新颖性的公开或丧失新颖性的例外
(细则4.17(v))

本国际公布:

- 包括国际检索报告(条约第21条(3))。

取第一目标对象的多帧第一人脸图像; 将每帧第一人脸图像输入至对应的系数回归网络, 得到每帧第一人脸图像的人脸重建系数, 各系数回归网络具有相同的网络参数, 人脸重建系数包括: 身份系数、表情系数、纹理系数、位姿系数和光照系数; 根据人脸重建系数得到对应帧第一人脸图像的二维重建图像; 根据第一人脸图像、相应的二维重建图像和人脸重建系数构建系数回归网络的损失函数, 损失函数包括各帧第一人脸图像的无监督训练损失函数和身份约束损失函数; 根据损失函数更新系数回归网络的网络参数。采用上述方法可以解决重建三维人脸图像时表情准确性低的技术问题。

用于人脸重建的网络训练方法、装置、设备及存储介质

本申请要求在 2021 年 09 月 30 日提交中国专利局，申请号为 202111161955.1 的中国专利申请的优先权，该申请的全部内容通过引用结合在本申请中，以及要求在 2021 年 11 月 22 日提交中国专利局，申请号为 202111389685.X 的中国专利申请的优先权，该申请的全部内容通过引用结合在本申请中。

技术领域

本申请实施例涉及图像处理技术领域，尤其涉及一种用于人脸重建的网络训练方法、装置、设备及存储介质。

背景技术

随着网络通信技术（尤其是 5G 通信技术）的发展，用户足不出户便可利用通信设备（如手机、电脑等）享受视频通话、网上课堂、网上会议等网络通信资源。目前，视频交流是网络通信技术中一种常见的场景，视频交流时通话双方可以看到对方当前的人脸图像。然而，随着视频交流的广泛应用，用户的安全意识也在提高，有些场景下，用户不希望将自己的真实人像展示给其他人看，此时，可以使用虚拟的卡通图像替代真实人像。

使用卡通图像时，可以将用户当前真实人像的表情、头部动作姿势迁移到卡通图像上，以使卡通图像与真实人像具有相同的表情、头部动作姿势。一些相关技术中，迁移时，先利用真实人像重建出三维人脸图像，以得到表情系数和位姿系数，再将表情系数和位姿系数迁移到对应的卡通图像。人脸表情的准确拟合是迁移时比较重要的环节，因此，在重建三维人脸图像时，需要注意三维人脸图像中表情准确性的问题。

发明内容

本申请实施例提供了一种用于人脸重建的网络训练方法、装置、设备及存储介质，以解决相关技术中重建三维人脸图像时表情重建准确性低的技术问题。

第一方面，本申请一个实施例提供了一种用于人脸重建的网络训练方法，包括：

获取第一目标对象的多帧第一人脸图像；

将每帧所述第一人脸图像分别输入至对应的系数回归网络，以得到每帧所述第一人脸图像的人脸重建系数，各所述系数回归网络具有相同的网络参数，所述人脸重建系数包括身份系数、表情系数、纹理系数、位姿系数和光照系数；

根据所述人脸重建系数得到对应帧所述第一人脸图像的二维重建图像；

根据所述第一人脸图像、相应的二维重建图像和所述人脸重建系数构建所述系数回归网络的损失函数，所述损失函数包括各帧第一人脸图像的无监督训练损失函数和身份约束损失函数，所述无监督训练损失函数包括：图像重构损失函数、图像感知损失函数、关键点损失函数和正则损失函数；

根据所述损失函数更新所述系数回归网络的网络参数，以训练所述系统回归网络。

第二方面，本申请一个实施例提供了一种用于人脸重建的网络训练方法，包括：

获取第一训练集，所述第一训练集包括多个人脸样本图像；

通过第一参数估计网络模型得到所述人脸样本图像对应的第一训练 3DMM 参数，所述第一训练 3DMM 参数包括身份系数、表情系数、纹理系数、位姿系数和光照系数；

通过预设第一损失函数训练所述第一参数估计网络模型,所述预设第一损失函数包括图像重构损失函数、图像感知损失函数、关键点重构损失函数和正则损失函数,所述预设第一损失函数通过所述人脸样本图像、所述人脸样本图像对应的渲染图像和所述第一训练 3DMM 参数得到,所述渲染图像是将所述第一训练 3DMM 参数引入预设的 3DMM 模型后渲染到图像平面而得到;

获取第二训练集,所述第二训练集包括多个人脸样本图像和每个人脸样本图像对应的粗糙样本 3D 表情模板,所述粗糙样本 3D 表情模板通过训练好的所述第一参数估计网络模型输出的所述人脸样本图像的第一训练 3DMM 参数构建;

通过第二参数估计网络模型得到所述人脸样本图像的第二训练 3DMM 参数,所述第二训练 3DMM 参数包括表情系数、纹理系数、位姿系数和光照系数;

通过预设第二损失函数训练所述第二参数估计网络模型,所述预设第二损失函数包括图像重构损失函数、图像感知损失函数、关键点重构损失函数、正则损失函数和表情梯度损失函数,所述预设第二损失函数通过所述人脸样本图像、所述人脸样本图像对应的渲染图像、所述人脸样本图像对应的粗糙样本 3D 表情模板和所述粗糙样本 3D 表情模板对应的精细样本 3D 表情模板得到,所述渲染图像是将基于所述第二训练 3DMM 参数得到的 3D 人脸渲染到图像平面后得到,所述精细样本 3D 表情模板通过所述粗糙样本 3D 表情模板得到。

第三方面,本申请一个实施例还提供了一种用于人脸重建的网络训练装置,包括:

图像获取模块,用于获取第一目标对象的多帧第一人脸图像;

图像输入模块,用于将每帧所述第一人脸图像分别输入至对应的系数回归网络,以得到每帧所述第一人脸图像的人脸重建系数,各所述系数回归网络具有相同的网络参数,所述人脸重建系数包括身份系数、表情系数、纹理系数、位姿系数和光照系数;

图像重建模块,用于根据所述人脸重建系数得到对应帧所述第一人脸图像的二维重建图像;

损失构建模块,用于根据所述第一人脸图像、相应的二维重建图像和所述人脸重建系数构建所述系数回归网络的损失函数,所述损失函数包括各帧所述第一人脸图像的无监督训练损失函数和身份约束损失函数,所述无监督训练损失函数包括:图像重构损失函数、图像感知损失函数、关键点损失函数和正则损失函数;

参数更新模块,用于根据所述损失函数更新所述系数回归网络的网络参数,以训练所述系统回归网络。

第四方面,本申请一个实施例还提供了一种用于人脸重建的网络训练装置,包括:

第一训练模块,用于获取第一训练集,所述第一训练集包括多个人脸样本图像;通过第一参数估计网络模型得到所述人脸样本图像对应的第一训练 3DMM 参数,所述第一训练 3DMM 参数包括身份系数、表情系数、纹理系数、位姿系数和光照系数;通过预设第一损失函数训练第一参数估计网络模型,所述预设第一损失函数包括图像重构损失函数、图像感知损失函数、关键点重构损失函数和正则损失函数,所述预设第一损失函数通过所述人脸样本图像、所述人脸样本图像对应的渲染图像和所述第一训练 3DMM 参数得到,所述渲染图像是将所述第一训练 3DMM 参数引入预设的 3DMM 模型后渲染到图像平面而得到;

第二训练模块,用于获取第二训练集,所述第二训练集包括多个人脸样本图像和每个人脸样本图像对应的粗糙样本 3D 表情模板,所述粗糙样本 3D 表情模板通过训练好的所述第一参数估计网络模型输出的所述人脸样本图像的第一训练 3DMM 参数构建;通过第二参数估计网络模型得到所述人脸样本图像的第二训练 3DMM 参数,所述第二训练 3DMM 参数包括表情系数、纹理系数、位姿系数和光照系数;通过预设第二损失函数训练所述第二参数估计网络模型,所述预设第二损失函数包括图像重构损失函数、图像感

知损失函数、关键点重构损失函数、正则损失函数和表情梯度损失函数，所述预设第二损失函数通过所述人脸样本图像、所述人脸样本图像对应的渲染图像、所述人脸样本图像对应的粗糙样本 3D 表情模板和所述粗糙样本 3D 表情模板对应的精细样本 3D 表情模板得到，所述渲染图像是将基于所述第二训练 3DMM 参数得到的 3D 人脸渲染到图像平面后得到，所述精细样本 3D 表情模板通过所述粗糙样本 3D 表情模板得到。

第五方面，本申请一个实施例还提供了一种用于人脸重建的网络训练设备，包括：

一个或多个处理器；

存储器，用于存储一个或多个程序；

当所述一个或多个程序被所述一个或多个处理器执行，使得所述一个或多个处理器实现如第一方面所述的用于人脸重建的网络训练方法或如第二方面所述的用于人脸重建的网络训练方法。

第六方面，本申请实施例还提供了一种计算机可读存储介质，其上存储有计算机程序，该程序被处理器执行时实现如第一方面所述的用于人脸重建的网络训练方法或如第二方面所述的用于人脸重建的网络训练方法。

在本申请一个实施例中，通过获取多帧包含同一目标对象的第一人脸图像，并利用具有相同网络参数的各系数回归网络处理各帧第一人脸图像，以得到人脸重建系数，之后，根据人脸重建系数重建出第一人脸图像对应的二维重建图像，根据第一人脸图像、二维重建图像和人脸重建系数构建无监督训练损失函数和身份约束损失函数，进而基于无监督训练损失函数和身份约束损失函数更新系数回归网络的网络参数的技术手段，解决了重建三维人脸图像时表情重建准确性低的技术问题。通过系数回归网络分别输出身份系数和表情系数，可以实现身份系数和表情系数的分别确定，并且，基于身份约束损失函数可以保证身份系数的准确性，即促使系数回归网络基于同一个人的不同图像可以得到相同的人脸身份，该人脸身份无关表情，进而保证了表情系数的准确性，即保证表情准确性。同时，除了对身份系数的准确性进行了约束，还结合第一人脸图像、二维重建图像和人脸重建系数构建了无监督训练损失函数，提高了损失函数的合理性，进而在基于损失函数训练系数回归网络时，提高了系数回归网络的性能，进一步保证了表情系数的准确性。

附图说明

图 1 为本申请一个实施例提供的一种用于人脸重建的网络训练方法的流程图；

图 2 为本申请一个实施例提供的一种用于人脸重建的网络训练方法的流程图；

图 3 为本申请一个实施例提供的一种用于人脸重建的网络训练方法的流程图；

图 4 为本申请一个实施例提供的一种互斥表情对示意图；

图 5 为本申请一个实施例提供的一种用于人脸重建的网络训练方法的框架示意图；

图 6 为本申请一个实施例提供的一种系数回归网络的应用流程示意图；

图 7 为本申请一个实施例中提供的虚拟图像生成框架图；

图 8 为本申请一个实施例提供的一种生成 3D 表情模板的流程图；

图 9 为本申请一个实施例提供的一种生成 3D 表情模板的流程图；

图 10 为本申请一个实施例提供的一种粗糙表情模板的生成过程示意图；

图 11 为本申请一个实施例提供的一种精细表情模板的生成过程示意图；

图 12 为本申请一个实施例提供的一种用于人脸重建的网络训练装置的结构示意图；

图 13 为本申请实施例提供的一种用于人脸重建的网络训练设备的结构示意图。

具体实施方式

下面结合附图和实施例对本申请作进一步的详细说明。可以理解的是，此处所描述的具体实施例用于解释本申请，而非对本申请的限定。另外还需要说明的是，为了便于描述，附图中仅示出了与本申请相关的部分而非全部结构。

基于真实人脸重建三维人脸图像时，可以采用传统方法和基于深度学习的方法。其中，传统方法是：将真实人脸中的人脸关键点和重建的三维人脸图像中的人脸关键点投影到相同的二维平面，并根据平面中同一语义的人脸关键点（如真实人脸中表示鼻尖的人脸关键点和三维人脸图像中表示鼻尖的人脸关键点具有相同的语义）间的位置误差进行模型拟合，即调整三维人脸图像中的关键点，以使得三维人脸图像仅可能贴近真实人脸。基于深度学习的方法是：根据大量的二维真实人脸图像训练出三维人脸重建模型，通过该三维人脸重建模型重建出三维人脸图像。无论上述哪种方式重建三维人脸图像，都无法保证重建时表情系数的准确性。

基于此，本申请实施例中提供一种用于人脸重建的网络训练方法、装置、设备及存储介质，以在重建三维人脸图像时，提高表情系数的准确性。

其中，本申请一个实施例提供的用于人脸重建的网络训练方法可以由用于人脸重建的网络训练设备执行，该用于人脸重建的网络训练设备可以通过软件和/或硬件的方式实现，该用于人脸重建的网络训练设备可以是两个或多个物理实体构成，也可以是一个物理实体构成。例如，用于人脸重建的网络训练设备可以是电脑、手机、平板电脑或交互智能平板等智能设备，还可以是服务器。

实施例一

图 1 为本申请一个实施例提供的一种用于人脸重建的网络训练方法的流程图。参考图 1，该用于人脸重建的网络训练方法具体包括：

步骤 110、获取第一目标对象的多帧第一人脸图像。

目标对象是指需要重建自身人脸图像的对象，当前，将神经网络训练过程中使用的目标对象记为第一目标对象。神经网络用于基于第一目标对象的人脸图像得到重建人脸时需要一些系数。可理解，一次训练过程使用的第一目标对象可以为一个或多个，当前，以使用一个第一目标对象为例。并且，每次训练时可以使用不同的第一目标对象（即具有不同人脸的对象）。将第一目标对象的人脸图像记为第一人脸图像。一次训练过程中，使用第一目标对象的多帧第一人脸图像，该帧数可以根据实际情况设置。其中，第一人脸图像的获取方式当前不作限定，例如，在包含第一目标对象的视频数据中，随机截取多帧包含第一目标对象的人脸图像作为第一人脸图像。又如，从截取的包含第一目标对象的人脸图像中，提取人脸区域，剔除背景区域，以得到多帧第一人脸图像。再如，在同一背景中，对第一目标对象进行拍摄，以得到多帧第一人脸图像。还如，从现有公开的一个或多个数据集中，提取多帧同一第一目标对象的第一人脸图像，其中，使用的数据集可以是 Voxceleb2、300W-LP 和/或 FEAFa 等。Voxceleb2 为具有大规模人脸视频数据的数据集，300W-LP 为具有丰富人脸姿势信息的数据集，FEAFa 为具有丰富人脸表情信息的数据集。一个实施例中，第一人脸图像为单目采集的图像，单目采集是指通过一个摄像头采集。一个实施例中，第一人脸图像为 RGB 图像。

示例性的，各帧第一人脸图像均为已知人脸身份的人脸图像。其中，人脸身份是对于人脸形态的具体描述，如人脸身份描述了人脸的关键部位，人脸的关键部位包括眉毛、眼睛、鼻子、嘴巴以及脸颊等。举例而言，人脸身份所描述的关键部位为大眼睛、高鼻梁、薄嘴唇。此时，由于不同目标对象的人脸不同，那么，不同目标对象的人脸身份所描述的关键部位间会存在差异。可理解，将不同的表情叠加到人脸身份上可以得到各种表情的人脸。第一人脸图像也可以是未知人脸身份的人脸图像。

步骤 120、将每帧第一人脸图像分别输入至对应的系数回归网络，以得到每帧第一人脸图像的人脸重建系数，各系数回归网络具有相同的网络参数，人脸重建系数包括身份系数、表情系数。

系数回归网络为得到重建人脸所需系数的神经网络。系数回归网络的具体结构可以根据实际情况设置。例如，采用 MobileNet-V2 轻量级神经网络或其他轻量级神经网络作为系数回归网络的网络骨架，以加快网络推理的速度，满足应用场景下实时性的需求。训练过程中，使用的系数回归网络的数量与多帧第一人脸图像的帧数相等，以保证每个第一人脸图像均有对应的系数回归网络进行处理。一个实施例中，多个系数回归网络共享网络参数，即各系数回归网络具有相同的网络参数，其中，网络参数是指系数回归网络中使用的卷积核权重等参数。可理解，系数回归网络构建时，采用默认（如人为设定）的网络参数，且随着训练次数的增加，网络参数不断更新，以提高系数回归网络的性能，使得系数回归网络输出更准确的结果。

示例性的，将一帧第一人脸图像输入到一个系数回归网络后，系数回归网络对第一人脸图像进行处理，以输出重建该第一人脸图像中人脸所需的系数。一个实施例中，将系数回归网络输出的系数记为人脸重建系数，人脸重建系数包含的系数类型可以根据实际情况选择，一个实施例中，人脸重建系数包括：身份系数和表情系数。

身份系数用于描述当前输入的第一人脸图像的人脸身份，可选的，预先设置可出现在人脸上的各个身份基，每个身份基具有一个关键部位的语义，每个身份基可认为是该关键部位会出现的一个类别，例如，关键部位为嘴巴时，其对应的各身份基包括上厚嘴唇、上薄嘴唇、上小（短）嘴唇、上大（长）嘴唇、下厚嘴唇、下薄嘴唇、下小（短）嘴唇、上大（长）嘴唇等。身份系数可以体现第一人脸图像出现各身份基的权重（也可以理解为置信度），身份系数可以为矩阵向量的形式，身份系数中的每个数值对应一个身份基，数值越大，表明人脸中出现该身份基的可能性越高。在重建过程中，根据身份系数，将各身份基叠加在预先构建的人脸模型上，便可以重建出第一人脸图像的人脸身份。

表情系数用于描述当前输入的第一人脸图像的人脸表情，可选的，预先设置可出现在人脸上的各个表情基，每个表情基对应一个表情。可理解，由于人脸的表情是很丰富的，如果想要表示出人脸的全部表情，则需要构建大量的表情基，这会大大增加数据处理量。因此，一个实施例中，仅构建基本表情的表情基，其中，基本表情的具体内容可以根据实际情况设定，通过将各基本表情组合叠加到人脸身份上可以得到人脸的各种表情。例如，针对眼睛的基本表情包括：左眼闭、左眼瞪大、右眼闭和右眼瞪大，此时，根据上述四个基本表情和人脸身份中的眼睛便可以得到眼睛的各种表情，例如，双眼微眯的表情可以通过左眼闭、右眼闭和人脸身份中的眼睛线性叠加后获得。表情系数可以体现第一人脸图像出现各表情基的权重（也可以理解为置信度），表情系数可以为矩阵向量的形式，表情系数中的每个数值对应一个表情基，数值越大，表明人脸中出现该表情基的可能性越高。以眼部为例，在重建过程中，根据得到的眼部各表情基的系数，将对应的表情基应用在人脸身份的眼部（系数越大，表情基占比越大）后，可以重建出第一人脸图像中带表情的眼睛。

可理解,上述人脸重建系数仅为示例性描述,实际应用中,可以结合实际需求增加或减少人脸重建系数的类别。

需要说明的是,当前每个系数回归网络均输出对应的人脸重建系数,即每帧第一人脸图像均存在对应的人脸重建系数。

可选的,人脸重建系数还可以包括纹理系数(与面部纹理有关)、位姿系数(与头部动作有关)和光照系数(与环境光照有关)。

步骤 130、根据人脸重建系数得到对应帧第一人脸图像的二维重建图像。

示例性的,根据人脸重建系数重建第一人脸图像中人脸的三维人脸图像,并将三维人脸图像映射到二维平面中,便可以得到第一人脸图像重建后的二维人脸图像。一个实施例中,将当前得到的二维人脸图像记为二维重建图像,可理解,二维重建图像可以认为是基于人脸重建系数重建的人脸图像,系数回归网络输出的人脸重建系数越准确,二维重建图像与对应的第一人脸图像越相似。

一个实施例中,人脸重建系数包括身份系数和表情系数时,通过人脸重建系数得到的三维重建图像可以描述第一目标对象的人脸身份以及在对应第一人脸图像中出现的表情。

示例性的,预先构建平均的三维人脸形状、身份基底以及表情基底,其中,三维人脸形状可以理解为预先构建的人脸模型,其可以通过对大量的三维人脸图像进行平均得到,身份基底可以认为是预先构建的身份基的集合,表情基底可以认为是预先构建的表情基的集合,之后,将身份系数应用在身份基底上,将表情系数应用在表情基底上,再与平均的三维人脸形状进行叠加,便可以得到三维重建图像。

之后,将三维重建图像映射到二维平面上,便可以得到二维重建图像。一个实施例中,使用软件渲染器进行图像的映射。其中,软件渲染器的具体类型当前不作限定。

步骤 140、根据第一人脸图像、相应的二维重建图像和人脸重建系数构建系数回归网络的损失函数,损失函数包括各帧第一人脸图像的无监督训练损失函数和身份约束损失函数。

示例性的,系数回归网络的性能越高,基于系数回归网络输出的人脸重建系数得到的二维重建图像与输入系数回归网络的第一人脸图像越相似,因此,可比较第一人脸图像和对应的二维重建图像间的差异程度,进而根据差异程度调整系数回归网络的网络参数,以提高系数回归网络的性能。在此基础上,为了进一步提高系数回归网络的性能,还结合人脸重建系数,利用有监督训练方式(即使用先验知识)调整系数回归网络的网络参数。例如,基于第一人脸图像的已知身份系数和系数回归网络输出的身份系数,可以明确系数回归网络输出的身份系数是否准确,其中,已知身份系数和身份系数间的差异越小,身份系数越准确,之后,可以根据差异程度来调整系数回归网络的网络参数,以提高身份系数的准确度,进而提高系数回归网络的性能。

一个实施例中,在调整系数回归网络的网络参数前,先基于第一人脸图像和相应的二维重建图像以及人脸重建系数,构建系数回归网络的损失函数,并通过损失函数来调整网络参数。其中,损失函数用来评价模型(这里指系数回归网络)的预测值和真实值不一样的程度,损失函数越好,模型的性能越好。一般而言,预先构建损失函数的表达式,该表达式中的变量与人脸重建系数有关,该变量可以是人脸重建系数中的相关数据或者是基于人脸重建系数得到的二维重建图像的相关数据,将每次训练过程中得到的变量代入表达式中,便可以得到本次训练过程中的损失函数(这里指损失函数的具体值),进而基于损失函数调整网络参数。可理解,随着训练次数的增加,损失函数逐渐收敛。

损失函数的类型可以根据实际情况设置,当前损失函数包括各帧第一人脸图像的无监督训练损失函数

和身份约束损失函数。

示例性的，无监督训练也可理解为无监督学习，其可以根据没有类别（即没有被标记）的样本解决神经网络训练过程中的问题，即学习过程中，没有用到先验知识。无监督训练损失函数可理解为无监督训练过程中使用的损失函数，其可以基于第一人脸图像和二维重建图像构建，也可以基于人脸重建系数构建，且在构建过程中，没有使用任何的先验知识。一个实施例中，无监督训练损失函数包括：第一人脸图像的图像重构损失函数、图像感知损失函数、关键点损失函数和正则损失函数中的至少一种。其中，图像重构损失函数可以体现第一人脸图像和对应的二维重建图像间的像素差异，一个实施例中，利用 L2 范数计算第一人脸图像和二维重建图像中具有相同像素索引（相同像素索引在两个图像中对应的像素的位置相同）的像素间的差值（这里是指像素值的差），之后，利用各像素索引对应的差值得到图像重构损失函数，可理解，像素差异越小，图像重构损失函数越小。图像感知损失函数可以体现第一人脸图像和对应的二维重建图像间的图像特征差异。一个实施例中，使用预先训练好的人脸识别网络（该网络用于提取图像中与人脸有关的特征）分别提取第一人脸图像的图像特征和二维重建图像的图像特征，之后，计算两个图像特征的相似度，并根据相似度得到图像感知损失函数，可理解，相似度越高，图像感知损失函数越小。关键点损失函数可以体现第一人脸图像和对应的二维重建图像间人脸关键点的位置差值。一个实施例中，提取第一人脸图像和二维重建图像中的各人脸关键点，其中，人脸关键点是指位于关键部位中的点，每个人脸关键点都为人脸的语义信息。人脸关键点的选择可以根据实际情况设定，例如，当前设置一个人脸上共有 68 个人脸关键点，其中，左眉和右眉分别包含 5 个关键点，左眼和右眼分别包含 6 个关键点，嘴部包含 20 个关键点，脸颊包含 17 个关键点，鼻子包含 9 个关键点。对于第一人脸图像和对应的二维重建图像，均有 68 个人脸关键点。之后，计算第一人脸图像和二维重建图像中具有相同语义信息的人脸关键点间的位置差值，并基于位置差值构建关键点损失函数，可理解，位置差值越小，关键点损失函数越小。正则损失函数用于约束系数回归网络，以提高系数回归网络的性能。例如，为了防止三维人脸图像中人脸形状的退化，可以使用 L2 范数对身份系数进行约束。再如，为了防止表情系数的稀疏化，可以使用 L1 范数对表情系数进行约束。得到上述损失函数后，将上述损失函数相加，便可得到无监督训练损失函数，在相加时，可以结合实际需求为每个损失函数设置对应的超参数，以保证无监督训练损失函数更加合理。每帧第一人脸图像均有对应的无监督训练损失函数。

除了构建无监督训练损失函数，还需要构建身份约束损失函数。为了使得系数回归网络输出准确的表情系数，需要保证系数回归网络输出准确的身份系数，即重建的三维人脸图像准确的前提下，只有拟合准确的身份系数（即输出准确的身份系数）才能保证拟合准确的表情系数。示例性的，为了保证身份系数的准确，设计了身份约束损失函数，以通过身份约束损失函数使系数回归网络对同一第一目标对象的不同第一人脸图像预测出一致且准确的身份系数。其中，身份约束损失函数可以理解为对身份系数进行约束的损失函数。身份约束损失函数使用了先验信息，即当前使用的每帧第一人脸图像均存在已知身份系数，之后，根据已知身份系数和系数回归网络输出的身份系数构建身份约束损失函数。一个实施例中，使用 L2 范数计算已知身份系数和身份系数间的误差，进而基于该误差得到身份约束损失函数，或者是，计算各帧第一人脸图像的已知身份系数的平均系数，并使用 L2 范数计算平均系数和身份系数间的误差，进而基于该误差得到身份约束损失函数。可理解，已知身份系数和系数回归网络输出的身份系数差异越小，身份约束损失函数越小，系数回归网络输出的身份系数越准确。再示例性的，身份约束损失函数还可以不使用先验信息，即计算系数回归网络输出的各身份系数的平均值，使用 L2 范数计算平均值和身份系数间的误差，进

而基于该误差得到身份约束损失函数,通过该身份约束损失函数,可以使得各帧第一人脸图像经过对应的系数回归网络后输出的各身份系数尽量贴近各身份系数的平均值,即各身份系数尽量相等,以实现各系数回归网络对不同的第一人脸图像预测出一致且准确的身份系数,避免各帧第一人脸图像因为表情的不同而影响了身份系数的预测。每帧第一人脸图像均有对应的身份约束损失函数。可选的,还可以确定各系数回归网络输出的身份系数的平均系数,并基于平均系数和各身份系数间的误差得到身份约束损失函数。

一个实施例中,各系数回归网络共用相同的损失函数,并基于相同的损失函数更新各系数回归网络的网络参数,以在训练系数回归网络时,使得各系数回归网络的网络参数相同。系数回归网络的损失函数可以由每帧第一人脸图像的无监督训练损失函数以及身份约束损失函数组成,且可以为身份约束损失函数设置超参数,以提高损失函数的精确性。此时,各系数回归网络的损失函数的表达式为:

$$L = \sum_{t=1}^T (L_{\text{loss}}^t + \lambda_{\text{isc}} L_{\text{isc}}^t) \quad (1)$$

其中, L 表示系数回归网络的损失函数, T 为当前获取的第一人脸图像的总帧数, t 表示第 t 帧, L_{loss}^t 表示第 t 帧第一人脸图像对应的无监督训练损失函数, L_{isc}^t 表示第 t 帧第一人脸图像对应的身份约束损失函数, λ_{isc} 表示身份约束损失函数的超参数。

步骤 150、根据损失函数更新系数回归网络的网络参数。

示例性的,根据系数回归网络的损失函数更新各系数回归网络的网络参数,提高系数回归网络的性能,使得系数回归网络输出更准确的人脸重建系数。网络参数更新完成后,确定一次训练过程结束。之后,再次获取新的多帧第一人脸图像(该第一人脸图像的第一目标对象可以发生变化),之后,重复上述过程,以得到系数回归网络的损失函数,并根据损失函数再次更新系数回归网络的网络参数,直到满足停止训练的条件为止。其中,停止训练的条件可以根据实际情况设置,例如,设置训练总次数,并在每次训练结束后,更新当前记录的训练次数,之后,判断当前记录的训练次数是否达到训练总次数,若达到,则结束训练。否则,开始新一次的训练。停止训练的条件还可以是损失函数收敛,损失函数收敛是指连续训练次数得到的损失函数均较小且其各损失函数间的差值在设定的范围内。满足停止训练的条件时,可以认为系数回归网络输出的人脸重建系数满足了准确重建人脸图像的需求。

上述,通过获取多帧包含同一目标对象的第一人脸图像,并利用具有相同网络参数的各系数回归网络处理各帧第一人脸图像,以得到人脸重建系数,之后,根据人脸重建系数重建出第一人脸图像对应的二维重建图像,根据第一人脸图像、二维重建图像和人脸重建系数构建无监督训练损失函数和身份约束损失函数,进而基于无监督训练损失函数和身份约束损失函数更新系数回归网络的网络参数的技术手段,解决了重建三维人脸图像时表情重建准确性低的技术问题。通过系数回归网络分别输出身份系数和表情系数,可以实现身份系数和表情系数的分别确定,并且,基于身份约束损失函数可以保证身份系数的准确性,即促使系数回归网络基于同一个人的不同图像可以得到相同的人脸身份,该人脸身份无关表情,进而保证表情系数的准确性。同时,除了对身份系数的准确性进行了约束,还构建了无监督训练损失函数,即在比较第一人脸图像和二维重建图像时,考虑了像素差异、特征差异和关键点差异三个维度,使用人脸重建系数时,分别对身份系数、表情系数进行约束,提高了损失函数的合理性,进而在基于损失函数训练系数回归网络时,提高了系数回归网络的性能,进一步保证了表情系数的准确性。

实施例二

图2为本申请一个实施例提供的一种用于人脸重建的网络训练方法的流程图。图2所示的方法是在上述方法的基础上，进行具体化。

图2所示的方法中，在重建三维人脸图像时，为了提高三维人脸图像的准确性，丰富了人脸重建系数的内容，例如，除了表情系数和身份系数外，人脸重建系数还可以包括纹理系数和位姿系数，通过纹理系数可以重建出三维人脸图像中的皮肤纹理，通过位姿系数可以重建出三维人脸图像中头部动作姿势，进而基于三维人脸图像得到二维重建图像时，使二维重建图像更加贴近于真实的第一人脸图像，以在计算损失函数，使得损失函数更能精准表达第一人脸图像与二维重建图像间的差异。

参考图2，该方法具体包括：

步骤210、获取第一目标对象的多帧第一人脸图像。

步骤220、将每帧第一人脸图像分别输入至对应的系数回归网络，以得到每帧第一人脸图像的人脸重建系数，各系数回归网络具有相同的网络参数，人脸重建系数包括身份系数、表情系数、纹理系数和位姿系数。

人脸重建系数除了包括身份系数和表情系数外，还包括纹理系数和位姿系数。

纹理系数用于描述当前输入的第一人脸图像的皮肤纹理，可选的，与身份系数和表情系数类似，预先设置可出现在人脸皮肤表面的各纹理基，每个纹理基表示一类可出现在人脸皮肤表情的纹理。纹理系数可以体现第一人脸图像的皮肤表面属于各纹理基的权重，纹理系数可以为矩阵向量的形式，纹理系数中的每个数值对应一个纹理基，数值越大，表明人脸中出现该纹理基的可能性越高。在重建过程中，根据纹理系数，将对应的纹理基应用在预先设置的纹理模型后，可以重建出第一人脸图像的皮肤纹理。

位姿系数用于描述当前输入的第一人脸图像中人脸头部动作的位姿，其中，第一人脸图像的位姿包括人脸的位置和姿态。在重建过程中，将位姿系数应用在重建的三维人脸图像中，可以重建出第一人脸图像中人脸的位姿。一般而言，通过位姿系数进行位姿变换时，通常包括旋转变换和平移变换，因此，位姿系数包括旋转系数和平移系数。旋转系数和平移系数均作用于三维空间中的三维坐标。一个实施例中，旋转系数采用欧拉旋转系数。

可理解，上述人脸重建系数仅为示例性描述，实际应用中，可以结合实际需求增加或减少人脸重建系数的类别。例如，实际应用中，当不需要重建出皮肤纹理时，可以删除纹理系数。又如，实际应用中，在人脸重建系数中增加用于描述当前输入的第一人脸图像中光照的光照系数，以在重建人脸图像时，模拟出对应的光照。

步骤230、根据人脸重建系数得到对应帧第一人脸图像的二维重建图像。

示例性的，根据人脸重建系数对第一目标对象进行重建时，不仅加入了身份系数和表情系数，还加入了纹理系数和位姿系数。一个实施例中，根据身份系数和表情系数重建出的三维人脸图像可以反映人脸的三维形状（即人脸的表面形状），因此，将身份系数和表情系数重建出的三维人脸图像记为三维人脸形状，即三维人脸形状可以描述第一目标对象的人脸身份以及在对应第一人脸图像中出现的表情。根据纹理系数可以重建出人脸的三维纹理，因此，将纹理系数重建出的三维人脸图像记为三维人脸纹理，即三维人脸纹理可以描述第一目标对象的皮肤纹理。可理解，通过每帧第一人脸图像对应的人脸重建系数均可重建出对应的三维人脸形状和三维人脸纹理。

之后，根据三维人脸形状、三维人脸纹理可以得到包含纹理、身份和表情的二维重建图像，并将位姿系数应用在二维重建图像中，以使二维重建图像包含第一人脸图像的头部动作。示例性的，得到三维人脸

形状和三维人脸纹理后,将三维人脸形状和三维人脸纹理映射到二维平面中,即将带有纹理的三维人脸(即三维人脸形状和三维人脸纹理)渲染到第一人脸图像的图像平面(即二维平面),便可以得到表示第一目标对象的二维重建图像,每帧第一人脸图像对应一个二维重建图像。在映射时,还可以引入位姿系数,以使二维人脸图像中的人脸具有一定的位姿。

基于三维人脸形状和三维人脸纹理得到二维重建图像时,步骤230包括步骤231-步骤232:

步骤231、根据身份系数和表情系数重建对应帧第一人脸图像的三维人脸形状,并根据纹理系数重建对应帧第一人脸图像的三维人脸纹理。

示例性的,根据人脸重建系数对第一目标对象的人脸进行重建。一个实施例中,重建的内容包括三维人脸形状和三维人脸纹理。

一个实施例中,三维人脸形状可以通过身份基和表情基叠加获取。此时,步骤231中,根据所身份系数和表情系数重建对应帧第一人脸图像的三维人脸形状具体包括:将身份系数和表情系数分别应用于对应的身份基和表情基,并结合预先构建的平均人脸形状,得到对应帧第一人脸图像的三维人脸形状。

三维人脸形状可以表示为:

$$\mathbf{S}(\alpha, \beta) = \bar{\mathbf{S}} + \mathbf{B}_{\text{id}} \alpha + \mathbf{B}_{\text{exp}} \beta \quad (2)$$

其中, $\mathbf{S}(\alpha, \beta)$ 表示重建的三维人脸形状, $\bar{\mathbf{S}}$ 表示预先构建的平均人脸形状, $\mathbf{B}_{\text{id}}$ 表示身份基底, $\mathbf{B}_{\text{exp}}$ 表示表情基底, α 表示身份系数, β 表示表情系数, 上述各符号对应的参数均可采用矩阵向量的形式表示。

其中, $\mathbf{S}(\alpha, \beta) \in \mathbb{R}^{n \times 3}$, $\bar{\mathbf{S}} \in \mathbb{R}^{n \times 3}$, $\mathbb{R}^{n \times 3}$ 表示包括 n 个三维坐标的数据集或矩阵, n 表示三维人脸中顶点的数量, 通常, 通过公式(2)在三维空间中构建的三维人脸可以通过三角网格的方式呈现, 三角网格可以由在三维空间中的顶点及三个顶点间的三角面片组成, 各三角网格组成了三维人脸。可理解, 每个顶点均有对应的三维坐标(即三维空间中的坐标), 此外, 每个顶点还可以有对应的颜色、法线等信息。

可选的, 平均人脸形状(即 $\bar{\mathbf{S}}$)的构建过程可以是: 将当前已公开的数据库(如 Basel Face Mode(BFM)数据库)中一定数量(如全部、200 或 100)的三维人脸的平均坐标(具体为三维坐标)数据确定为平均人脸形状, 该平均人脸形状为三维的人脸形状, 其可以认为是预先构建的人脸模型。

身份基底(即 α)是身份基的集合, 可选的, 身份基的构建过程是: 使用已公开的数据库(如 BFM 数据库), 基于主成分分析(Principal Component Analysis, PCA)得到各身份基。一个实施例中, 基于 PCA 在 BFM 数据库中构建了 80 个身份基。其中, 每个身份基均在身份系数中存在对应的系数。可理解, 人脸图像在一些情况下, 还会显示耳朵和颈部, 当前构建的身份基和对应的身份系数中均不包含对耳朵和颈部的限定。

表情基底是表情基的集合, 通常, 表情基可以表示人脸身份作出相应表情时, 所需要的形变量。可选的, 表情基的构建过程是: 基于 FaceWarehouse 数据集(包含人脸表情的数据集), 利用面部动作编码系统(Facial Action Coding System, FACS)通过肌肉动作对各基础表情进行分类以得到表示各基础表情的表情模板。FACS 对现实生活中人类的表情进行了分类, 它是面部表情的肌肉运动较常用的参照标准。根据人脸的解剖学特点, 将人脸划分成即相互独立又相互联系的运动单元(AU), 并分析了这些运动单元的运动特征以及在人脸中控制的主要区域及相关的表情, 进而描绘出不同的面部肌肉动作和不同的表情的对

应关系。其中，FACS 构建的每个表情模板可包括其表示的基础表情以及表明肌肉运动幅度的精细化值，例如，一个表情模板为 AU1 (0.2)，其中，AU1 表示内侧眉毛向上拉升的基础表情，0.2 为肌肉的拉升程度，且表情模板 AU1 (0.2) 中肌肉拉升幅度小于表情模板 AU1 (0.3)。一般而言，表明肌肉运动幅度的精细化值的范围是由 0 至 1，且值越大，表明肌肉运动幅度越强。一个实施例中，通过 FACS 构建了 46 表情模板。之后，对表情模板进行迁移，以使表情模板与 BFM 数据库的三维人脸的拓部结构一致，基于迁移后的表情模板便可得到表情基。其中，迁移可以理解为形变迁移，其是将一种拓扑结构下的三维人脸迁移到另一种拓扑结构的三维人脸。迁移的过程可以是：对 FaceWarehouse 数据集的三维人脸（记为源三维人脸）和 BFM 数据库的三维人脸（记为目标三维人脸）进行配准，建立源三维人脸和目标三维人脸的顶点关联，其中，顶点的位置可以根据实际情况选择，如将三维人脸划分成多个三角面片，各三角面片的顶点就是需要关联的顶点，之后，求解源三维人脸对应的各表情模板与源三维人脸的人脸身份间的形变量，该形变量可以通过仿射变换矩阵表示，之后，通过顶点的关联，利用形变量对目标三维人脸的人脸身份进行形变，便可得到各 BFM 数据库下的各表情模板。需说明的是，重建三维人脸形状时，需要迁移具有明确语义信息的表情基，例如，迁移表示“闭眼”的表情基、表示“皱眉”的表情基等，而使用已公开的数据库（如 BFM 数据库），基于主成分分析得到的表情基并没有明确的语义信息，因此，当前构建表情基时，使用明确人脸表情的 FaceWarehouse 数据集，并将 FaceWarehouse 数据集中各表情的语义信息迁移到 BFM 数据库的三维表情，以使重建三维人脸形状时使用的表情基具有语义信息。表情基底中每个表情基均在表情系数中找到对应的系数。

示例性的，系数回归网络输出身份系数和表情系数后，将身份系数和表情系数代入上述公式，便可以得到三维人脸形状。上述公式也可以认为是重建三维人脸形状的模型，该模型为线性的模型。

一个实施例中，重建三维人脸纹理时，预先构建平均的三维人脸纹理和纹理基底，之后，将纹理系数应用在纹理基底上，再与平均的三维人脸纹理进行叠加，便可以得到三维人脸纹理。此时，步骤 231 中根据纹理系数重建对应帧第一人脸图像的三维人脸纹理具体包括：将纹理系数应用于对应的纹理基，并结合预先构建的平均人脸纹理，得到对应帧第一人脸图像的三维人脸纹理。

三维人脸纹理可以表示为：

$$\mathbf{T}(\delta) = \bar{\mathbf{T}} + \mathbf{B}_{tex} \delta \quad (3)$$

其中， $\mathbf{T}(\delta)$ 表示重建的三维人脸纹理， $\bar{\mathbf{T}}$ 表示预先构建的平均人脸纹理， $\mathbf{B}_{tex}$ 表示纹理基底， δ 表示纹理系数，上述各符号对应的参数均可采用矩阵向量的形式表示。其中， $\mathbf{T}(\delta) \in \mathbb{R}^{n \times 3}$ ， $\bar{\mathbf{T}} \in \mathbb{R}^{n \times 3}$ 。

其中，平均人脸纹理与平均人脸形状的构建过程相似，平均人脸纹理的构建过程是：将当前已公开的数据库（如 Basel Face Mode (BFM) 数据库）中一定数量（如全部、200 或 100）的三维人脸的平均纹理数据确定为平均人脸纹理。

纹理基底是纹理基的集合，纹理基的构建过程与身份基的构建相似，纹理基的构建过程是：使用已公开的数据库（如 BFM 数据库），基于主成分分析（Principal Component Analysis, PCA）得到纹理基。一个实施例中，基于 PCA 在 BFM 数据库中构建了 80 个纹理基。其中，每个纹理基均在纹理系数中存在对应的系数。可理解，人脸图像在一些情况下，还会显示耳朵和颈部，当前构建的纹理基和对应的纹理系数中均不包含对耳朵和颈部的限定。

示例性的,系数回归网络输出纹理系数后,将纹理系数代入上述公式,便可以得到三维人脸纹理。上述公式也可以认为是重建三维人脸纹理的模型,该模型为线性的模型。

需要说明的是,公式(2)和公式(3)是三维人脸统计模型(即3DMM模型)常用的表达式,即当前将身份系数、表情系数和纹理系数代入3DMM模型后,便可以得到三维人脸形状和三维人脸纹理。3DMM模型是一种较为基础的三维人脸统计模型,可以根据二维的人脸图像(如第一人脸图像)对应的人脸重建系数生成三维人脸形状和三维人脸纹理,并可通过调整该3DMM模型的参数(即人脸重建系数)可以得到与二维的人脸图像最为近似的三维人脸形状和三维人脸纹理。

可理解,上述三维人脸形状和三维人脸纹理的重建方式仅是示例性描述,实际应用中,还可以采用其他的重建方式,当前不作限定。

每帧第一人脸图像均对应一个重建的三维人脸形状和三维人脸纹理。得到三维人脸形状和三维人脸纹理后,执行步骤232。

步骤232、根据位姿系数、三维人脸形状和三维人脸纹理渲染得到对应帧第一人脸图像的二维重建图像。

将三维人脸形状和三维人脸纹理映射到二维平面,并在映射过程中添加位姿系数,以得到二维重建图像。

一个实施例中,使用渲染器渲染二维重建图像。此时,步骤232包括步骤2321-步骤2322:

步骤2321、将位姿系数应用于透视投影模型。

透视投影(Perspective Projection)是为了获得接近真实三维物体的视觉效果而在二维的平面上绘图或者渲染的一种方法。透视投影模型是指使用透视投影的模型,该透视投影模型可以是指透视投影相机模型,其可以利用相机的内参和外参渲染得到二维图像,其中,内参是指相机内部参数,如内参包括相机的焦距(可根据经验选择)和第一人脸图像中心点的坐标。外参是指相机外部参数,当前,外参是指系数回归网络输出的位姿系数,即将位姿系数应用于透射投影模型。可选的,透视投影模型的数量可以与第一人脸图像的帧数相同,以将每帧第一人脸图像对应的位姿系数应用在对应的透视投影模型。

步骤2322、将透视投影模型、三维人脸形状和三维人脸纹理输入至渲染器,通过渲染器渲染得到对应帧第一人脸图像的二维重建图像。

示例性的,渲染器是软件渲染器。一个实施例中,渲染器为可微分渲染器,可微分渲染可理解为计算渲染过程的导数,实现将渲染过程放入神经网络中以解决更复杂的视觉问题,该渲染器也可以理解为差分渲染器。一个实施例中,通过可微分渲染器可以重建出二维重建图像。可理解,渲染器可以为一个或多个。当渲染器为多个时,渲染器的数量可以等于第一人脸图像的帧数,此时,每个渲染器渲染一张二维重建图像。当渲染器为一个时,渲染器依次重建出各第一人脸图像对应的二维重建图像。

一个实施例中,将应用位姿系数的透视投影模型、三维人脸形状和三维人脸纹理一同作为渲染器的输入,经过渲染器的处理输出二维重建图像。示例性的,渲染器工作过程中,将三维人脸形状投影到二维平面上,得到二维的图像,并将三维人脸纹理渲染在二维的图像上,在投影和渲染的过程中,使用应用位姿系数的透视投影模型,以得到包含头部动作姿态的二维重建图像。

一个实施例中,为了提高系数回归网络的性能,即提高二维重建图像的准确性,在人脸重建系数中还加入了光照系统。其中,光照系数可以体现采集第一人脸图像时的环境光照,进而增强三维人脸纹理。可选的,预先设置采集图像时出现的各光照类别,光照系数可以体现第一人脸图像的环境光照属于各光照类

别的权重，某类光照类别在光照系数中的数值越大，其出现在第一人脸图像中的可能性越高。基于此，人脸重建系数还包括：光照系数；步骤 232 之前，还包括：将光照系数应用于光照模型；将三维人脸纹理输入至光照模型，通过光照模型在三维人脸纹理中添加第一人脸图像所在的环境光照。

光照模型是根据光学的有关定律，模拟自然界中光照明的物理过程的计算机模型。当前使用的光照模型的类型可以根据实际情况设置，一个实施例中，以采用球谐（Spherical Harmonics）光照模型为例进行描述。球谐光照模型用于三维人脸纹理，以在三维人脸纹理中添加采集第一人脸图像时的环境光照，进而达到增强皮肤纹理的效果。此时，将系数回归网络输出的光照系数作为球谐光照模型的系数，之后，应用球谐光照模型处理三维人脸纹理，在处理过程中，球谐光照模型根据光照系数模拟光照环境，以实现在三维人脸纹理中添加第一人脸图像所在的环境光照。可理解，光照模型可以为一个或多个。当光照模型为多个时，光照模型的数量可以等于第一人脸图像的帧数，此时，每个光照模型渲染一张二维重建图像。当光照模型为一个时，光照模型依次处理各第一人脸图像对应的三维人脸纹理。实施例中，通过球谐光照模型，可以假设三维人脸的朗伯表面以接近真实场景的光照。其中，朗伯表面具有规则的物理模型和数学计算公式，并且在不破坏光的原有性质的前提下，能很好地近似光源和物体反射的光学现象（实施例中用于模拟对三维人脸纹理上添加真实场景的光照）。球谐光照模型应用过程中，使用球谐函数。其中，球谐函数是拉普拉斯方程的球坐标系形式解的角度部分，被广泛应用在计算机图形学中的渲染光照处理。球谐函数表记录有各球谐函数。球谐光照模型可以结合需要模拟的环境光照的精细程度，选择合适数量的球谐函数，一般而言，球谐函数越多，模拟的环境光照精细程度越高。一个实施例中，球谐光照模型使用球谐函数表的前三段球谐函数，即使用球谐函数表中 l 分别为 0、1 和 2 时的球谐函数，前三段球谐函数的数量为 9。一个实施例中，球谐光照模型处理三维人脸纹理时，将三维人脸纹理和三维人脸的表面法线一同输入至球谐光照模型。其中，使用 3DMM 重建三维人脸形状和三维人脸纹理时，还可以计算三维人脸的各顶点（顶点总数为 V ）的法向量，这些法向量可以认为是三维人脸的表面法线，通过调整表面法线，可以改变三维人脸表白的平滑程度。相关技术中，法向量是实现球谐光经常用的参数。

需说明的是，应用光照模型时，渲染器在重建二维重建图像时，向其输入的三维人脸纹理是经过光照系数和光照模型处理后的三维人脸纹理。可理解，实际应用中，身份系数、表情系数、纹理系数、光照系数和位姿系数均为向量的形式，并应用于通过渲染器重建二维重建图像，并在重建时使用 3DMM、光照模型和透视投影模型。也可以将身份系数、表情系数、纹理系数、光照系数和位姿系数记为 3DMM 系数，通过 3DMM 系数得到的二维重建图像也可以记为渲染图像。

可理解，对系数回归网络训练完成后，不同应用场景下，可以通过不同的系数重建出不同的人脸，例如，对人脸图像中的人脸进行重建并生成表情和动作一致的虚拟图像时，只需要使用表情系数和位姿系数。再如，对人脸图像中的人脸纹理、表情和身份进行重建时，只需要纹理系数、光照系数、表情系数和身份系数。

得到二维重建图像后，执行步骤 240。

步骤 240、根据第一人脸图像、相应的二维重建图像和人脸重建系数构建系数回归网络的损失函数，损失函数包括各帧第一人脸图像对应的无监督训练损失函数和身份约束损失函数。

示例性的，当前使用的二维重建图像是添加纹理和位姿的图像，还可以是添加纹理、位姿和光照的图像。

基于第一人脸图像、相应的二维重建图像和人脸重建系数构建的损失函数包括各帧第一人脸图像对应

的无监督训练损失函数和身份约束损失函数。其中，无监督训练损失函数包括：第一人脸图像的图像重构损失函数、图像感知损失函数、关键点损失函数和正则损失函数中的至少一种。

一个实施例中，无监督训练损失函数包括：第一人脸图像的图像重构损失函数，此时，根据第一人脸图像、相应的二维重建图像和人脸重建系数构建无监督训练损失函数包括步骤 241：

步骤 241、利用 L2 范数计算第一人脸图像和相应二维重建图像间的像素误差，并根据像素误差得到第一人脸图像的图像重构损失函数。

可理解，人脸重建系数越准确，第一人脸图像和相应的二维重建图像间相同位置（即坐标相同）的像素越相似，因此，可以基于第一人脸图像和二维重建图像间像素误差构建图像重构损失函数。一个实施例中，构建图像重构损失函数时使用 L2 范数。此时，图像重构损失函数的表达式为：

$$L_{pho} = \frac{\sum_{i \in \mathcal{M}} A_i \|I_i - \tilde{I}_i\|_2}{\sum_{i \in \mathcal{M}} A_i} \quad (4)$$

其中， L_{pho} 表示图像重构损失函数， $\mathcal{M}$ 表示二维重建图像中人脸所在的区域，即渲染器渲染二维重建图像时人脸所在的投影区域，通过 $\mathcal{M}$ 可以明确人脸区域中的各像素， i 表示像素索引， $i \in \mathcal{M}$ ，通过 i 可以在 $\mathcal{M}$ 中查找到对应的像素， A_i 表示人脸掩码中像素索引为 i 的像素的掩码值，其中，通过人脸分割算法在第一人脸图像中检测人脸掩码，人脸掩码可以体现第一人脸图像中显示人脸的像素，当 A_i 在第一人脸图像中对应的像素（即第一人脸图像中像素索引为 i 的像素）表示人脸时， A_i 为 1，反之， A_i 为 0。 I_i 表示第一人脸图像中像素索引为 i 的像素的像素值， $\tilde{I}_i$ 表示二维重建图像中像素索引为 i 的像素的像素值， $I_i - \tilde{I}_i$ 表示第一人脸图像和二维重建图像中像素索引为 i 的像素的像素值的差， $\|\cdot\|_2$ 表示 L2 范数。

可理解，每次得到第一人脸图像对应的二维重建图像后，便可以基于上述表达式得到 L_{pho} 的具体值。一次训练过程中，每帧第一人脸图像均对应一个图像重构损失函数。

一个实施例中，无监督训练损失函数包括：第一人脸图像的图像感知损失函数，此时，根据第一人脸图像、相应的二维重建图像和人脸重建系数构建无监督训练损失函数包括步骤 242-步骤 243：

步骤 242、利用人脸识别网络分别提取第一人脸图像的第一图像特征和对应二维重建图像的第二图像特征。

人脸识别网络为预先训练好的一个神经网络，其可以为卷积神经网络、残差神经网络等，且用于对输入的图像进行特征提取。将第一人脸图像输入至人脸识别网络后，人脸识别网络可以提取第一人脸图像中与人脸有关的深度特征，当前记为第一图像特征。同样的，将第一人脸图像对应的二维重建图像输入至人脸识别网络后，人脸识别网络可以提取二维重建图像中与人脸有关的深度特征，当前记为第二图像特征。

步骤 243、计算第一图像特征和第二图像特征之间的相似度，并根据相似度得到第一人脸图像的图像感知损失函数。

第一人脸图像与对应的二维重建图像越相似时，基于同一人脸识别网络提取的第一图像特征和第二图像特征越相似。因此，通过计算第一图像特征和第二图像特征之间的相似度，可以确定第一人脸图像与二维重建图像间的相似程度。据此，当前设计图像感知损失函数，该图像感知损失函数可以体现第一图像特征和第二图像特征之间的相似程度。可理解，对于本次训练过程而言，每帧第一人脸图像均有一个图像感

知损失函数。

一个实施例中，采用余弦相似度计算第一图像特征和第二图像特征之间的相似度，进而得到图像感知损失函数。其中，相似度越高，图像感知损失函数越小。示例性的，图像感知损失函数的表达式为：

$$L_{\text{per}} = 1 - \frac{f(I_1)f(I_2)}{\|f(I_1)\|\|f(I_2)\|} \quad (5)$$

其中， L_{per} 表示图像感知损失函数， $f(I_1)$ 表示第一人脸图像的第一图像特征， $f(I_2)$ 表示二维重建图像的第二图像特征， $f(I_1)f(I_2)$ 表示 $f(I_1)$ 和 $f(I_2)$ 的内积， $\|f(I_1)\|$ 表示 $f(I_1)$ 的长度， $\|f(I_2)\|$ 表示 $f(I_2)$ 的长度。

可理解，每次得到第一人脸图像对应的二维重建图像后，便可以基于上述表达式得到 L_{per} 的具体值。

一个实施例中，无监督训练损失函数包括第一人脸图像的关键点损失函数，此时，根据第一人脸图像、相应的二维重建图像和人脸重建系数构建无监督训练损失函数包括步骤 244-步骤 245：

步骤 244、获取第一人脸图像对应的第一人脸关键点和相应二维重建图像的第二人脸关键点。

示例性的，通过人脸关键点可以确定各关键部位的动作，一个实施例中，将第一人脸图像的人脸关键点记为第一人脸关键点，将二维重建图像的人脸关键点记为第二人脸关键点。可选的，为了便于计算，为每个人脸关键点设置一个关键点索引，且第一人脸图像和二维重建图像中具有相同语义信息的人脸关键点具有相同的关键点索引，如均表示左眉最左侧的人脸关键点具有相同的关键点索引。

人脸关键点的确定方式当前不作限定。例如，对于第一人脸图像而言，利用人脸识别算法检测出第一人脸图像中第一目标对象人脸所在的位置区域，之后，利用人脸关键点检测技术，在位置区域中检测到第一人脸关键点。再如，对于二维重建图像而言，得到三维人脸形状后，便可利用人脸关键点检测技术，在三维人脸形状中检测到三维的人脸关键点，之后，渲染器将三维人脸形状投影到二维平面时，可以将三维的人脸关键点投影到二维平面，以得到二维重建图像的第二人脸关键点。又如，利用人脸对齐方法 FAN 和 Dilb 获取第一人脸关键点和第二人脸关键点，FAN 和 Dilb 为不同的网络结构，通过 FAN 处理二维图像后可以得到人脸中三维的人脸关键点，通过 Dilb 可以得到人脸中二维的人脸关键点。FAN 对于面部轮廓的检测比较准确，但是，对于人脸眼部的区域检测准确度不如 Dilb，尤其是二维图像中的人脸图像为侧脸的图像时，Dilb 检测的眼部的人脸关键点的准确度高于 FAN，因此，二维图像显示侧脸时，使用 FAN 检测面部轮廓的人脸关键点，使用 Dilb 检测面部内部的人脸关键点。一个实施例中，获取第一人脸图像对应的第一人脸关键点包括：使用头部姿态估计算法处理第一人脸图像，以得到对应的偏航角初始位置；偏航角初始位置不超过角度阈值时，使用 Dilb 检测第一人脸图像中位于人脸内部的第一人脸关键点，并使用 FAN 检测第一人脸图像中位于人脸脸颊的第一人脸关键点；偏航角起始值超过角度阈值时，使用 FAN 检测第一人脸图像中的第一人脸关键点。其中，头部姿态估计算法可以获取二维图像中头部的姿态角。在 3D 空间中，物体的旋转可以由三个欧拉角来表示，三个欧拉角分别是计算俯仰角（pitch）、偏航角（yaw）和滚转角（roll），其中 pitch 表示围绕 X 轴旋转，对于头部而言，可理解为抬头时旋转的角，yaw 表示围绕 Y 轴旋转，对于头部而言，可理解为摇头时旋转的角，roll 表示围绕 Z 轴旋转，对于头部而言，可理解为转头时旋转的角。pitch、yaw 和 roll 均有对应的初始位置和目标位置。当前，通过头部姿态估计算法处理第一人脸图像后，获取 roll 的初始位置（即 roll yew），当前记为偏航角初始位置。之后，将偏航角初始位置与预设的角度阈值（可根据实际情况设置，如 15°）进行比较，若未超过角度阈值（即小于或等于角

度阈值), 则说明第一人脸图像中的面部基本为侧脸, 因此, 使用 Dilb 检测位于人脸内部 (即鼻子、嘴巴、眼镜和眉毛) 的第一人脸关键点, 使用 FAN 检测位于人脸轮廓 (即人脸的脸颊) 的第一人脸关键点。若超过角度阈值, 则说明第一人脸图像中的面部基本为正脸, 因此, 使用 FAN 检测位于人脸全部的第一人脸关键点。可理解, 通过 FAN 检测的人脸关键点为三维坐标, 此时, 只保留 X 轴和 Y 轴上的坐标, 以得到第一人脸关键点。通过 Dilb 检测的人脸关键点为二维坐标, 可以直接使用。

同样的, 获取第一人脸图像相应二维重建图像的第二人脸关键点包括: 使用头部姿态估计算法处理第一人脸图像相应的二维重建图像, 以得到对应的偏航角初始位置; 偏航角初始位置不超过角度阈值时, 使用 Dilb 检测二维重建图像中位于人脸内部的第二人脸关键点, 并使用 FAN 检测二维重建图像中位于人脸脸颊的第二人脸关键点; 偏航角起始值超过角度阈值时, 使用 FAN 检测二维重建图像中的第二人脸关键点。

步骤 245、计算第一人脸关键点和第二人脸关键点间位置的均方误差, 并根据均方误差得到第一人脸图像的关键点损失函数。

第一人脸图像与对应的二维重建图像越相似, 具有相同语义信息的第一人脸关键点在第一人脸图像的位置与第二人脸关键点在二维重建图像的位置越相似。因此, 可以通过计算第一人脸关键点和第二人脸关键点之间的位置差异, 确定第一人脸图像与二维重建图像间的相似程度。据此, 当前设计关键点损失函数, 该关键点损失函数可以体现第一人脸关键点和第二人脸关键点之间的位置差异程度。可理解, 对于本次训练过程而言, 每帧第一人脸图像均有一个关键点损失函数。

一个实施例中, 采用均方误差的方式计算具有相同语义的第一人脸关键点和第二人脸关键点之间的位置差异程度, 进而得到关键点损失函数。其中, 位置差异程度越小, 关键点损失函数越小。示例性的, 关键点损失函数的表达式为:

$$L_{i_{mn}} = \sum_{i=1}^n \omega_i \cdot \|Q_i - \hat{Q}_i\|^2 \quad (6)$$

其中, $L_{i_{mn}}$ 表示关键点损失函数, n 表示人脸关键点的总数量 (即人脸关键点的最大关键点索引的值), i 表示当前的关键点索引, ω_i 表示关键点索引为 i 的人脸关键点的权重, 权重可以根据实际需要预先设置, 例如, 将脸颊处的人脸关键点的权重设置为 1, 其余的人脸关键点的权重设置为 10, 以重点关注与人脸身份和表情有关的眼、鼻、嘴、眉。 Q_i 表示关键点索引为 i 的第一人脸关键点在第一人脸图像的位置 (即坐标), $\hat{Q}_i$ 表示关键点索引为 i 的第二人脸关键点在二维重建图像的位置 (即坐标)。 $\|\cdot\|^2$ 表示二范数的平方。可理解, 每次得到第一人脸图像对应的二维重建图像后, 便可以基于上述函数公式得到 $L_{i_{mn}}$ 的具体值。

一个实施例中, 无监督训练损失函数包括: 第一人脸图像的正则损失, 此时, 根据第一人脸图像、相应的二维重建图像和人脸重建系数构建无监督训练损失函数包括步骤 246-步骤 248:

步骤 246、计算第一人脸图像的身份系数的第一范数平方以及纹理系数的第二范数平方, 并根据第一范数平方和第二范数平方得到第一人脸图像的身份及纹理约束损失函数。

为了防止三维人脸形状和三维人脸纹理的退化, 对人脸重建系数中的身份系数和纹理系数进行约束,

以构建对应的损失函数,当前,将该损失函数记为身份及纹理约束损失函数。一个实施例中,身份及纹理约束损失函数的表达式为:

$$L_{reg} = \sum_{n=1}^M (\lambda_{\alpha} \|\alpha_n\|^2 + \lambda_{\delta} \|\delta_n\|^2) \quad (7)$$

其中, L_{reg} 表示身份及纹理约束损失函数, M 为身份基以及纹理基的总数量 (即总分类数), 一个实施例中, 身份基和纹理基的总数量相同, 均为 80, 即 $M=80$, n 表示纹理基的索引和身份基的索引, α_n 表示索引为 n 的身份基在身份系数中对应的系数, δ_n 表示索引为 n 的纹理基在纹理系数中对应的系数。 $\|\cdot\|^2$ 表示二范数的平方。 λ_{α} 和 λ_{δ} 均为超参数, 其具体值可以根据实际情况设置, 例如, 当前分别设置为 1 和 0.001, 再如, 当前分别设置为 1 和 $1.7e-3$ (约等于 0.0017)。可理解, 每次得到第一人脸图像对应的人脸重建系数后, 便可以基于上述表示得到 L_{reg} 的具体值。每帧第一人脸图像均对应一个身份及纹理约束损失函数。

步骤 247、使用 L1 范数对第一人脸图像的表情系数进行约束, 以得到第一人脸图像的表情约束损失函数。

为了防止表情系数的稀疏化, 对人脸重建系数中的表情系数使用 L1 正则来进行约束, 以构建对应的损失函数, 当前, 将该损失函数记为表情约束损失函数。一个实施例中, 表情约束损失函数的表达式为:

$$L_{sp} = \sum_{i=1}^m \lambda_{\alpha} \|\beta_i\| \quad (8)$$

其中, L_{sp} 表示表情约束损失函数, m 为表情基的总数量 (即总分类数), 一个实施例中, 表情基共有 46 个, 即 $m=46$, i 表示表情基的索引, β 表示索引为 i 的表情基在表情系数中对应的系数。 $\|\cdot\|$ 表示 L1 范数。 λ_{α} 为超参数, 其具体值可以根据实际情况设置, 例如, λ_{α} 的值可以与上述身份及纹理约束损失函数中 λ_{α} 的值相同或不同。可理解, 每次得到第一人脸图像对应的人脸重建系数后, 便可以基于上述表达式到 L_{sp} 的具体值。每帧第一人脸图像均对应一个表情约束损失函数。

可理解, 步骤 246 和步骤 247 可以同时执行或者前后执行, 其具体的执行顺序当前不作限定。

步骤 248、根据身份及纹理约束损失函数和表情约束损失函数得到第一人脸图像的正则损失。

示例性的, 正则损失由身份及纹理损失函数和表情约束损失函数组成。一个实施例中, 可以分别为身份及纹理损失函数和表情约束损失函数设置对应的超参数。超参数的具体值可以根据实际情况设置, 当前不作限定。

可理解, 当无监督训练损失函数包含第一人脸图像的图像重构损失函数、图像感知损失函数、关键点损失函数和正则损失函数中的多种时, 各损失函数可以同时计算, 也可以按照设定的顺序进行计算, 当前不作限定。可理解, 上述无监督训练损失函数也可以认为是自监督损失函数, 即无监督训练损失函数中使用的标签并不是人工标注的, 而是从无监督的数据 (即人脸图像) 中挖掘出监督信息, 即构建损失函数需要的标签。例如, 图像重构损失函数使用的 I_i 、图像感知损失函数使用的 $f(I_i)$ 、关键点损失函数使用的 Q_i 均

是从第一人脸图像中挖掘出的标签。

并且,可以为每个损失函数设置对应的超参数,以得到最终的无监督训练损失函数。举例而言,无监督训练损失函数包含第一人脸图像的图像重构损失函数、图像感知损失函数、关键点损失函数和正则损失函数。此时,无监督训练损失函数可以表示为:

$$\mathcal{L}_{\text{com}} = \lambda_{\text{phs}} \mathcal{L}_{\text{phs}} + \lambda_{\text{per}} \mathcal{L}_{\text{per}} + \lambda_{\text{im}} \mathcal{L}_{\text{im}} + \lambda_{\text{reg}} \mathcal{L}_{\text{reg}} + \lambda_{\text{sp}} \mathcal{L}_{\text{sp}} \quad (9)$$

其中, $\mathcal{L}_{\text{com}}$ 表示无监督训练损失函数, $\mathcal{L}_{\text{phs}}$ 表示图像重构损失函数, λ_{phs} 为 $\mathcal{L}_{\text{phs}}$ 的超参数, $\mathcal{L}_{\text{per}}$ 表示图像感知损失函数, λ_{per} 为 $\mathcal{L}_{\text{per}}$ 的超参数, $\mathcal{L}_{\text{im}}$ 表示关键点损失函数, λ_{im} 为 $\mathcal{L}_{\text{im}}$ 的超参数, $\mathcal{L}_{\text{reg}}$ 表示身份及纹理约束损失函数, λ_{reg} 为 $\mathcal{L}_{\text{reg}}$ 的超参数, $\mathcal{L}_{\text{sp}}$ 表示表情约束损失函数, λ_{sp} 为 $\mathcal{L}_{\text{sp}}$ 的超参数, $\mathcal{L}_{\text{reg}}$ 和 $\mathcal{L}_{\text{sp}}$ 组成正则损失函数。各超参数的具体值可以结合对应损失函数的重要程度以及系数回归网络的训练结果设置,例如,设置 λ_{phs} 为 1.9, λ_{per} 为 0.2, λ_{im} 为 0.1, λ_{reg} 为 0.0001, λ_{sp} 为 0.1。

按照上述公式便可以得到每帧第一人脸图像对应的无监督训练损失函数。

除了构建无监督训练损失函数,还需要构建身份约束损失函数。为了使得系数回归网络输出准确的表情系数,需要保证系数回归网络输出准确的身份系数,即重建的三维人脸形状准确的前提下,只有拟合准确的身份系数(即输出准确的身份系数)才能保证拟合准确的表情系数。一个实施例中,根据第一人脸图像、相应的二维重建图像和人脸重建系数构建系数回归网络的身份约束损失函数包括步骤 249-步骤 2411:

步骤 249、获取各帧第一人脸图像对应的身份标签系数。

示例性的,第一人脸图像的身份标签系数是指第一人脸图像已知的身份系数,其可以作为先验信息。即获取第一人脸图像时,同时明确第一人脸图像的身份标签系数。

步骤 2410、将各身份标签系数的平均系数作为伪标签监督系数。

一个实施例中,计算各帧第一人脸图像的身份标签系数的平均系数,并将平均系数作为伪标签来监督系数回归网络输出准确的身份系数。当前,将平均系数记为伪标签监督系数。可选的,对于第一目标对象而言,伪标签监督系数减少了多帧第一人脸图像中不同的表情所引起的人脸身份的变化,因此,伪标签监督系数的准确度高于单帧第一人脸图像的身份标签系数。实际应用中,也可以使用每个第一人脸图像的身份标签系数来监督对应系数回归网络输出准确的身份系数。

步骤 2411、计算第一人脸图像的身份系数与伪标签监督系数之间的均方误差,并根据均方误差得到身份约束损失函数。

一个实施例中,一帧第一人脸图像的身份约束损失函数定义为伪标签监督系数与该帧第一人脸图像对应的身份系数之间的均方误差。此时,身份约束损失函数的表达式为:

$$\mathcal{L}_{\text{idc}}^t = \|\alpha_t - \bar{\alpha}\|^2 \quad (10)$$

其中, $\mathcal{L}_{\text{idc}}^t$ 表示第 t 帧的第一人脸图像的身份约束损失函数, α_t 表示第 t 帧的第一人脸图像对应的身份系数,其为矩阵向量的形式, $\bar{\alpha}$ 表示伪标签监督系数,其为矩阵向量的形式。 $\|\cdot\|^2$ 表示二范数的平方。可理解,身份系数越准确, $\alpha_t - \bar{\alpha}$ 间的差值越小。该身份约束损失函数可以理解为进行身份一致性的约束,通

过身份一致性约束，可以从同一人脸的多帧人脸图像预测出一致性的身份系数。

系数回归网络输出身份系数后，便可以根据上述表达式得到 $\hat{z}_{idc}^t$ 的具体值。并且，每帧第一人脸图像均存在一个对应的身份约束损失函数。

一个实施例中，得到各帧第一人脸图像的身份约束损失函数和无监督训练损失函数后，便可以构建各系数回归网络的损失函数，且各系数回归网络的损失函数相同。

步骤 250、根据损失函数更新系数回归网络的网络参数。

上述，通过获取多帧包含同一目标对象的第一人脸图像，并利用具有相同网络参数的各系数回归网络处理各帧第一人脸图像，以得到人脸重建系数，之后，根据人脸重建系数重建出第一人脸图像对应的二维重建图像，根据第一人脸图像、二维重建图像和人脸重建系数构建无监督训练损失函数和身份约束损失函数，进而基于无监督训练损失函数和身份约束损失函数更新系数回归网络的网络参数的技术手段，解决了重建三维人脸图像时表情系数重建准确性低的技术问题。通过系数回归网络分别输出身份系数、表情系数、纹理系数和位姿系数，以得到更为准确的二维重建图像，进而使得二维重建图像和第一人脸图像间像素差异、特征差异和关键点差异更为准确，提高了无监督训练损失函数的准确性，并且，基于身份约束损失函数可以保证身份系数的准确性，进而保证表情系数的准确性。

实施例三

图 3 为本申请一个实施例提供的一种用于人脸重建的网络训练方法的流程图。图 3 所示的方法是在上述方法的基础上，进行具体化。

可理解，对于人脸的表情而言，有些表情是不可能同时出现在人脸上的，例如，表示左眼闭的表情基和表示左眼瞪眼的表情基不能同时出现在人脸上，这类不能同时出现的两个表情基可以记为互斥表情对，而系数回归网络输出的表情系数包含了互斥表情对中两个表情基的系数，即两个表情基在系数回归网络的预测过程中都处于激活状态，此时，为了避免互斥表情对对表情系数准确性的影响，可以基于系数回归网络输出的表情系数，构建与互斥表情对有关的损失函数，以避免互斥表情对包含的表情基同时出现，提高表情系数的准确性，进而提高系数回归网络的性能。

据此，参考图 3，该用于人脸重建的网络训练方法包括：

步骤 310、获取第一目标对象的多帧第一人脸图像。

步骤 320、将每帧第一人脸图像分别输入至对应的系数回归网络，以得到每帧第一人脸图像的人脸重建系数，各系数回归网络具有相同的网络参数，人脸重建系数包括：身份系数、表情系数、纹理系数、位姿系数和光照系数。

步骤 330、根据人脸重建系数得到对应帧第一人脸图像的二维重建图像。

步骤 340、根据第一人脸图像、相应的二维重建图像和人脸重建系数构建系数回归网络的损失函数，损失函数包括各帧第一人脸图像的无监督训练损失函数、身份约束损失函数和和表情互斥约束损失函数。

其中，无监督训练损失函数可参考前述内容。

一个实施例中，带有身份标签系数的图像数量有限，不利于系数回归网络的训练，因此可以使用未知人脸身份的图像进行训练，即第一人脸图像没有对应的身份标签系数。此时，构建身份约束损失函数时，可以使用系数回归网络输出的身份系数的平均作为伪标签监督系数，以构建身份约束损失函数。此时，根据所述第一人脸图像、相应的二维重建图像和所述人脸重建系数构建系数回归网络的身份约束损失函数包

括：根据各系数回归网格输出的身份系数，构建伪标签监督系数；根据第一人脸图像的身份系数与伪标签监督系数得到第一人脸图像的身份约束损失函数。

可理解，对于同一人脸来说，其人脸身份在不同人脸图像中应该是一致的，但是，由于表情等因素，可能使得同一人脸在不同人脸图像中识别出的人脸身份存在偏差，因此，通过多帧第一人脸图像对应的身份系数的平均，可以克服不同表情所引起的人脸身份的变化。此时，根据各系数回归网格输出的身份系数，构建伪标签监督系数包括：计算各系数回归网格输出的身份系数的平均系数，并将平均系数作为伪标签监督系数；根据第一人脸图像的身份系数与伪标签监督系数得到第一人脸图像的身份约束损失函数包括：计算第一人脸图像的身份系数与伪标签监督系数之间的均方误差，并根据均方误差得到第一人脸图像的身份约束损失函数。此时，第 t 帧对应的身份约束损失函数仍然可表示为公式 (10)，其中， $L_{i\alpha\epsilon}^t$ 表示第 t 帧的第一人脸图像的身份约束损失函数， α_t 表示第 t 帧的第一人脸图像对应的身份系数，其为矩阵向量的形式，

$\bar{\alpha}$ 表示伪标签监督系数，其为矩阵向量的形式。 $\|\cdot\|^2$ 表示二范数的平方。此时， $\bar{\alpha} = \frac{1}{T} \sum_{t=1}^T \alpha_t$ 。其中， T 表示第一人脸图像的总帧数。通过这种方式，无需获取已知人脸身份的训练数据，也可以使得系数回归网络得到较为准确的人脸身份。此外，也可以采用前述实施例的身份约束损失函数的计算过程。

一个实施例中，在构建的各表情基中，有一些表情基是互斥的，其不能同时出现在人脸中。但是，在系数回归网络的训练过程中，由于对表情系数没有直接的监督信息（即预先不知道第一人脸图像中出现的表情基），因此，互斥的表情基不可避免地会被系数回归网络同时预测成激活状态，即表情系数同时包含互斥的表情基对应的系数，这样会阻碍表情系数的准确拟合，影响表情系数的准确性。因此，系数回归网络的损失函数还包括：各帧第一人脸图像的表情互斥约束损失函数，以指导系数回归网络抑制那些不应该同时出现的互斥表情集。

一个实施例中，构建表情互斥约束损失函数时，根据所述第一人脸图像、相应的二维重建图像和所述人脸重建系数构建表情互斥约束损失函数包括步骤 341-步骤 343：

步骤 341、获取预先构建的全部互斥表情对，每个互斥表情对包含至少两个互斥的表情基；

示例性的，将互斥的表情基记为互斥表情对，每个互斥表情对包含至少两个不可能同时出现在人脸上的互斥的表情基。一个实施例中，以互斥表情对包含两个互斥的表情基为例进行描述。当前，预先在构建的表情基中查找出全部的互斥表情对。例如，图 4 为本申请一个实施例提供的一种互斥表情对示意图。参考图 4，其示出了查到的部分（共 6 组）互斥表情对，以左上互斥表情对为例进行描述，其示出了表示右眼闭眼的表情基和表示右眼瞪眼的表情基为一组互斥表情对。其中，B1 表示表情基索引为 1 的表情基。B9 表示表情基索引为 9 的表情基。可理解，训练系数回归网络前，先查找全部的互斥表情对，在训练过程中，可以直接使用查找的互斥表情对，无需重复查找。

步骤 342、查找各互斥表情对包含的表情基在第一人脸图像中的表情系数，得到第一人脸图像的表情互斥约束损失函数。

示例性的，系数回归网络输出表情系数后，可以在表情系数中查找互斥表情对中全部表情基对应的系数，例如，互斥表情对包含 B1 和 B9 表情基，在表情系数中可以找到 B1 对应的系数和 B9 对应的系数。

步骤 343、根据查找到的表情系数，利用指数函数得到第一人脸图像的表情互斥约束损失函数。

示例性的,根据查找到的各互斥表情对的系数,构建表情互斥约束损失函数。一个实施例中,利用指示函数构建表情互斥约束损失函数,此时,表情互斥约束损失函数的表达式为:

$$L_{exp} = \sum_{(i,j) \in O} \|\mathbb{I}\{\beta_i > \beta_j\} \cdot \beta_j\|^2 + \|\mathbb{I}\{\beta_j > \beta_i\} \cdot \beta_i\|^2 \quad (11)$$

其中, L_{exp} 表示表情互斥约束损失函数, O 表示各互斥表情对组成的集合, (i, j) 为 O 中的一个互斥表情对且所述互斥表情对包含第 i 个表情基 (索引为 i 的表情基) 和第 j 个表情基 (索引为 j 的表情基), β_i 和 β_j 分别表示第 i 个表情基和第 j 个表情基在第一人脸图像中的表情系数, $\mathbb{I}\{\cdot\}$ 表示指数函数, $\|\cdot\|^2$ 表示二范数的平方。可理解,对于指数函数而言,若其为真,则 $\mathbb{I}\{\cdot\}=1$, 若其为假,为 $\mathbb{I}\{\cdot\}=0$ 。对应到表达式中,若 $\beta_i > \beta_j$ 为真,则 $\mathbb{I}\{\beta_i > \beta_j\}$ 等于 1, 若 $\beta_i > \beta_j$ 为假,则 $\mathbb{I}\{\beta_i > \beta_j\}$ 等于 0。

通过上述公式,可以使得表情互斥约束损失函数仅保留互斥表情对中在第一人脸图像中更可能出现的表情基的系数,进而基于表情互斥约束损失函数调整系数回归网络的网络参数时,使系数回归网络学习不应同时出现的互斥表情对,进而促进了表情系数的准确拟合,即提高表情系数的准确性。

可理解,无监督训练损失函数、身份约束损失函数和表情互斥约束损失函数可以同时计算,也可以前后计算,当前不限定具体的计算顺序。

得到各帧第一人脸图像对应的无监督训练损失函数、身份约束损失函数和表情互斥损失函数后,确定系数回归网络的损失函数。

此时,各系数回归网络的损失函数的表达式为:

$$L = \sum_{t=1}^T (L_{com}^t + \lambda_{idc} L_{idc}^t + \lambda_{exp} L_{exp}^t) \quad (12)$$

其中, L 表示系数回归网络的损失函数, T 为当前获取的第一人脸图像的总帧数, t 表示第 t 帧, L_{com}^t 表示第 t 帧第一人脸图像对应的无监督训练损失函数, L_{idc}^t 表示第 t 帧第一人脸图像对应的身份约束损失函数, L_{exp}^t 表示第 t 帧第一人脸图像对应的表情互斥约束损失函数, λ_{idc} 表示身份约束损失函数的超参数, λ_{exp} 表示表情互斥约束损失函数的超参数。为身份约束损失函数以及表情互斥约束损失函数设置超参数,可提高损失函数的精确性。各超参数的具体值可以根据实际情况设置,例如, L_{idc}^t 的重要性高于 L_{exp}^t , 所以,设置 λ_{idc} 为 1000, λ_{exp} 为 10。即损失函数中所述身份约束损失函数的超参数大于所述表情互斥约束损失函数的超参数。

步骤 350、根据损失函数更新系数回归网络的网络参数。

上述,利用先验知识收集了互斥表情对,并针对互斥表情对中表情基共同出现的问题设计了表情互斥约束损失函数,以将互斥表情的知识注入到系数回归网络的训练过程中,使系数回归网络学习抑制互斥表情的共同出现,从而进一步提高了表情系数的准确性。并且,通过系数回归网络输出的身份系数构建身份约束损失函数的伪标签监督系数,可以在标签数据 (即已知人脸身份的人脸图像) 较少时,仍实现对系数回归网络的训练。

下面对于用于人脸重建的网络训练方法进行示例性描述。其中,图 5 为本申请一个实施例提供的一种用

于人脸重建的网络训练方法的框架示意图。参考图 5，从一个视频数据中随机截取多帧第一人臉图像（当前为 4 帧第一人臉图像），并分别输入至对应的系数回归网络，各系数回归网络共享网络参数，由系数回归网络输出人脸重建系数，该人脸重建系数包括身份系数 α ， $\alpha \in R^{80}$ （有 80 个身份基）、表情系数 β ， $\beta \in R^{46}$ （有 46 个表情基）、纹理系数 δ ， $\delta \in R^{80}$ （有 80 个纹理基）、光照系数 γ ， $\gamma \in R^{27}$ （有 27 个光照类型）、位姿系数 p ， $p \in R^6$ （三个方向上的坐标和角度）。之后，根据人脸重建系数和三维形变人脸模型（包括平均人脸形状和平均人脸纹理）得到三维人脸图像（包括三维人脸形状和三维人脸纹理），之后，利用可微分渲染器渲染出二维重建图像，之后，根据各系数回归网络输出的身份系数构建身份约束损失函数（图 5 中记为 L_{idc} ），根据各系数回归网络输出的表情系数构建表情互斥约束损失函数（图 5 中记为 L_{exp} ），结合二维重建图像、第一人臉图像和人脸重建系数构建无监督训练损失函数，并结合无监督训练损失函数、身份约束损失函数和表情互斥约束损失函数更新系数回归网络的网络参数，之后，继续训练，直到满足迭代次数（即停止训练的次数）。

一个实施例中，对系数回归网络的训练结束后，可以应用系数回归网络。即将系数回归网络部署在需要的设备中，该设备可以是用于人脸重建的网络训练设备，还可以是其他的设备，当前以用于人脸重建的网络训练设备应用系数回归网络为例进行描述。

图 6 为本申请一个实施例提供的一种系数回归网络的应用流程示意图。参考图 6，系数回归网络应用时，包括如下步骤：

步骤 410、获取第二目标对象的第二人脸图像。

示例性的，第二目标对象为应用过程中需要构建虚拟画像的对象。第二人脸图像为包含第二目标对象人脸的图像。第二人脸图像的来源当前不作限定。例如，视频交流场景中，通过摄像头进行拍摄，此时，拍摄的视频数据中显示的对象（即人类）为第二目标对象，对于视频数据中当前帧图像而言，先判断其中是否包含第二目标对象的人脸图像，若包含第二目标对象的人脸图像，则提取其中的人脸图像作为第二人脸图像，否则，不对当前帧图像进行处理。其中，判断是否包含人脸图像的方式当前不作限定，例如，使用基于深度学习的人脸检测算法确定是否在当前帧图像中检测到人脸区域，若检测到人脸区域，则确定包含人脸图像，否则，确定不包含人脸图像。检测到人脸区域后，便可以提取该人脸区域，即对当前帧图像进行剪裁，以得到第二人脸图像。再如，利用人脸检测器检测当前帧图像中是否包含人脸框，若包含人脸框，则确定包含人脸图像，否则，确定不包含人脸图像。之后，对人脸框中的人脸区域进行剪裁，以得到第二人脸图像。

步骤 420、将第二人脸图像输入至系数回归网络，以得到第二人脸图像的人脸迁移系数，人脸迁移系数包括表情系数和位姿系数。

示例性的，由于训练的各系数回归网络具有相同的网络参数，因此，应用过程中，可以直接选择其中一个系数回归网络。之后，将第二人脸图像输入至该系数回归网络，并获取系数回归网络的输出。一个实施例中，将系数回归网络应用过程中输出的人脸重建系数记为人脸迁移系数。可理解，可以根据实际情况，设置部分人脸重建系数作为人脸迁移系数。例如，当前需要进行表情迁移，因此，选择表情系数作为人脸迁移系数，此时，系数回归网络可以只输出表情系数，或者是，系数回归网络还是输出全部的人脸重建系数，之后，设备仅获取表情系数作为人脸迁移系数。再如，当前需要进行表情和位姿迁移，因此，选择表

情系数和位姿系数作为人脸迁移系数，此时，系数回归网络可以只输出表情系数和位姿系数，或者是，系数回归网络还是输出全部的人脸重建系数，之后，设备仅获取表情系数和位姿系数作为人脸迁移系数。

步骤 430、根据表情系数和位姿系数生成与第二人脸图像对应的虚拟画像。

示例性的，预先存储虚拟图像，该虚拟图像可以为卡通图像，其可以为二维虚拟图像也可以为三维虚拟图像。一个实施例中，以三维虚拟图像进行为例，并且，当前还存储三维虚拟图像的各表情基。示例性的，根据表情系数和位姿系数对虚拟图像进行处理，以生成与第二人脸图像具有相同表情以及相同头部动作姿态的虚拟画像，该虚拟画像生成后，可以显示在设备中。一个实施例中，还可以由其他的设备生成虚拟画像，此时，将表情系数和位姿系数发送至其他的设备，由其他的设备将表情系数和位姿系数应用在虚拟图像中，以生成虚拟画像。

一个实施例中，生成与第二人脸图像相对应的虚拟画像时，可通过开源图形库（Open Graphics Library，OpenGL）的图形渲染框架对预先设定的三维虚拟图像进行渲染和显示。其中，在渲染时具体为根据表情系数对三维虚拟图像的表情基进行线性加权，得到包含表情的三维虚拟图像。生成包含表情的三维虚拟图像后，图形渲染框架根据位姿参数对包含表情的三维虚拟图像进行相应的刚性变换，以得到虚拟画像，并进行显示。

举例而言，图 7 为本申请一个实施例中提供的虚拟图像生成框架图。参考图 7，对于一帧图像而言，利用人脸检测器检测人脸框，并对人脸区域进行剪裁，以得到仅包含人脸的第二人脸图像，之后，将第二人脸图像输入至系数回归网络，以得到位姿系数和表情系数，之后，将位姿系数和表情系数迁移到虚拟图像上，以得到虚拟画像。

可理解，对于视频交流场景而言，生成当前帧的虚拟画面后，便可以获取下一帧图像，并将下一帧图像作为当前帧图像，重复上述过程，以持续显示虚拟画像。

上述，系数回归网络应用过程中，使用系数回归网络得到的表情系数和位姿系数具有较高的准确度，进而基于表情系数和位姿系数得到的虚拟画像与第二人脸图像间的表情和头部动作姿态更为相似，有效保护了目标对象的隐私，防止信息泄露，同时，虚拟图像对第二人脸图像中的表情和位姿进行准确跟随。

实施例四

本申请一个实施例还提供了一种用于人脸重建的网络训练方法，该方法与前述实施例所述的方法均是基于神经网络输出人脸图像的身份系数、表情系数、纹理系数、光照系数和位姿系数，并重建出人脸图像，进而基于重建的人脸图像和原人脸图像构造出适合的损失函数以训练神经网络，进而基于神经网络重建出准确表情的人脸图像。

该方法基于比较成熟的 3DMM 模型（人脸 3D 形变统计模型）实现，先通过训练好的第一参数估计网络模型和预设的 3DMM 模型确定输入的二维人脸图像对应的粗糙三维人脸形状（为了与前述实施例的三维人脸形状区分，这里记为粗糙 3D 表情模板），再根据该粗糙三维人脸形状（即粗糙 3D 表情模板）确定对应的精细三维人脸形状（为了与前述实施例的三维人脸形状区分，这里记为精细 3D 表情模板），并且，根据输入的二维人脸图像，通过训练好的第二参数估计网络模型输出 3DMM 参数，该 3DMM 参数包含准确的表情系数。该方法充分考虑了不同人之间同一个表情的差异性，为每个人构造不同的精细化表情形状，充分体现了不同个体间的个性化，基于本方法生成的精细三维人脸形状（即精细 3D 表情模板），在拟合二维人脸图像的表情时能够更好地捕捉待人脸的准确表情，从而提高人脸驱动的准确性。

其中，3DMM 模型是一种较为基础的三维人脸统计模型。一个实施例中，3DMM 模型根据输入的二维

维人脸图像重建生成三维人脸形状（为了与前述实施例的三维人脸形状区分，这里记为 3D 表情模板），并可通过调整该 3DMM 模型的参数（即 3DMM 参数）得到与输入的二维人脸图像最为近似的三维人脸形状（即 3D 表情模板）。对于每个人的二维人脸图像，都有一组对应的 3DMM 参数，使通过该组模型参数下的 3DMM 模型生成的三维人脸形状（即 3D 表情模板）与该二维人脸图像中的人脸最为近似。

3DMM 模型可由网格（Mesh）组成，网格通常是指三角网格，三角网格可以由在三维空间中的顶点及三个顶点间的三角面片组成。每个顶点除了位置坐标以外，还可以包含颜色、法线等信息。3DMM 参数可以包括但不限于身份系数、表情系数、纹理（色彩亮度）系数、光照系数及头部位姿系数（即位姿系数），可以理解为 3DMM 模型中关于身份、表情、纹理、光照及头部位姿的加权值。3DMM 模型的每一维系数控制人脸的局部的变化。

实施例中，3DMM 模型可以采用如下公式（13）和（14）表示，并使用 Basel Face Model 数据集的基于主成分分析(PCA)的身份基底和纹理基底，以具有明确语义信息的表情模板（如瞪眼、闭眼、皱眉、挑眉等表情单元）的 FaceWarehouse 数据集基于面部动作编码系统 FACS 的 46 个表情模板偏移量（即三维人脸形状相对于 3DMM 模型的预设标准人脸形状的形变量）作为本实施例使用的表情基底对该 3DMM 模型及本实施例提供的人脸形状生成方法进行详细说明。

$$\mathbf{S}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \bar{\mathbf{S}} + \mathbf{B}_{id}\boldsymbol{\alpha} + \mathbf{B}_{exp}\boldsymbol{\beta} \quad (13)$$

$$\mathbf{T}(\boldsymbol{\delta}) = \bar{\mathbf{T}} + \mathbf{B}_{tex}\boldsymbol{\delta} \quad (14)$$

其中， $\bar{\mathbf{S}} \in \mathbb{R}^{n \times 3}$ ， $\bar{\mathbf{T}} \in \mathbb{R}^{n \times 3}$ 分别表示该 3DMM 模型的平均人脸的形状（即预设标准人脸形状）坐标值和纹理像素值， n 表示 3D 人脸的顶点数， $\mathbb{R}^{n \times 3}$ 表示包括 n 个三维坐标的数据集或矩阵， $\mathbf{S}(\boldsymbol{\alpha}, \boldsymbol{\beta})$ 表示 3D 人脸顶点的三维坐标（也可理解为重建的三维人脸形状，该三维人脸形状可以描述二维图像中的人脸身份以及出现的表情）， $\mathbf{T}(\boldsymbol{\delta})$ 表示 3D 人脸顶点的 RGB 顶点颜色的像素值（也可理解为重建的三维人脸纹理，该三维人脸纹理可以描述二维图像中人脸的皮肤纹理）。 $\mathbf{B}_{id} \in \mathbb{R}^{2n \times 80}$ 、 $\mathbf{B}_{exp} \in \mathbb{R}^{2n \times 46}$ 和 $\mathbf{B}_{tex} \in \mathbb{R}^{2n \times 80}$ 分别表示身份基底、表情基底和纹理基底， $\boldsymbol{\alpha} \in \mathbb{R}^{80}$ 、 $\boldsymbol{\beta} \in \mathbb{R}^{46}$ 与 $\boldsymbol{\delta} \in \mathbb{R}^{80}$ 分别是相应的 3DMM 系数。可理解，关于公式（13）和公式（14）与公式（2）和公式（3）为相同的公式，当前不再赘述。

基于 3DMM 模型和参数估计网络预测的 3DMM 系数 $\boldsymbol{\alpha}$ 、 $\boldsymbol{\beta}$ 、 $\boldsymbol{\delta}$ ，可以通过上述公式(13)和公式(14)与 3DMM 基底相结合，重建出 3D 人脸的形状和纹理（相当于步骤 231 中的三维人脸形状和三维人脸纹理）。

本实施例使用的表情基底、身份基底和纹理基底与前述实施例使用的表情基底、身份基底和纹理基底的构建方式相同。由于本实施例使用的表情基底可体现 Facewarehouse 模型的 46 个 3D 表情模板偏移量（即表情模板相对于无表情人脸模板（即人脸身份）的形变量）。因此，在参数估计网络对输入人脸图像预测得到身份系数后，我们可以将公式（13）改写为：

$$\mathbf{S}(\boldsymbol{\beta}) = \mathbf{B}_0 + \mathbf{B}_{exp}\boldsymbol{\beta} \quad (15)$$

其中， $\mathbf{B}_0 = \bar{\mathbf{S}} + \mathbf{B}_{id}\boldsymbol{\alpha}$ ，表示输入人脸图像重建的预设标准人脸，即无表情的 3D 人脸，如前述实施例

所述的人脸身份。再进一步，我们可以将公式（15）改写为：

$$S(\beta) = B_0 + \sum_{i=1}^{46} \beta_i (B_i - B_0) \quad (16)$$

其中， B_i 表示第*i*个3D表情模板， $B_i = B_0 + B_{\text{exp}_i}$ 。 B_{exp_i} 表示 B_i 相对于 B_0 的形变量，相当于前述实施例所使用的表情基。即人脸图像重建所使用的3D表情模板结合了相应人脸图像中的预设标准人脸。

由上述3D表情模板构造公式可知，对于任意一个人脸图像，其3D表情模板相对于中性3D人脸的形变都是一样的，均通过FaceWarehouse模型的46个表情基得到。然而，在实际中不同人的同一个表情模板会有一定差异性（即对于不同的人来说，同一表情相对于不同的人脸身份的偏移量存在差异），而基于公式（16）得到的3D表情模板不能建模这种差异性，在拟合人脸表情的时候会损失一些精度，进而影响了表情系数的准确性。所以，本实施例在通过第一参数估计网络模型，计算输入的二维人脸图像对应的第一3DMM参数之前，先对第一参数估计网络模型进行训练。在通过训练好的第二参数估计网络模型，计算二维人脸图像对应的第二3DMM参数之前，先对第二参数估计网络模型进行训练。以提高第一参数估计网络模型和第二参数估计网络模型预测3DMM参数的准确性。

在一实施方式中，上述训练第一参数估计网络模型的操作，可以包括以下步骤：

步骤A1，获取第一训练集；第一训练集包括多个人脸样本图像，每个人脸样本图像对应一组粗糙3DMM参数。

可以直接从现有的图像集（如CelebA名人数据集、Public Figures Face Database哥伦比亚大学公众人物面部数据库、color FERET Database彩色FERET数据库、MTFL dataset多任务面部地标数据集、Voxceleb2名人采访视频数据集等等）中选取若干人脸图像，也可以根据实际需要拍摄多张人脸图像。并可通过试验得到选取的每张人脸图像对应的，能够生成与该人脸样本图像最为近似的三维人脸形状的一组3DMM参数。将选取的人脸图像作为人脸样本图像，将每张人脸样本图像对应的一组3DMM参数作为该人脸样本图像的粗糙3DMM参数，并将该人脸样本图像和粗糙3DMM参数组成样本对，由多个该样本对组成第一训练集，用于对第一参数估计网络模型进行训练，以使得第一参数估计网络模型预测的3DMM系数更加准确。可理解，前述实施例中使用的第一人脸图像也可以作为本实施例中使用的人脸样本图像。需要说明的是，下述对第一参数估计网络模型的训练过程中，所使用的预设第一损失函数不涉及粗糙3DMM参数，即训练过程不需要粗糙3DMM参数的参与，因此，当前也可以不计算粗糙3DMM参数，即获取的第一训练集只包括多个人脸样本图像。

步骤A2，根据第一训练集，对第一参数估计网络模型进行训练。

从第一训练集中选取上述样本对，每个训练周期从训练集中获取的样本对的数目可以为多个。将样本对中的人脸样本图像输入第一参数估计网络模型中，该第一参数估计网络模型可以为任意的深度神经网络，只需估计出输入的人脸图像对应的3DMM参数即可。

一个实施例中，第一参数估计网络模型与前述实施例所述的系数回归网络的功能相同，均是根据包含人脸的二维图像，输出对应的3DMM参数，即前述实施例的系数回归网络可以作为第一参数估计网络，第一参数估计网络输出的3D参数与前述实施例提出的人脸重建系数实质相同，仅是命名上的不同。当前，为了与后续第二参数估计网络输出的3DMM参数进行区分，将训练过程中第一参数估计网络输出的3DMM参数记为第一训练3DMM参数、第二参数估计网络输出的3DMM参数记为第二训练3DMM参数。

在一实施方式中，上述步骤A2具体可以包括以下步骤：步骤A21，将第一训练集中的每个人脸样本

图像输入第一参数估计网络模型中,得到人脸样本图像对应的 3DMM 参数。步骤 A22,通过预设第一损失函数训练第一参数估计网络模型,使人脸样本图像对应的 3DMM 参数等于对应的粗糙 3DMM 参数。

示例性的,预设第一损失函数为:

$$L_{\text{com}} = \lambda_{\text{pho}} L_{\text{pho}} + \lambda_{\text{per}} L_{\text{per}} + \lambda_{\text{im}} L_{\text{im}} + \lambda_{\text{reg}} L_{\text{reg}} + \lambda_{\text{sp}} L_{\text{sp}} \quad (17)$$

其中, L_{pho} 、 L_{per} 、 L_{im} 和 $\{L_{\text{reg}}, L_{\text{sp}}\}$ 分别为通过图像重构损失函数、图像感知损失函数、关键点重构

损失函数和正则损失函数计算的损失值; λ_{pho} 、 λ_{per} 、 λ_{im} 、 λ_{reg} 、 λ_{sp} 均大于 0,分别为相应损失函数的超

参数。可以理解,结合实施例二的相关描述可知,图像重构损失函数、图像感知损失函数、关键点重构损失函数和正则损失函数也可以记为无监督损失函数,即预设第一损失函数为无监督损失函数,预设第一损失函数和无监督损失函数实质上表达的是相同的内容。

示例性的,图像重构的过程中,通常图像重构损失相对于其他几个损失对每个人脸图像的影响较大,相应地, λ_{pho} 的值可以相对较大,明显大于其它几个参数,例如可设置为 1-10 之间的任意数值。而关键点

重构损失函数相对于其他几个损失对每个人脸图像的影响不是很大,相应地, λ_{im} 的值可以相对较小,明显小于其它几个参数,例如可设置为 10-k, k 可大于或等于 3。而和可取值 0-1 之间。举例而言, λ_{pho} ,

λ_{per} 、 λ_{im} 、 λ_{reg} 、 λ_{sp} ,可分别设置为 1.9、0.2、0.1、0.0001 和 0.1。

本实施例使用图像重构损失计算输入图像 I 与相应的渲染图像 $\hat{I}$ (相当于前述实施例中的二维重建图像)之间的像素误差。图像重构损失函数如下面公式 (18) 所示:

$$L_{\text{pho}} = \frac{\sum_{i \in \mathcal{M}} A_i \cdot \|I_i - \hat{I}_i\|_2}{\sum_{i \in \mathcal{M}} A_i} \quad (18)$$

其中, i 表示像素索引, $\mathcal{M}$ 表示 3D 人脸在图像中的脸部渲染区域。 A 表示通过现有的人脸分割算法检测的人脸掩码,人脸掩码可理解为当前像素所在位置位于人脸的概率,当该像素位置是人脸皮肤时其值为 1,反之为 0。公式 (18) 与公式 (4) 为相同的公式,公式 (18) 中所使用的参数以及获取过程可以参考公式 (4) 所在步骤 241 的相关描述。

本实施例利用预训练好的人脸识别网络提取输入图像 I 与相应的渲染图像 $\hat{I}$ 的网络特征,并使用余弦距离来计算两个网络特征之间的相似性,即图像感知损失。该损失定义如下面公式 (19) 所示:

$$L_{\text{per}} = 1 - \frac{\langle f(I), f(\hat{I}) \rangle}{\|f(I)\| \|f(\hat{I})\|} \quad (19)$$

其中 $f(\cdot)$ 代表从人脸识别网络提取的深层特征, $\langle \cdot, \cdot \rangle$ 代表向量内积。公式 (19) 与公式 (5) 为相同的

公式，公式（19）中所使用的参数以及获取过程可以参考公式（5）对应步骤 242-步骤 243 的相关描述。

关键点重构损失定义为从 3D 关键点检测器检测的真实人脸关键点 $\mathbf{Q}$ 与 3D 人脸投影的关键点 $\hat{\mathbf{Q}}$ 之间的均方误差，其中，3D 人脸的关键点是指人脸面部的关键区域位置，包括眉毛、眼睛、鼻子、嘴巴、脸部轮廓等，可以通过顶点索引得到，并通过投影模型投影到 2D 图像平面上（即得到渲染图像），该损失定义如下面公式（20）所示：

$$L_{\text{key}} = \sum_{i=1}^n \omega_i \cdot \|\mathbf{Q}_i - \hat{\mathbf{Q}}_i\|^2 \quad (20)$$

其中， i 代表关键点索引， n 代表人脸关键点的数量，可根据关键点检测器确定，可以为 68、81、106 等， ω_i 代表关键点的权重，本方案将人脸轮廓关键点权重设置为 1，其它人脸关键点权重设置为大于 1 的自然数，例如 10。公式（20）与公式（6）为相同的公式，公式（20）中所使用的参数以及获取过程可以参考公式（6）对应步骤 244-步骤 245 的相关描述。

为防止 3D 人脸的形状和纹理退化，致使得不到有意义的 3D 人脸，本实施例对人脸形状与纹理的系数进行正则损失约束，该约束定义如下面公式（21）所示：

$$L_{\text{reg}} = \sum_{n=1}^N \lambda_{\alpha} \|\alpha_n\|^2 + \lambda_{\varepsilon} \|\varepsilon_n\|^2 \quad (21)$$

其中， λ_{α} 和 λ_{ε} 分别表示系数和的超参数，设置为 1 和 0.001。

此外，本实施例还使用正则损失来促使表情系数的稀疏化表达，该损失定义如下面公式（22）所示：

$$L_{\text{sp}} = \sum_{i=1}^m \lambda_{\beta} \|\beta_i\| \quad (22)$$

其中， m 代表表情模板数（例如 $m=46$ ）， i 代表表情模板的索引， λ_{β} 表示系数 α_n 的超参数， β_i 表示第 i 个 3D 表情模板。公式（21）和公式（22）与公式（7）和公式（8）为相同的公式，公式（20）中所使用的参数以及获取过程可以参考公式（6）对应步骤 246-步骤 248 的相关描述。

本实施例为了对参数估计网络进行训练，本实施例将带有纹理的 3D 人脸渲染到图像平面上，该过程引入 3DMM 模型的光照系数与位姿系数。重建的 3D 人脸纹理进一步经过球谐(Spherical Harmonics)光照模型处理（结合光照系数），建模人脸图像的环境光照。为了将 3D 人脸投影到图像平面，本实施例采用透视投影相机模型（结合位姿系数）。最后，我们将经过光照处理的 3D 人脸通过投影模型渲染到 2D 图像上，以得到渲染到图像平面的渲染图像，参与到网络的训练当中。即通过所述第一训练 3DMM 参数得到所述渲染图像的过程包括：将第一训练 3DMM 参数引入预设的 3DMM 模型，所述 3DMM 模型包括人脸形状和人脸纹理；将 3DMM 模型中的人脸纹理经过球谐光照模型处理，所述球谐光照模型处理过程中结合光照系数，以建模环境光照；将经过球谐光照模型处理后的 3DMM 模型经过透视投影相机模型，渲染到图像平面上，以得到渲染图像，所述透视投影相机模型处理过程中结合位姿系数。其中，渲染图像的渲染过程与前述实施例二维重建图像的渲染过程相同，当前不另做描述。可理解，预设第一损失函数中所提及的渲染图像均可理解为基于第一训练 3DMM 参数得到的渲染图像。

基于第一参数估计网络模型的训练过程（即步骤 A1-步骤 A2）可知，训练第一参数估计网络模型时，

通过预设第一损失函数调整第一参数估计网络模型的网络参数,以使第一参数估计网络模型输出的第一训练 3DMM 参数尽可能准确,而粗糙 3DMM 参数仅解释了第一参数估计网络模型训练完成后的效果(即第一训练 3DMM 参数等于对应的粗糙 3DMM 参数),而非训练所需的必要数据。一般对于神经网络而言,训练停止条件通常是达到一定的训练次数或者是所使用的损失函数收敛,即第一参数估计网络模型的训练停止条件可以与粗糙 3DMM 参数无关,因此,没有粗糙 3DMM 参数也可以完成对第一参数估计网络模型的训练。此时,对于上述第一参数估计网络模型来说,训练过程也可以表述为:获取第一训练集,所述第一训练集包括多个人脸样本图像;通过第一参数估计网络模型得到所述人脸样本图像对应的第一训练 3DMM 参数,所述第一训练 3DMM 参数包括身份系数、表情系数、纹理系数、位姿系数和光照系数;通过预设第一损失函数训练第一参数估计网络模型,所述预设第一损失函数包括图像重构损失函数、图像感知损失函数、关键点损失函数和正则损失函数,所述预设第一损失函数通过所述人脸样本图像、所述人脸样本图像对应的渲染图像和所述第一训练 3DMM 参数得到,所述渲染图像是将所述第一训练 3DMM 参数引入预设的 3DMM 模型后渲染到图像平面而得到。举例而言,先获取包含多张人脸样本图像,将人脸样本图像输入至第一参数估计网络模型,第一参数估计网络模型根据人脸样本图像输出第一训练 3DMM 参数,之后,将第一训练 3DMM 参数应用于 3DMM 模型,并结合光照系数、位姿系数,使用球谐光照模型和透视投影相机模型,渲染得到二维的渲染图像,该渲染图像可认为是人脸样本图像的二维的重建图像。之后,结合渲染图像和人脸样本图像构建图像重构损失函数、图像感知损失函数、关键点重构损失函数,根据 3DMM 参数构建正则损失函数,以得到预设第一损失函数,并基于预设第一损失函数的具体值对第一参数估计网络模型的网络参数进行调整,之后,再次重复上述过程,直到预设第一损失函数收敛或者训练次数达到一定次数(即到达一定的训练)。训练结束后,得到的第一参数估计网络模型输出的 3DMM 参数可以认为是等于粗糙 3DMM 参数。不使用粗糙 3DMM 参数时,可以认为不使用先验信息,这样会降低对人脸样本图像的要求,不需要额外得到粗糙 3DMM,使得第一参数估计网络模型的训练过程更容易开展。

可理解,第一参数估计网络模型的训练过程与前述实施例系数回归网络的训练过程区别在于,系数回归网络训练过程中,系数回归网络构建的损失函数包括多帧第一人脸图像的无监督训练损失函数和身份约束损失函数,还可包括表情互斥约束损失函数,第一参数估计网络模型训练过程中构建的损失函数包括一帧人脸样本图像的无监督训练损失函数,其他的环节均相同,当前不另做描述。

在另一实施方式中,上述训练第二参数估计网络模型的操作,可以包括以下步骤:

步骤 B1,获取第二训练集,第二训练集包括多个人脸样本图像和每个人脸样本图像对应的粗糙三维人脸样本形状(即粗糙样本 3D 表情模板)、精细 3DMM 参数。

第二训练集中的人脸样本图像和对应的精细 3DMM 参数的获取方式可以参照第一训练集中的获取方式,在此不再赘述。第二训练集中的粗糙三维人脸样本形状(即粗糙样本 3D 表情模板)即,将选取的人脸样本图像输入系数为对应的粗糙 3DMM 参数的 3DMM 模型中输出的粗糙三维人脸样本形状(即粗糙样本 3D 表情模板)。然后将选组的人脸样本图像和对应的精细 3DMM 参数及粗糙三维人脸样本形状组成样本组,多个样本组构成该第二训练集,用于对第二参数估计网络模型进行训练,以使得第二参数估计网络模型预测的精细 3DMM 参数更加准确。

可理解,第一训练集中的人脸样本图像或前述实施例使用的第一人脸图像也可以作为第二训练集中的人脸样本图像。由公式(16)可知,3D 表情模板可表示为 $\mathbf{B}_i = \mathbf{B}_0 + \mathbf{B}_{exp_i}$, 且 $\mathbf{B}_0 = \bar{\mathbf{S}} + \mathbf{B}_{id} \alpha$, 且步骤 S2

指出：粗糙三维人脸形状（即粗糙 3D 表情模板）通过训练好的第一参数估计网络输出的 3DMM 参数和 3DMM 模型确定，基于此可知，将第二训练集中的人脸样本图像输入至训练好的第一参数估计网络模型后，将输出的第一训练 3DMM 参数中的身份系数 α 代入 $\bar{S} + B_{id}\alpha$ 可以得到 B_G ，之后，通过 $B_G + B_{exp_i}$ 得到的 B_i 可以认为是第 i 个表情对应的粗糙样本 3D 表情模板，粗糙样本 3D 表情模板可以理解是训练过程中构造的带有人脸身份的粗糙表情模状。每个表情均有对应的粗糙样本 3D 表情模板，46 个表情对应的粗糙样本 3D 表情模板组成了人脸样本图像中人脸的一套三维表情模板。此时，获取第二训练集可包括：获取多个人脸样本图像；将每个人脸样本图像分别输入至训练好的第一参数估计网络模型，得到每个人脸样本图像对应的第一训练 3DMM 参数；基于第一训练 3DMM 参数和预设的 3DMM 模型确定人脸样本图像对应的粗糙样本 3D 表情模板。其中，基于所述第一训练 3DMM 参数和预设的 3DMM 模型确定所述人脸样本图像对应的粗糙样本 3D 表情模板包括：将所述第一训练 3DMM 参数中的身份系数应用于表情基底，并结合 3DMM 模型的平均人脸的形状，得到所述人脸样本图像重建的预设标准人脸；根据预设标准人脸和表情基底的叠加结果，得到粗糙样本 3D 表情模板。

当前第二训练集包括精细 3DMM 参数，精细 3DMM 参数与前述的粗糙 3DMM 的作用相同，仅用于解释第二参数估计网络模型的训练效果，即训练过程不需要精细 3DMM 参数的参与，因此，当前也可以不计算精细 3DMM 参数，即获取的第二训练集可以只包括多个人脸样本图像和其对应的粗糙样本 3D 表情模板。

步骤 B2，根据第二训练集，对第二参数估计网络模型进行训练。

在一个实施方式中（即实施方式一），步骤 B2 包括：基于粗糙样本 3D 表情模板生成精细样本 3D 表情模板，可理解，精细样本 3D 表情模板可以认为是人脸样本图像中人脸所能出现的表情（这里指 46 个表情）的个性化描述，比通用的表情基或粗糙样本 3D 表情模板更准确体现人脸样本图像中人脸的表情。每个粗糙样本 3D 表情模板均有对应的精细样本 3D 表情模板，精细样本 3D 表情模板的生成方式是计算精细样本 3D 表情模板相对于粗糙样本 3D 表情模板的形变（当前，将该形变记为三维表情样本形变）并添加到粗糙样本 3D 表情模板中。训练第二参数估计网络模型时，第二参数估计网络模型输出第二训练 3DMM 参数后，基于精细样本 3D 表情模板和第二训练 3DMM 参数重建人脸样本图像对应的三维人脸（即在 3DMM 模型中使用精细样本 3D 表情模板），之后，基于重建的三维人脸得到二维的渲染图像，并基于渲染图像和人脸样本图像构建当前训练所需的损失函数。可理解，由于当前渲染图像使用的精细样本 3D 表情模板，因此，通过渲染图像构建的损失函数可以是第二参数估计网络学习到精细样本 3D 表情模板，进而保证了第二参数估计网络输出的表情系数的准确性。

当前，三维表情样本形变的确定方式可以根据实际情况设定，例如，构建并训练一个神经网络，将粗糙样本 3D 表情模板输入至该神经网络后，可以输出三维表情样本形变，以在训练第二参数估计网络模型时，通过该神经网络得到的三维表情样本形变生成精细样本 3D 表情模板。再如，构建并训练一个神经网络，该神经网络用于输出粗糙样本 3D 表情模板对应的精细样本 3D 表情模板。还如，只构建一个神经网络，将粗糙样本 3D 表情模板输入至该神经网络可以得到三维表情样本形变，将该神经网络与第二参数估计网络模型一同训练，训练后，神经网络输出准确的三维表情样本形变。

当前，以构建用于输出三维表情样本形变的神经网络且该神经网络与第二参数估计网络模型一同训练为例。将神经网络记为映射网络，此时，方法还包括：通过映射网络得到粗糙样本 3D 表情模板对应的三

维表情样本形变；根据三维表情样本形变和粗糙样本 3D 表情模板得到精细样本 3D 表情模板。该映射网络为卷积神经网络，可理解，粗糙样本 3D 表情模板位于三维空间中，这样不利于卷积神经网络的学习，因此，一个实施例中，将粗糙样本 3D 表情模板映射到一个二维空间中，以得到二维的图像，然后使映射网络对二维的图像进行学习。可选的，使用 UV 映射，将粗糙样本 3D 表情模板映射到 2D 的 UV 空间中，供映射网络（如卷积神经网络）进行学习，以使映射网络输出 UV 空间中精细样本 3D 表情模板相对于粗糙样本 3D 表情模板的形变量，学习完再从 UV 空间映射回三维空间，以得到三维表情样本形变。当前，UV 空间为二维的空间，UV 空间中的图为二维的 UV 图，UV 空间中 U 坐标代表二维的 UV 图的横轴，V 表示纵轴。需要说明，人脸样本图像和对应的渲染图像也处于二维空间，但是其处于的二维空间为 X,Y 坐标系所在的空间，而非 U,V 坐标系所在的空间。此时，通过映射网络得到粗糙样本 3D 表情模板对应的三维表情样本形变包括：将粗糙样本 3D 表情模板映射到 UV 空间，得到二维样本 UV 图；将二维样本 UV 图输入至映射网络，得到二维表情样本形变；将二维表情样本形变映射到粗糙样本 3D 表情模板所在的三维空间，得到三维表情样本形变。其中，映射时，将粗糙样本 3D 表情模板中每个顶点（即三角网格的顶点）映射到 UV 空间中，得到二维的 UV 坐标，这些坐标组成的 UV 图记为二维样本 UV 图，该图中每个顶点仍然保留其在三维空间中的信息（如顶点的三维坐标等信息），只是为了便于映射网络的学习，将三维空间中的信息暂时用二维样本 UV 图的形式存储。之后，将二维样本 UV 图输入至映射网络，映射网络基于二维样本 UV 图得到 UV 空间中精细样本 3D 表情模板相对于粗糙样本 3D 表情模板的形变量，当前，将该形变量记为二维表情样本形变。可理解，由于二维样本 UV 图保留了三维空间中的信息，映射网络从二维样本 UV 图中学习三维空间中的信息，因此，其输出的二维表情样本形变也包含了三维的形变信息，只是其在 UV 空间中通过 UV 图的形式体现，即二维表情样本形变记录有 UV 空间中的三维的形变信息。之后，将二维表情样本形变由 UV 空间逆映射到三维空间中，以得到三维表情样本形变。之后，将三维表情样本形变添加到粗糙样本 3D 表情模板便可以得到精细样本 3D 表情模板。可定义从 UV 空间映射到三维空间的函数为 F 。基于精细 3D 表情模板重建三维人脸时，公式 (16) 可变换为下述公式：

$$\mathbf{S} = \mathbf{B}_0 + \sum_{i=1}^{46} \beta_i (\mathbf{B}_i - \mathbf{B}_0 + F(\mathbf{A}_i \Delta_i)) \quad (23)$$

其中， $F(\mathbf{A}_i \Delta_i)$ 表示第 i 个表情（共有 46 个表情）的三维表情样本形变。基于前述内容可知，粗糙样本 3D 表情模板 $\mathbf{B}_i = \mathbf{B}_0 + \mathbf{B}_{exp_i}$ 。有公式 (23) 可知，精细样本 3D 表情模板为 $\mathbf{B}_i' = \mathbf{B}_i + F(\mathbf{A}_i \Delta_i)$ ，相对于粗糙三维人脸形状 $\mathbf{B}_i$ ，增加了三维表情样本形变 $F(\mathbf{A}_i \Delta_i)$ 。其中， Δ_i 为通过映射网络确定的二维表情样本形变， $\mathbf{A}_i$ 是第 i 个表情的注意力掩码。

可以理解的是，对于当前使用的 46 个表情来说，其 3D 表情模板相对于中性人脸的形变应该是局部的，例如，具有闭眼意义的 3D 表情模板对应于中性人脸的形变应该集中在眼部区域。而映射网络是基于整个面部进行学习，输出的二维表情样本形变对应于整个面部的形变，因此，为每个表情构建一个注意力掩码，该注意力掩码可体现基于对应表情产生的形变所集中的局部区域，将注意力掩码应用在二维表情样本形变后，可以使得二维表情样本形变只关注与当前表情有关的局部区域。一个实施例中，注意力掩码的生成过程为：计算粗糙样本 3D 表情模板对于预设标准人脸的形变中每一个顶点的欧式距离，预设标准人

脸通过身份系数得到；根据欧式距离，形成注意力掩码，并将注意力掩码应用于二维表情样本形变，注意力掩码大于等于 0 且小于等于 1。

示例来说，预设标准人脸是指中性人脸，当前，对应于人脸样本图像，预设标准人脸是人脸样本图像中人脸对应的中性人脸，如公式（15）、公式（16）以及公式（23）中的 B_0 均为预设标准人脸，预设标准人脸通过 3DMM 参数中的身份系数得到。可理解，对预设标准人脸进行形变是指改变其中相应顶点的位置，为了便于理解，为每个顶点设置一个索引或编号，若形变后得到粗糙样本 3D 表情模板，那么，粗糙样本 3D 表情模板中各顶点在预设标准人脸中均存在形变前的顶点，这两个顶点具有相同的索引或编号，粗糙样本 3D 表情模板对于预设标准人脸的形变中每一个顶点的欧式距离，便是粗糙样本 3D 表情模板和预设标准人脸中索引或编号相同的顶点的欧式距离，其可以通过三维坐标计算得到。可理解，欧式距离越大，说明顶点所在区域的形变越大。此时，通过欧式距离可以体现出粗糙样本 3D 表情模板对于预设标准人脸的形变量（即 $B_i - B_0$ ）。每个顶点均有对应的欧式距离。之后，根据欧式距离可以确定形变较大的局部区域，进而生成只关注该局部区域的注意力掩码。此时，根据欧式距离，形成注意力掩码可包括：若欧式距离大于或等于设定的阈值，则保留欧式距离，若欧式距离小于设定的阈值，则将欧式距离置为零；将各顶点的欧式距离进行归一化，形成注意力掩码。可理解，由于欧式距离可以体现对应顶点的形变量，因此，将各欧式距离与设定的阈值进行比较，该阈值可以根据实际需求设置，例如，设置为 0.001。欧式距离小于该阈值时，可以认为对应顶点所在的区域中，粗糙样本 3D 表情模板对于预设标准人脸的形变很小，与相应的表情无关，因此，该顶点对应的欧式距离设置为 0。欧式距离大于或等于该阈值时，可以认为对应顶点所在的区域中，粗糙样本 3D 表情模板对于预设标准人脸的形变不小，与相应的表情有关，因此，保留该欧式距离，即保留该形变。全部欧式距离均与阈值比较后，将处理后的各欧式距离进行归一化处理，以将欧式距离变化到 0-1 的范围，归一化后的数值便可以作为对应顶点的注意力掩码，其可以表示该顶点所在区域被关注的程度，数值越高，关注程度越大。之后，按照各顶点的索引或编号对归一化后的数值进行组合，形成的矩阵或向量便是当前使用的注意力掩码，即公式（23）中的 A_i 。

得到注意力掩码后，将其作用于二维表情样本形变，即得到公式（23）中的 $A_i \Delta_i$ ，进而基于 $A_i \Delta_i$ 得到精细样本 3D 表情模板相对于粗糙样本 3D 表情模板的形变主要约束在粗糙样本 3D 表情模板中与表情有关的局部区域。

得到精细样本 3D 表情模板后，便可以基于精细样本 3D 表情模板和第二参数估计网络模型输出的第二训练 3DMM 参数，结合 3DMM 模型，得到人脸样本图像对应的渲染图像。其中，得到渲染图像的过程与通过第一训练 3DMM 参数得到渲染图像的过程相同，当前不作赘述。之后，结合渲染图像和人脸样本图像构建训练所需的预设第二预设损失函数。一个实施例中，训练第二参数估计网络模型时，还同时训练映射网络，即通过所述预设第二损失函数训练所述映射网络，以使映射网络输出准确的二维样本表情形变。此时，映射网络和第二参数估计网络模型共用预设第二损失函数。一个实施例中，预设第二损失函数包括预设第一损失函数（也可记为无监督训练损失函数）。另一个实施例中，预设第二损失函数包括预设第一损失函数（也可记为无监督训练损失函数）以及表情梯度损失函数。其中，表情梯度损失函数通过精细样本 3D 表情模板和粗糙样本 3D 表情模板构建，用于使精细样本 3D 表情模板和粗糙样本 3D 表情模板的梯度相接近，目的在于使用注意力掩码后，防止精细样本 3D 表情模板对应的语义（即对应的表情的语义）发生改变，举例而言，渲染图像中的人脸参考了精细样本 3D 表情模板，精细样本 3D 表情模板与三维表情

样本形变有关,若三维表情样本形变不准确(如开始训练时,因为学习次数较少,基于映射网络输出的二维表情样本形变准确度不高,进而导致三维表情样本形变准确度不高),则会在学习过程中,使映射网络学习到错误的语义,如粗糙样本 3D 表情模板的语义是闭眼,通过注意力掩码将三维表情样本形变关注在眼部区域,而当前三维表情样本形变更倾向睁眼的语义(如闭眼时上眼皮的形变应向下,当前三维表情样本形变中上眼皮并未向下),这样在训练过程中便会学习到错误的语义,进而导致精细样本 3D 表情模板的语义发生改变,使用表情梯度损失函数后,通过梯度约束三维表情样本形变(如闭眼语义下,通过梯度约束上眼皮的形变应向下),就可以避免精细样本 3D 表情模板的语义发生改变。可理解,46 个表情对应的精细样本 3D 表情模板组成了人脸样本图像中人脸的另一套三维表情模板。可理解,第二预设损失函数包括第一预设损失函数是指在第二预设损失函数中使用了第一预设损失函数的计算公式。

除了上述实施方式外,还可以从第二训练集中选取上述样本组,每个训练周期从训练集中获取的样本组的数目可以为多个。将样本组中的人脸样本图像和粗糙三维人脸样本形状(即粗糙样本 3D 表情模板)输入第二参数估计网络模型中,该第二参数估计网络模型可以为任意能够通过深度学习,估计出输入的人脸图像和粗糙三维人脸样本形状(即粗糙样本 3D 表情模板)对应的精细 3DMM 参数即可。

此时,在另一实施方式中(即实施方式二),上述步骤 B2 具体可以包括以下步骤:

步骤 B21,确定第二训练集中的每个人脸样本图像对应的粗糙三维人脸样本形状(即粗糙样本 3D 表情模板)相对于预设标准人脸的三维表情样本形变,预设标准人脸根据二维人脸图像自 3DMM 模型的预设标准人脸集中选择。

本实施例为了生成精细化的三维人脸形状(即精细样本 3D 表情模板),在粗糙三维人脸形状(即粗糙样本 3D 表情模板)的基础上针对不同人学习个性化表情形变。首先通过 UV 映射将粗糙三维人脸样本形状(即粗糙样本 3D 表情模板)映射到 UV 空间,得到粗糙三维人脸样本形状(即粗糙样本 3D 表情模板)对应的二维样本 UV 图。然后确定二维样本 UV 图相对于预设标准人脸的二维 UV 图的二维表情样本形变,并根据二维表情样本形变,通过映射网络确定粗糙三维人脸样本形状(即粗糙样本 3D 表情模板)相对于预设标准人脸的三维表情样本形变。即当前的映射网络通过学习粗糙样本 3D 表情模板相对于预设标准人脸的形变(该形变位于 UV 空间中),得到用于生成精细样本 3D 表情模板的三维表情样本形变(位于三维空间中)。与前述实施方式的区别在于,当前的二维表情样本形变是粗糙样本 3D 表情模板相对于预设标准人脸的形变,其供映射网络进行学习,以输出用于生成精细样本 3D 表情模板的三维表情样本形变,前述实施方式的二维表情样本形变是精细样本 3D 表情模板相对于粗糙样本 3D 表情模板的形变,其为映射网络的输出结果。可理解,由于映射网络为卷积神经网络,因此,通过为映射网络设置不同的任务,可以实现不同的功能。

确定二维样本 UV 图相对于预设标准人脸的二维 UV 图的二维表情样本形变之后,还可以计算粗糙三维人脸形状(即粗糙样本 3D 表情模板)对于预设标准人脸的形变 $\mathbf{B}_i - \mathbf{B}_0$ 每一个顶点的欧式距离,并设定一个阈值 0.001 将低于该阈值的顶点位置设置为 0,再将这个形变值归一化到 0-1 范围,作为注意力掩码 A_i (注意力掩码大于等于 0,且小于等于 1)。注意力掩码 A_i 反映二维表情形变局部区域重要性,可以将待学习的个性化二维表情形变约束在跟粗糙三维形状(即粗糙样本 3D 表情模板)相似的局部区域,因此,可将注意力掩码 A_i 加到个性化二维表情形变 Δ_i 上,定义从 UV 空间映射到 3D 空间的函数为 F ,则三维表

情样本形变可表示为 $F(A_i \Delta_i)$ 。可理解,将注意力掩码作用在二维表情样本形变后,基于当前使用的函数 F ,可以得到用于生成精细样本3D表情模板的三维表情样本形变。

步骤B22,通过预设第二损失函数对第二参数估计网络模型进行训练,使基于人脸样本图像和对应的粗糙三维人脸形状(即粗糙样本3D表情模板)得到的第二训练3DMM参数等于对应的精细3DMM参数。

为了更好地利用神经网络进行训练,可首先将粗糙三维人脸形状映射到2D的UV空间中进行学习,学习完再从UV空间映射回3D空间。可定义从UV空间映射到3D空间的函数为 F 。则通过公式(23)可知精细三维人脸形状的表达式:

$$S = B_0 + \sum_i^k \beta_i (B_i - B_0 + F(A_i \Delta_i)) \quad (23)$$

其中, $F(A_i \Delta_i)$ 表示第 i 个表情的三维表情样本形变,有公式(23)可知,经过精细化的三维人脸形状(即精细3D表情模板) $B'_i = B_i + F(A_i \Delta_i)$,相对于粗糙三维人脸形状(即粗糙3D表情模板) B_i ,添加了个性化的表情形变。

两种实施方式下,计算第二训练3DMM参数的过程中,与计算第一训练3DMM参数阶段相比,除了3D人脸的形状表达不一样外,其他3DMM模型的表达(例如纹理模型、光照模型、投影模型)均相同。并且,渲染图像的生成方式、预设第一损失函数的计算方式均相同。

在精细的三维人脸形状(即精细3D表情模板)生成阶段,同样可以采用粗糙三维人脸形状生成(即粗糙3D表情模板)阶段的训练方式,通过在2D图像空间进行自监督训练范式,来监督第二参数估计网络和映射网络的训练。

示例性的,为了防止加了个性化表情形变后会改变原来表情模板的语义信息,还引入了一个额外的表情模板梯度损失。因此,上述预设第二损失函数的定义可如下面公式(24)所示:

$$L = L_{com} + \lambda_{gra} L_{gra} \quad (24)$$

其中, L_{com} 为预设第一损失函数, L_{gra} 为表情梯度损失函数, λ_{gra} 为表情梯度损失函数的超参数。

如果直接向粗糙三维人脸形状(即粗糙样本3D表情模板)添加自由的形变,即使在使用注意力掩码将其限制在局部区域之后,也有可能可以改变人脸表情的语义。为了进行人脸表情驱动,不同的人脸表情应该具有相同的语义信息。因此,可使用表情模板梯度损失促使变形后的精细化的三维人脸形状(即精细3D表情模板)的梯度和粗糙三维人脸形状(即粗糙3D表情模板)的梯度相接近,该表情梯度损失函数定义如下面公式(25)所示:

$$L_{gra} = \|G_{s_a \rightarrow (s_i + \Delta i)} - G_{s_a \rightarrow s_i}\|^2 \quad (25)$$

其中, $G_{a \rightarrow b}$ 表示同一张人脸图像的精细三维人脸图像 b 对于粗糙三维人脸图像 a 的梯度。 $G_{s_a \rightarrow (s_i + \Delta i)}$ 可以认为是精细样本3D人脸模板相对于预设标准人脸的梯度, $G_{s_a \rightarrow s_i}$ 可以认为是粗糙样本3D人脸模板相对于预设标准人脸的梯度,通过L2范数对两个梯度进行约束,可以使得精细样本3D表情模板的梯度和粗糙

样本 3D 表情模板的梯度相接近。

需要说明的是，第一参数估计网络模型和第二参数估计网络可以是结构相同的两个模型。并且，由于第二参数估计网络模型基于第一参数估计网络模型得到的粗糙人脸形状进行训练，可以确定第二参数估计网络模型预测的身份系数与第一参数估计网络模型预测的身份系数相同，所以，在第二参数估计网络模型训练过程中，可仅训练表情系数、纹理（色彩亮度）系数、光照系数及头部位姿系数。即第二参数估计网络模型训练过程中使用第一参数估计网络模型输出的身份系数。

基于第二参数估计网络模型的训练过程（即步骤 B1-步骤 B2）可知，训练第二参数估计网络模型时的过程可以表述为：获取第二训练集，第二训练集包括多个人脸样本图像和每个人脸样本图像对应的粗糙样本 3D 表情模板，粗糙样本 3D 表情模板过训练好的第一参数估计网络模型输出的人脸样本图像的 3DMM 参数构建；通过第二参数估计网络模型得到人脸样本图像的 3DMM 参数；通过预设第二损失函数训练第二参数估计网络模型，预设第二损失函数包括：无监督训练损失函数（即预设第一损失函数）和表情梯度损失函数，预设第二损失函数通过人脸样本图像、重建后的渲染图像、粗糙样本 3D 表情模板和相应的精细样本 3D 表情模板构建，渲染图像通过第二参数估计网络输出的 3DMM 参数得到，精细样本 3D 表情模板通过样本 3D 表情模板得到。并且，精细样本 3D 表情模板可以基于映射网络实现，映射网络和第二参数估计网络一同训练，共用预设第二损失函数。

通过上述方式训练得到上述训练好的第一参数估计网络模型和第二参数估计网络模型后，通过以下步骤利用该第一参数估计网络模型和第二参数估计网络模型来生成三维人脸形状（这里为 3D 表情模板），以下步骤使用的第二参数估计网络模型的训练方式是实施方式二，参考图 8，生成 3D 表情模板可包括：

步骤 S1，通过训练好的第一参数估计网络模型，计算输入的二维人脸图像对应的第一 3DMM 参数。

该三维人脸形状生成方法的执行主体可以为服务器，服务器接收到终端输入的任意一张二维人脸图像，可通过上述训练好的第一参数估计网络模型，计算输入的二维人脸图像对应的第一 3DMM 参数，具体计算过程可参照上述第一参数估计网络模型的训练过程，在此不再赘述。当前，第一 3DMM 参数包括身份系数、表情系数、纹理系数、光照系数和头部位姿系数。二维人脸图像是应用模型应用过程中使用的图像。

步骤 S2，基于第一 3DMM 参数和预设的 3DMM 模型确定二维人脸图像对应的粗糙三维人脸形状（即粗糙 3D 表情模板）。

服务器在计算出第一 3DMM 参数之后，可将计算出的第一 3DMM 参数赋值于预设的 3DMM 模型，形成输入的二维人脸图像对应的特定粗糙 3DMM 模型，将该二维人脸图像输入至该特定粗糙 3DMM 模型，便可输出该二维人脸图像对应的粗糙三维人脸形状（即粗糙 3D 表情模板）。即将身份系数代入公式（15）或公式（16）中，计算出的 B_i 便是第 i 个粗糙 3D 表情模板。

步骤 S3，根据二维人脸图像和粗糙三维人脸形状（即粗糙 3D 表情模板），通过训练好的第二参数估计网络模型，计算二维人脸图像对应的第二 3DMM 参数。

服务器在生成粗糙三维人脸形状（即粗糙 3D 表情模板）之后，可先确定粗糙三维人脸形状（即粗糙 3D 表情模板）相对于预设标准人脸的三维表情形变，其中，预设标准人脸根据二维人脸图像自 3DMM 模型的预设标准人脸集中选择。

其中，为了减轻个性化三维表情形变学习的难度，可通过 UV 映射将粗糙三维人脸形状（即粗糙 3D 表情模板）映射到二维的 UV 空间，利用卷积神经网络学习在 UV 空间的精细化三维表情形变，即先得到

粗糙三维人脸形状对应的粗糙二维 UV 图，然后，确定粗糙二维 UV 图相对于预设标准人脸的二维 UV 图的二维表情形变；根据二维表情形变，通过映射网络确定粗糙三维人脸形状相对于预设标准人脸的三维表情形变。

在一实施方式中，确定粗糙二维 UV 图相对于预设标准人脸的二维 UV 图的二维表情形变之后，还可以包括：计算二维表情形变每一个顶点的欧式距离，根据欧式距离，形成二维表情形变的注意力掩码，注意力掩码大于等于 0，且小于等于 1。

需要说明的是，上述通过 UV 空间确定三维表情形变的方法只是本实施例的较佳实施方式，本实施例并不以此为限，例如，也可以直接利用其他 3D 网络在 3D 空间进行个性化表情模板形变的学习。

服务器确定粗糙三维人脸形状相对于预设标准人脸的三维表情形变之后，通过训练好的第二参数估计网络模型，计算二维人脸图像对应的第二 3DMM 参数。具体计算过程可参照上述第二参数估计网络模型的训练过程，在此不再赘述。

步骤 S4，基于第二 3DMM 参数和预设的 3DMM 模型确定二维人脸图像对应的精细三维人脸形状（即精细 3D 表情模板）。

服务器在计算出上述第二 3DMM 参数之后，可将计算出的第二 3DMM 参数赋值于预设的 3DMM 模型，形成输入的二维人脸图像对应的特定精细 3DMM 模型，将该二维人脸图像输入至该特定精细 3DMM 模型，便可输出该二维人脸图像对应的精细三维人脸形状。即通过公式（23）可以构建精细 3D 表情模板。

通过上述方式训练得到上述训练好的第一参数估计网络模型和第二参数估计网络模型后，还可以通过以下步骤利用该第一参数估计网络模型和第二参数估计网络模型来生成 3D 表情模板，以下步骤使用的第二参数估计网络模型的训练方式是实施方式一，参考图 9，生成 3D 表情模板可包括：

步骤 S5、通过训练好的第一参数估计网络，计算输入的二维人脸图像对应的第一 3DMM 参数。

步骤 S6、基于第一 3DMM 参数和预设的 3DMM 模型，得到二维人脸图像对应的粗糙三维人脸形状。

步骤 S7、根据粗糙三维人脸形状，通过训练好的映射网络，得到三维表情形变。

步骤 S8、根据粗糙三维人脸形状和三维表情形变得到二维人脸图像对应的精细三维人脸形状。

即通过训练好的第一参数估计网络，计算二维人脸图像对应的 3DMM 参数，当前记为第一 3DMM 参数，之后，将第一 3DMM 参数中的身份系数代入公式（15）或公式（16）中，计算出 B_i ，即粗糙 3D 表情模板。之后，将粗糙 3D 表情模板映射到 UV 空间后，输入至映射网络，之后，将映射网络输出的二维表情形变（即映射网络应用过程中输出的形变）加上计算得到的注意力掩码后映射至三维空间中，以得到三维空间中精细 3D 表情模板相对于粗糙 3D 表情模板的形变（这里记为三维表情形变），之后，将三维表情形变应用在粗糙 3D 表情模板上，以得到精细 3D 表情模板。之后，基于精细 3D 表情模板可以得到适用于对应人脸的表情基，进而基于表情基进行表情拟合时，得到更精准的表情系数。

通过上述方式训练得到上述训练好的第一参数估计网络模型和第二参数估计网络模型后，还可以通过第二参数估计网络模型输出的第二 3DMM 参数，重建出二维人脸图像对应的虚拟图像，例如，获取第二 3DMM 参数中的表情系数和位姿系数，之后，将表情系数和位姿系数应用到的虚拟图像中，以得到与二维人脸图像表情和头部动作一致的虚拟图像中。

在一实施方式中，构建粗糙 3D 表情模板时，同时构建 46 个表情对应的粗糙 3D 表情模板；构建精细 3D 表情模板时，同时构建 46 个表情对应的精细 3D 表情模板。通过该方法生成表情模板，对于不同的人，生成不同的表情模板，充分考虑了每个人在相同表情下的特异性，可以提高 3D 人脸重建和表情拟合的准

确性，从而加强了人脸驱动的效果，可使得人脸驱动过程更加灵敏。

需要说明的是，上述各实施方式中应用 FaceWarehouse 数据集、Basel Face Model 数据集、CelebA 名人数据集及 Voxceleb2 名人采访视频数据集仅用于对该方法进行详细说明，并不是对本实施例的限定，该方法具体实施过程中也可以采用其它数据集，只要能实现该方法就行。

为了便于理解本申请实施例提供的方法，下面结合附图进行说明。如图 10 和图 11 所示，服务器接收到输入的二维人脸图像之后，可以通过训练好的第一参数估计网络，预测该二维人脸图像的粗糙 3DMM 系数（包括身份系数、表情系数、纹理系数、光照系数及头部位姿系数），然后基于该粗糙 3DMM 系数和预设的 3DMM 模型，可生成该二维人脸图像对应的 46 个粗糙 3D 表情模板，可重建得到该二维人脸图像的粗糙的 3D 人脸形状（将重建得到 3D 人脸渲染到输入的二维人脸图像上，通过多个自监督训练损失函数可监督第一参数估计网络的训练），3D 人脸形状包括人脸形状和人脸纹理。然后，可将粗糙三维表情模板映射到 UV 空间后输出至映射网络，得到二维表情形变再逆映射至三维空间，以得到三维表情形变，将三维表情形变应用于粗糙 3D 表情模板，可以得到 46 个粗糙 3D 表情模板对应的 46 个精细 3D 表情模板。将原始输入的二维人脸图像输入到第二参数估计网络，预测该二维人脸图像的精细 3DMM 系数（包括表情系数、纹理系数、光照系数及头部位姿系数），然后基于该精细 3DMM 系数和精细 3D 表情模板，可重建得到该二维人脸图像的精细化的 3D 人脸形状（将重建得到 3D 人脸渲染到输入的二维人脸图像上，通过多个自监督训练损失函数可监督第二参数估计网络的训练，上述精细三维表情模板可用于训练映射网络）。

本实施例提供的方法，先通过第一参数估计网络模型，计算输入的二维人脸图像对应的第一 3DMM 参数，并基于第一 3DMM 参数和预设的 3DMM 模型确定二维人脸图像对应的粗糙 3D 表情模板，再根据粗糙 3D 表情模板和映射网络，得到二维人脸图像对应的精细三维人脸形状，如此，从粗糙到精细通过的两阶段生成个性化的三维人脸形状，关注了不同人的个性化 3D 人脸形状的构建，充分考虑了每个人在相同表情下的特异性，从而会提高 3D 人脸重建和表情拟合的准确性，加强了人脸驱动的效果。

需说明，本实施例涉及的参数估计网络模型的相关说明均可适用于第一参数估计网络模型和第二参数估计网络模型。

可理解，实施例四与实施例一至三可以结合使用，以进一步提供表情重建的准确性。一个实施例中，获取包括多个人脸样本图像的第一训练集，通过第一参数估计网络模型得到人脸样本图像对应的第一训练 3DMM 参数，之后，通过预设第一损失函数训练第一参数估计网络模型（即基于第一训练 3DMM 参数重建得到人脸样本图像对应的渲染图像，以结合人脸样本图像、人脸样本图像对应的渲染图像以及第一训练 3DMM 参数得到预设第一损失函数，进而训练第一参数估计网络）；之后，获取第二训练集，第二训练集包括多个人脸样本图像和每个人脸样本图像对应的粗糙样本 3D 表情模板；通过第二参数估计网络模型得到所述人脸样本图像的第二训练 3DMM 参数，通过映射网络得到粗糙样本 3D 表情模板对应的精细样本 3D 表情模板；通过预设第二损失函数训练第二参数估计网络模型和映射网络（即基于第二训练 3DMM 参数重建得到人脸样本图像对应的渲染图像，以结合人脸样本图像、人脸样本图像对应的渲染图像、第二训练 3DMM 参数、粗糙样本 3D 表情模板和对应的精细样本 3D 表情模板得到预设第二损失函数，进而训练第二参数估计网络和映射网络）。之后，在训练系数回归网络的过程中，基于前述训练好的第一参数估计网络模型和映射网络得到第一人脸图像的精细 3D 表情模板（参考图 8 所示的方法），即获取包含第一目

标对象的第一人脸图像后,将其输入第一参数估计网络模型得到第一 3DMM 参数,之后,基于第一 3DMM 参数和预设的 3DMM 模型构造粗糙 3D 表情模板,根据粗糙 3D 表情模板和通过训练好的映射网络得到三维表情形变,根据三维表情形变和粗糙 3D 表情模板的得到对应的精细 3D 表情模板。之后,包含第一目标对象的多帧第一人脸图像共用相同的精细 3D 表情模板。之后,基于多帧第一人脸图像训练系数回归网络(参考实施例一至实施例三的相关描述),在生成三维人脸形状时,系数回归网络生成的表情系数应用的表情基是基于对应的精细 3D 表情模板得到的(即表情基为公式(23)所示的 $B_i - B_0 + F(A_i \Delta_i)$),以构建更为准确的三维人脸形状,之后,基于三维人脸形状和三维人脸纹理得到二维重建图像,并根据二维重建图像、第一人脸图像和人脸重建系数确定系数回归网络的损失函数,并根据损失函数对系数回归网络进行训练,即在训练系数回归网络时,使用适用于第一目标对象的更为精准的 3D 表情模板,以进一步提高系数回归网络输出的表情系数的准确性。上述技术方案的具体实时细节可参考前述实施例的相关描述。

实施例五

图 12 为本申请一个实施例提供的一种用于人脸重建的网络训练装置的结构示意图。参考图 12,该用于人脸重建的网络训练装置包括:图像获取模块 501、图像输入模块 502、图像重建模块 503、损失构建模块 504、参数更新模块 505。

其中,图像获取模块 501,用于获取第一目标对象的多帧第一人脸图像;图像输入模块 502,用于将每帧第一人脸图像分别输入至对应的系数回归网络,以得到每帧第一人脸图像的人脸重建系数,各系数回归网络具有相同的网络参数,人脸重建系数包括:身份系数和表情系数;图像重建模块 503,用于根据人脸重建系数得到对应帧第一人脸图像的二维重建图像;损失构建模块 504,用于根据第一人脸图像、相应的二维重建图像和人脸重建系数构建系数回归网络的损失函数,损失函数包括各帧第一人脸图像的无监督训练损失函数和身份约束损失函数;参数更新模块 505,用于根据损失函数更新系数回归网络的网络参数,以训练所述系数回归网络。

本申请一个实施例中,人脸重建系数还包括:纹理系数和位姿系数,图像重建模块 503 包括:三维重建单元,用于根据身份系数和表情系数重建对应帧第一人脸图像的三维人脸形状,并根据纹理系数重建对应帧第一人脸图像的三维人脸纹理;渲染单元,用于根据位姿系数、三维人脸形状和三维人脸纹理渲染得到对应帧第一人脸图像的二维重建图像。

本申请一个实施例中,三维重建单元包括:形状重建子单元,用于将身份系数和所述表情系数分别应用于对应的身份基和表情基,并结合预先构建的平均人脸形状,得到对应帧所述第一人脸图像的三维人脸形状;纹理重建子单元,用于将纹理系数应用于对应的纹理基,并结合预先构建的平均人脸纹理,得到对应帧第一人脸图像的三维人脸纹理。

本申请一个实施例中,渲染单元包括:位姿系数应用子单元,用于将位姿系数应用于透视投影模型;渲染器渲染子单元,用于将透视投影模型、三维人脸形状和三维人脸纹理输入至渲染器,通过渲染器渲染得到对应帧第一人脸图像的二维重建图像。

本申请一个实施例中,人脸重建系数还包括:光照系数;该装置还包括:光照系数应用模块,用于根据位姿系数、三维人脸形状和三维人脸纹理渲染得到对应帧第一人脸图像的二维重建图像之前,将光照系数应用于光照模型;光照添加模块,用于将三维人脸纹理输入至光照模型,通过光照模型在三维人脸纹理

中添加第一人脸图像所在的环境光照。

本申请一个实施例中，所述光照模型为球谐光照模型，所述球谐光照模型使用球谐函数表中前三段的球谐函数。

本申请一个实施例中，损失构建模块 504 构建身份约束损失函数时，包括：标签系数获取单元，用于获取各帧第一人脸图像对应的身份标签系数；系数平均单元，用于将各身份标签系数的平均系数作为伪标签监督系数；第一计算单元，用于计算第一人脸图像的身份系数与伪标签监督系数之间的均方误差，并根据均方误差得到身份约束损失函数。

本申请一个实施例中，损失构建模块 504 构建身份约束损失函数时，还包括：系数构建单元，用于根据各所述系数回归网格输出的身份系数，构建伪标签监督系数；损失函数计算单元，用于根据所述第一人脸图像的身份系数与所述伪标签监督系数得到所述第一人脸图像的身份约束损失函数。其中，根据各所述系数回归网格输出的身份系数，构建伪标签监督系数包括：计算各所述系数回归网格输出的身份系数的平均系数，并将所述平均系数作为伪标签监督系数；所述根据所述第一人脸图像的身份系数与所述伪标签监督系数得到所述第一人脸图像的身份约束损失函数包括：计算所述第一人脸图像的身份系数与所述伪标签监督系数之间的均方误差，并根据所述均方误差得到身份约束损失函数。

本申请一个实施例中，无监督训练损失函数包括：第一人脸图像的图像重构损失函数，损失构建模块 504 构建无监督训练损失函数时，具体用于：利用 L2 范数计算第一人脸图像和相应二维重建图像间的像素误差，并根据像素误差得到第一人脸图像的图像重构损失函数。无监督训练损失函数包括：第一人脸图像的图像感知损失函数，损失构建模块 504 构建无监督训练损失函数时包括：特征提取单元，用于利用人脸识别网络分别提取第一人脸图像的第一图像特征和对应二维重建图像的第二图像特征；第二计算单元，用于计算第一图像特征和第二图像特征之间的相似度，并根据相似度得到第一人脸图像的图像感知损失函数。无监督训练损失函数包括：第一人脸图像的关键点损失函数，损失构建模块 504 构建无监督训练损失函数时包括：关键点获取单元，用于获取第一人脸图像对应的第一人脸关键点和相应二维重建图像的第二人脸关键点；第三计算单元，用于计算第一人脸关键点和第二人脸关键点间位置的均方误差，并根据均方误差得到第一人脸图像的关键点损失函数。无监督训练损失函数包括：第一人脸图像的正则损失，损失构建模块 504 构建无监督训练损失函数时包括：第四计算单元，用于计算第一人脸图像的身份系数的第一范数平方以及纹理系数的第二范数平方，并根据第一范数平方和第二范数平方得到第一人脸图像的身份及纹理约束损失函数；第五计算单元，用于使用 L1 范数对第一人脸图像的表情系数进行约束，以得到第一人脸图像的表情约束损失函数；第六计算单元，用于根据身份及纹理约束损失函数和表情约束损失函数得到第一人脸图像的正则损失。

本申请一个实施例中，关键点获取单元具体包括：使用头部姿态估计算法处理所述第一人脸图像，以得到对应的偏航角初始位置；偏航角初始位置不超过角度阈值时，使用 Dilb 检测所述第一人脸图像中位于人脸内部的第一人脸关键点，并使用 FAN 检测所述第一人脸图像中位于人脸轮廓的第一人脸关键点；偏航角起始值超过角度阈值时，使用 FAN 检测所述第一人脸图像中的第一人脸关键点；使用头部姿态估计算法处理所述第一人脸图像相应的二维重建图像，以得到对应的偏航角初始位置；偏航角初始位置不超过所述角度阈值时，使用 Dilb 检测所述二维重建图像中位于人脸内部的第二人脸关键点，并使用 FAN 检测所述二维重建图像中位于人脸脸颊的第二人脸关键点；偏航角起始值超过角度阈值时，使用 FAN 检测所述二维重建图像中的第二人脸关键点。

本申请一个实施例中，所述图像重构损失函数的表达式为：

$$L_{pho} = \frac{\sum_{i \in \mathcal{M}} A_i \cdot \|I_i - \hat{I}_i\|_2}{\sum_{i \in \mathcal{M}} A_i}$$

其中， L_{pho} 表示图像重构损失函数， $\mathcal{M}$ 表示人脸所在的区域， i 表示像素索引， $i \in \mathcal{M}$ ， A_i 表示人脸掩码中像素索引为 i 的像素的掩码值， I_i 表示第一人脸图像中像素索引为 i 的像素的像素值， $\hat{I}_i$ 表示二维重建图像中像素索引为 i 的像素的像素值；

所述图像感知损失函数的表达式为：

$$L_{per} = 1 - \frac{\langle f(I_i), f(\hat{I}_i) \rangle}{\|f(I_i)\| \cdot \|f(\hat{I}_i)\|}$$

其中， L_{per} 表示图像感知损失函数， $f(I_i)$ 表示第一人脸图像的第一图像特征， $f(\hat{I}_i)$ 表示二维重建图像的第二图像特征；

所述关键点损失函数的表达式为：

$$L_{km} = \sum_{i=1}^n \omega_i \cdot \|Q_i - \hat{Q}_i\|^2$$

其中， L_{km} 表示关键点损失函数， n 表示人脸关键点的总数量， i 表示当前的关键点索引， ω_i 表示关键点索引为 i 的人脸关键点的权重， Q_i 表示关键点索引为 i 的第一人脸关键点在第一人脸图像的位置， $\hat{Q}_i$ 表示关键点索引为 i 的第二人脸关键点在二维重建图像的位置；

所述身份及纹理约束损失函数的表达式为：

$$L_{reg} = \sum_{n=1}^M (\lambda_\alpha \| \alpha_n \|^2 + \lambda_\delta \| \delta_n \|^2)$$

其中， L_{reg} 表示身份及纹理约束损失函数， M 为身份基、纹理基的总数量， n 表示纹理基的索引和身份基的索引， α_n 表示索引为 n 的身份基在身份系数中对应的系数， δ_n 表示索引为 n 的纹理基在纹理系数中对应的系数， λ_α 和 λ_δ 为超参数；

所述表情约束损失函数的表达式为：

$$L_{sp} = \sum_{i=1}^m \lambda_\beta |\beta_i|$$

其中， L_{sp} 表示表情约束损失函数， m 为表情基的总数量， i 表示表情基的索引， β 表示索引为 i 的表情基在表情系数中对应的系数， λ_β 为超参数。

本申请一个实施例中，损失函数还包括各帧第一人脸图像的表情互斥约束损失函数，损失构建模块 504

构建表情互斥约束损失函数时还包括：互斥获取单元，用于获取预先构建的全部互斥表情对，每个互斥表情对包含至少两个互斥的表情基；系数查找单元，用于查找各所述互斥表情对包含的表情基在所述第一人脸图像中的表情系数；第八计算单元，用于根据查找到的表情系数，利用指数函数得到第一人脸图像的表情互斥约束损失函数。

本申请一个实施例中，表情互斥约束损失函数的表达式为：

$$L_{exp} = \sum_{(i,j) \in O} \|\exp\{\beta_i > \beta_j\} \cdot \beta_j\|^2 + \|\exp\{\beta_j > \beta_i\} \cdot \beta_i\|^2$$

其中， L_{exp} 表示表情互斥约束损失函数， O 表示各互斥表情对组成的集合， (i, j) 为 O 中的一个互斥表情对且所述互斥表情对包含第 i 个表情基和第 j 个表情基， β_i 和 β_j 分别表示第 i 个表情基和第 j 个表情基在第一人脸图像中的表情系数， $\exp\{\cdot\}$ 表示指数函数， $\|\cdot\|^2$ 表示二范数的平方。

本申请一个实施例中，系数回归网络训练完成后，还包括：第二人脸获取模块，用于获取第二目标对象的第二人脸图像；迁移系数确定模块，用于将所述第二人脸图像输入至系数回归网络，以得到第二人脸图像的人脸迁移系数，人脸迁移系数包括：表情系数和位姿系数；虚拟画像生成模块，用于根据表情系数和位姿系数生成与第二人脸图像对应的虚拟画像。

本申请一个实施例中，所述损失函数中所述身份约束损失函数的超参数大于所述表情互斥约束损失函数的超参数。

上述提供的用于人脸重建的网络训练装置可用于执行上述实施例一至实施例三任一提供的用于人脸重建的网络训练方法，具备相应的功能和有益效果。

实施例六

本申请一个实施例还提供的一种用于人脸重建的网络训练装置。该装置包括：第一参数计算模块，用于通过第一参数估计网络模型，计算输入的二维人脸图像对应的第一3DMM参数；粗糙形状确定模块，用于基于第一3DMM参数和预设的3DMM模型确定二维人脸图像对应的粗糙三维人脸形状（即粗糙3D表情模板）；第二参数计算模块，用于根据二维人脸图像和粗糙三维人脸形状（即粗糙3D表情模板），通过第二参数估计网络模型，计算二维人脸图像对应的第二3DMM参数；精细形状确定模块，用于基于第二3DMM参数和预设的3DMM模型确定二维人脸图像对应的精细三维人脸形状（即精细3D表情模板）。

在一实施方式中，第二参数计算模块，具体用于：确定粗糙三维人脸形状（即粗糙3D表情模板）相对于预设标准人脸的三维表情形变，预设标准人脸根据二维人脸图像自3DMM模型的预设标准人脸集中选择；根据所粗糙三维人脸形状（即粗糙3D表情模板）和三维表情形变，通过训练好的第二参数估计网络模型，计算二维人脸图像对应的第二3DMM参数。

在另一实施方式中，第二参数计算模块，进一步用于：通过UV映射将粗糙三维人脸形状（即粗糙3D表情模板）映射到UV空间，得到粗糙三维人脸形状（即粗糙3D表情模板）对应的粗糙二维UV图；确定粗糙二维UV图相对于预设标准人脸的二维UV图的二维表情形变；根据二维表情形变，通过映射网络确定粗糙三维人脸形状（即粗糙3D表情模板）相对于预设标准人脸的三维表情形变。

在另一实施方式中，第二参数计算模块，还用于：计算二维表情形变每一个顶点的欧式距离，根据欧式距离，形成二维表情形变的注意力掩码，注意力掩码大于等于0，且小于等于1。

该装置还包括第一训练模块，该第一训练模块用于：获取第一训练集；第一训练集包括多个人脸样本图像，每个人脸样本图像对应一组粗糙3DMM参数；根据第一训练集，对第一参数估计网络模型进行训

练。

在另一实施方式中，该第一训练模块，具体用于：将第一训练集中的每个人脸样本图像输入第一参数估计网络模型中，得到人脸样本图像对应的 3DMM 参数；通过预设第一损失函数训练第一参数估计网络模型，使基于人脸样本图像得到的 3DMM 参数等于对应的粗糙 3DMM 参数。

在又一实施方式中，第一训练模块，具体用于：获取第一训练集，所述第一训练集包括多个人脸样本图像；通过第一参数估计网络模型得到所述人脸样本图像对应的第一训练 3DMM 参数，所述第一训练 3DMM 参数包括身份系数、表情系数、纹理系数、位姿系数和光照系数；通过预设第一损失函数训练第一参数估计网络模型，预设第一损失函数包括图像重构损失函数、图像感知损失函数、关键点重构损失函数和正则损失函数，预设第一损失函数通过所述人脸样本图像、所述人脸样本图像对应的渲染图像和所述第一训练 3DMM 参数得到，所述渲染图像是将所述第一训练 3DMM 参数引入预设的 3DMM 模型后渲染到图像平面而得到。

在另一实施方式中，该装置还包括第二训练模块，该第二训练模块用于：获取第二训练集，第二训练集包括多个人脸样本图像和每个人脸样本图像对应的粗糙三维人脸样本形状（即粗糙 3D 表情模板）、精细 3DMM 参数；根据第二训练集，对第二参数估计网络模型进行训练。

在另一实施方式中，该第二训练模块，具体用于：确定第二训练集中的每个人脸样本图像对应的粗糙三维人脸样本形状（即粗糙 3D 表情模板）相对于预设标准人脸的三维表情样本形变，预设标准人脸根据二维人脸图像自 3DMM 模型的预设标准人脸集中选择；通过预设第二损失函数对第二参数估计网络模型进行训练，使基于人脸样本图像和对应的粗糙三维人脸形状得到的 3DMM 参数等于对应的精细 3DMM 参数。

在又一实施方式中，第二训练模块，具体用于：获取第二训练集，所述第二训练集包括多个人脸样本图像和每个人脸样本图像对应的粗糙样本 3D 表情模板，所述粗糙样本 3D 表情模板通过训练好的所述第一参数估计网络模型输出的所述人脸样本图像的第一训练 3DMM 参数构建；通过第二参数估计网络模型得到所述人脸样本图像的第二训练 3DMM 参数，所述第二训练 3DMM 参数包括表情系数、纹理系数、位姿系数和光照系数；通过预设第二损失函数训练所述第二参数估计网络模型，所述预设第二损失函数包括图像重构损失函数、图像感知损失函数、关键点重构损失函数、正则损失函数和表情梯度损失函数，所述预设第二损失函数通过所述人脸样本图像、所述人脸样本图像对应的渲染图像、所述人脸样本图像对应的粗糙样本 3D 表情模板和所述粗糙样本 3D 表情模板对应的精细样本 3D 表情模板得到，所述渲染图像是将基于所述第二训练 3DMM 参数得到的 3D 人脸渲染到图像平面后得到，所述精细样本 3D 表情模板通过所述粗糙样本 3D 表情模板得到。

在一实施方式中，该装置还包括：精细模块确定模块，用于通过映射网络得到所述粗糙样本 3D 表情模板对应的三维表情样本形变；根据所述三维表情样本形变和所述粗糙样本 3D 表情模板得到精细样本 3D 表情模板。

在一实施方式中，第二训练模块用于通过预设第二损失函数训练所述第二参数估计网络模型时，还用于：通过所述预设第二损失函数训练所述映射网络。

在一实施方式中，精细模块确定模块在通过映射网络得到所述粗糙样本 3D 表情模板对应的三维表情样本形变时，具体包括：将粗糙样本 3D 表情模板映射到 UV 空间，得到二维样本 UV 图；将所述二维样本 UV 图输入至映射网络，得到二维表情样本形变；将所述二维表情样本形变映射到粗糙样本 3D 表情模

板所在的三维空间, 得到三维表情样本形变。

在一实施方式中, 第二训练模块训练第二参数估计网络模型时, 该装置还包括: 掩码确定模块, 用于计算粗糙样本 3D 表情模板对于预设标准人脸的形变中每一个顶点的欧式距离, 所述预设标准人脸通过所述身份系数得到; 根据所述欧式距离, 形成注意力掩码, 并将所述注意力掩码应用于所述二维表情样本形变, 所述注意力掩码大于等于 0 且小于等于 1。

在一实施方式中, 该掩码确定模块在根据所述欧式距离, 形成注意力掩码时, 具体包括: 若所述欧式距离大于或等于设定的阈值, 则保留所述欧式距离, 若所述欧式距离小于设定的阈值, 则将所述欧式距离置为零; 将各所述顶点的欧式距离进行归一化, 形成注意力掩码。

在一实施方式中, 所述精细样本 3D 表情模板表示为: $\mathbf{B}_i' = \mathbf{B}_i + F(\mathbf{A}_i \Delta_i)$ 。其中, $\mathbf{B}_i$ 表示第 i 个粗糙样本 3D 表情模板, $\mathbf{B}_i'$ 表示 $\mathbf{B}_i$ 对应的精细样本 3D 表情模板, Δ_i 表示 $\mathbf{B}_i$ 对应的二维表情样本形状, $\mathbf{A}_i$ 表示 Δ_i 对应的注意力掩码, $F(\mathbf{A}_i \Delta_i)$ 表示三维表情样本形变。

在一实施方式中, 第二训练模块在获取第二训练集时, 具体包括: 获取多个人脸样本图像; 将每个所述人脸样本图像分别输入至训练好的所述第一参数估计网络模型, 得到每个所述人脸样本图像对应的第一训练 3DMM 参数; 基于第一训练 3DMM 参数和预设的 3DMM 模型确定所述人脸样本图像对应的粗糙样本 3D 表情模板。

在一实施方式中, 第二训练模块在获取基于第一训练 3DMM 参数和预设的 3DMM 模型确定所述人脸样本图像对应的粗糙样本 3D 表情模板时, 具体包括: 将所述第一训练 3DMM 参数中的身份系数应用于表情基底, 并结合 3DMM 模型的平均人脸的形状, 得到所述人脸样本图像重建的预设标准人脸; 根据预设表情人脸和表情基底的叠加结果, 得到粗糙样本 3D 表情模板。

在一实施方式中, 所述预设第一损失函数为:

$$L_{\text{com}} = \lambda_{\text{pho}} L_{\text{pho}} + \lambda_{\text{per}} L_{\text{per}} + \lambda_{\text{im}} L_{\text{im}} + \lambda_{\text{reg}} L_{\text{reg}} + \lambda_{\text{sp}} L_{\text{sp}}$$

其中, L_{pho} 、 L_{per} 、 L_{im} 和 $(L_{\text{reg}}, L_{\text{sp}})$ 分别为通过图像重构损失函数、图像感知损失函数、关键点重构损失函数和正则损失函数计算的损失值; λ_{pho} 、 λ_{per} 、 λ_{im} 、 λ_{reg} 、 λ_{sp} 均大于 0, 分别为相应损失函数的超参数;

所述图像重构损失函数的表达式为:

$$L_{\text{pho}} = \frac{\sum_{i \in \mathcal{M}} A_i \cdot \|I_i - \hat{I}_i\|_2}{\sum_{i \in \mathcal{M}} A_i}$$

其中, $\mathcal{M}$ 表示人脸所在的区域, i 表示像素索引, $i \in \mathcal{M}$, A_i 表示人脸掩码中像素索引为 i 的像素的掩码值, I_i 表示人脸样本图像中像素索引为 i 的像素的像素值, $\hat{I}_i$ 表示重建后的渲染图像中像素索引为 i 的像素的像素值;

所述图像感知损失函数的表达式为:

$$L_{\text{per}} = 1 - \frac{\langle f(I_i), f(\hat{I}_i) \rangle}{\|f(I_i)\| \cdot \|f(\hat{I}_i)\|}$$

其中, $f(I_i)$ 表示人脸样本图像的图像特征, $f(\hat{I}_i)$ 表示渲染图像的图像特征;

所述关键点损失函数的表达式为:

$$L_{\text{km}} = \sum_{i=1}^n \omega_i \cdot \|Q_i - \hat{Q}_i\|^2$$

其中, n 表示人脸关键点的总数量, i 表示关键点索引, ω_i 表示关键点权重, Q_i 表示关键点索引为 i 的人脸关键点在人脸样本图像的位置, $\hat{Q}_i$ 表示关键点索引为 i 的人脸关键点在渲染图像的位置;

所述正则损失函数的表达式包括:

$$L_{\text{reg}} = \sum_{n=1}^M (\lambda_{\alpha} \|\alpha_n\|^2 + \lambda_{\delta} \|\delta_n\|^2)$$

其中, M 为身份基底、纹理基底对应的总数量, n 表示索引, α_n 表示索引为 n 的身份系数, δ_n 表示索引为 n 的纹理系数, λ_{α} 和 λ_{δ} 为超参数;

所述正则损失函数的表达式还包括:

$$L_{\text{xf}} = \sum_{i=1}^m \lambda_{\beta} |\beta_i|$$

其中, m 为表情模板板数, i 表示表情模板的索引, β_i 表示索引为 i 的表情系数, λ_{β} 为超参数。

在一实施方式中, 所述预设第二损失函数为:

$$L = L_{\text{edm}} + \lambda_{\text{grx}} L_{\text{grx}}$$

其中, L_{edm} 为预设第一损失函数, L_{grx} 为表情梯度损失函数, λ_{grx} 为表情梯度损失函数的超参数,

所述表情梯度损失函数为:

$$L_{\text{grx}} = \|G_{x_a \rightarrow (x_i + \Delta i)} - G_{x_a \rightarrow x_i}\|^2$$

其中, $G_{a \rightarrow b}$ 表示变形后的三维人脸图像 b 对于原三维人脸图像 a 的梯度。

在一实施方式中, 第一训练模块在通过第一训练 3DMM 参数得到渲染图像时, 具体用于: 将第一训练 3DMM 参数引入预设的 3DMM 模型, 所述 3DMM 模型包括人脸形状和人脸纹理; 将 3DMM 模型中的人脸纹理经过球谐光照模型处理, 所述球谐光照模型处理过程中结合光照系数, 以建模环境光照; 将经过球谐光照模型处理后的 3DMM 模型经过透视投影相机模型, 渲染到二维图像上, 以得到渲染图像, 所述透视投影相机模型处理过程中结合位姿系数。

在一实施方式中, 该装置还包括: 精细状态确定模块, 用于通过训练好的第一参数估计网络, 计算输

入的二维人脸图像对应的第一 3DMM 参数；基于第一 3DMM 参数和预设的 3DMM 模型，得到二维人脸图像对应的粗糙三维人脸形状；根据粗糙三维人脸形状，通过训练好的映射网络，得到三维表情形变；根据所述粗糙三维人脸形状和三维表情形变得到所述二维人脸图像对应的精细三维人脸形状。

在一实施方式中，所述第二参数估计网络模型的训练过程中使用第一参数估计网络模型输出的身份系数。

上述提供的用于人脸重建的网络训练装置可用于执行上述实施例四提供的用于人脸重建的网络训练方法，具备相应的功能和有益效果。

值得注意的是，上述用于人脸重建的网络训练装置的实施例中，所包括的各个单元和模块只是按照功能逻辑进行划分的，但并不局限于上述的划分，只要能够实现相应的功能即可；另外，各功能单元的具体名称也只是为了便于相互区分，并不用于限制本申请的保护范围。

实施例七

图 14 为本申请实施例提供的一种用于人脸重建的网络训练设备的结构示意图。如图 14 所示，该用于人脸重建的网络训练设备包括处理器 50、存储器 51、输入装置 52、输出装置 53；用于人脸重建的网络训练设备中处理器 50 的数量可以是一个或多个，图 14 中以一个处理器 50 为例。用于人脸重建的网络训练设备中处理器 50、存储器 51、输入装置 52、输出装置 53 可以通过总线或其他方式连接，图 9 中以通过总线连接为例。

存储器 51 作为一种计算机可读存储介质，可用于存储软件程序、计算机可执行程序以及模块，如本申请实施例中的用于人脸重建的网络训练方法对应的程序指令/模块（例如，用于人脸重建的网络训练装置中的图像获取模块 501、图像输入模块 502、图像重建模块 503、损失构建模块 504、参数更新模块 505，或者用于人脸重建的网络训练装置中的第一训练模块、第二训练模块）。处理器 50 通过运行存储在存储器 51 中的软件程序、指令以及模块，从而执行用于人脸重建的网络训练设备的各种功能应用以及数据处理，即实现上述的用于人脸重建的网络训练方法。

存储器 51 可主要包括存储程序区和存储数据区，其中，存储程序区可存储操作系统、至少一个功能所需的应用程序；存储数据区可存储根据用于人脸重建的网络训练设备的使用所创建的数据等。此外，存储器 51 可以包括高速随机存取存储器，还可以包括非易失性存储器，例如至少一个磁盘存储器件、闪存器件、或其他非易失性固态存储器件。在一些实例中，存储器 51 可进一步包括相对于处理器 50 远程设置的存储器，这些远程存储器可以通过网络连接至用于人脸重建的网络训练设备。上述网络的实例包括但不限于互联网、企业内部网、局域网、移动通信网及其组合。

输入装置 52 可用于接收输入的数字或字符信息，以及产生与用于人脸重建的网络训练设备的用户设置以及功能控制有关的键信号输入，还可包括图像采集装置、音频采集装置等。输出装置 53 可包括显示屏等显示设备，还可包括音频播放装置。另外，用于人脸重建的网络训练设备还可以包括通信装置，以实现与远端设备的数据通信，如进行视频交流。

上述用于人脸重建的网络训练设备包含用于人脸重建的网络训练装置，可以用于执行任意用于人脸重建的网络训练方法，具备相应的功能和有益效果。

实施例八

本申请实施例还提供一种包含计算机可执行指令的存储介质，所述计算机可执行指令在由计算机处理

器执行时用于执行本申请任意实施例所提供的用于人脸重建的网络训练方法中的相关操作，且具备相应的功能和有益效果。

本领域内的技术人员应明白，本申请的实施例可提供为方法、系统、或计算机程序产品。

因此，本申请可采用完全硬件实施例、完全软件实施例、或结合软件和硬件方面的实施例的形式。而且，本申请可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质（包括但不限于磁盘存储器、CD-ROM、光学存储器等）上实施的计算机程序产品的形式。本申请是参照根据本申请实施例的方法、设备（系统）、和计算机程序产品的流程图和/或方框图来描述的。应理解可由计算机程序指令实现流程图和/或方框图中的每一流程和/或方框、以及流程图和/或方框图中的流程和/或方框的结合。可提供这些计算机程序指令到通用计算机、专用计算机、嵌入式处理机或其他可编程数据处理设备的处理器以产生一个机器，使得通过计算机或其他可编程数据处理设备的处理器执行的指令产生用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的装置。这些计算机程序指令也可存储在能引导计算机或其他可编程数据处理设备以特定方式工作的计算机可读存储器中，使得存储在该计算机可读存储器中的指令产生包括指令装置的制造品，该指令装置实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能。这些计算机程序指令也可装载到计算机或其他可编程数据处理设备上，使得在计算机或其他可编程设备上执行一系列操作步骤以产生计算机实现的处理，从而在计算机或其他可编程设备上执行的指令提供用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的步骤。

在一个典型的配置中，计算设备包括一个或多个处理器(CPU)、输入/输出接口、网络接口和内存。存储器可能包括计算机可读介质中的非永久性存储器，随机存取存储器(RAM)和/或非易失性内存等形式，如只读存储器(ROM)或闪存(flash RAM)。存储器是计算机可读介质的示例。

计算机可读介质包括永久性和非永久性、可移动和非可移动媒体可以由任何方法或技术来实现信息存储。信息可以是计算机可读指令、数据结构、程序的模块或其他数据。计算机的存储介质的例子包括，但不限于相变内存(PRAM)、静态随机存取存储器(SRAM)、动态随机存取存储器(DRAM)、其他类型的随机存取存储器(RAM)、只读存储器(ROM)、电可擦除可编程只读存储器(EEPROM)、快闪记忆体或其他内存技术、只读光盘只读存储器(CD-ROM)、数字多功能光盘(DVD)或其他光学存储、磁盒式磁带，磁带磁磁盘存储或其他磁性存储设备或任何其他非传输介质，可用于存储可以被计算设备访问的信息。按照本文中的界定，计算机可读介质不包括暂存电脑可读媒体(transitory media)，如调制的数据信号和载波。

还需要说明的是，术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含，从而使得包括一系列要素的过程、方法、商品或者设备不仅包括那些要素，而且还包括没有明确列出的其他要素，或者是还包括为这种过程、方法、商品或者设备所固有的要素。在没有更多限制的情况下，由语句“包括一个……”限定的要素，并不排除在包括要素的过程、方法、商品或者设备中还存在另外的相同要素。

注意，上述仅为本申请的较佳实施例及所运用技术原理。本领域技术人员会理解，本申请不限于这里所述的特定实施例，对本领域技术人员来说能够进行各种明显的变化、重新调整和替代而不会脱离本申请的保护范围。因此，虽然通过以上实施例对本申请进行了较为详细的说明，但是本申请不仅仅限于以上实施例，在不脱离本申请构思的情况下，还可以包括更多其他等效实施例，而本申请的范围由所附的权利要求范围决定。

权 利 要 求 书

1、一种用于人脸重建的网络训练方法，其中，包括：

获取第一目标对象的多帧第一人脸图像；

将每帧所述第一人脸图像分别输入至对应的系数回归网络，以得到每帧所述第一人脸图像的人脸重建系数，各所述系数回归网络具有相同的网络参数，所述人脸重建系数包括身份系数、表情系数、纹理系数、位姿系数和光照系数；

根据所述人脸重建系数得到对应帧所述第一人脸图像的二维重建图像；

根据所述第一人脸图像、相应的二维重建图像和所述人脸重建系数构建所述系数回归网络的损失函数，所述损失函数包括各帧第一人脸图像的无监督训练损失函数和身份约束损失函数，所述无监督训练损失函数包括：图像重构损失函数、图像感知损失函数、关键点损失函数和正则损失函数；

根据所述损失函数更新所述系数回归网络的网络参数，以训练所述系数回归网络。

2、根据权利要求1所述的方法，其中，所述根据所述人脸重建系数得到对应帧所述第一人脸图像的二维重建图像包括：

根据所述身份系数和所述表情系数重建对应帧所述第一人脸图像的三维人脸形状，并根据所述纹理系数重建对应帧所述第一人脸图像的三维人脸纹理；

根据所述位姿系数、所述三维人脸形状和所述三维人脸纹理渲染得到对应帧所述第一人脸图像的二维重建图像。

3、根据权利要求2所述的方法，其中，所述根据所述身份系数和所述表情系数重建对应帧所述第一人脸图像的三维人脸形状包括：

将所述身份系数和所述表情系数分别应用于对应的身份基和表情基，并结合预先构建的平均人脸形状，得到对应帧所述第一人脸图像的三维人脸形状；

所述根据所述纹理系数重建对应帧所述第一人脸图像的三维人脸纹理包括：

将所述纹理系数应用于对应的纹理基，并结合预先构建的平均人脸纹理，得到对应帧第一人脸图像的三维人脸纹理。

4、根据权利要求2所述的方法，其中，所述根据所述位姿系数、所述三维人脸形状和所述三维人脸纹理渲染得到对应帧所述第一人脸图像的二维重建图像包括：

将所述位姿系数应用于透视投影模型；

将所述透视投影模型、所述三维人脸形状和所述三维人脸纹理输入至渲染器，通过所述渲染器渲染得到对应帧所述第一人脸图像的二维重建图像。

5、根据权利要求2或4所述的方法，其中，所述根据所述位姿系数、所述三维人脸形状和所述三维人脸纹理渲染得到对应帧所述第一人脸图像的二维重建图像之前，还包括：

将所述光照系数应用于光照模型；

将所述三维人脸纹理输入至所述光照模型，通过所述光照模型在所述三维人脸纹理中添加所述第一人脸图像所在的环境光照。

6、根据权利要求5所述的方法，其中，所述光照模型为球谐光照模型，所述球谐光照模型使用球谐函数表中前三段的球谐函数。

7、根据权利要求1所述的方法，其中，根据所述第一人脸图像、相应的二维重建图像和所述人脸重建系数构建系数回归网络的身份约束损失函数包括：

获取各帧所述第一人脸图像对应的身份标签系数;

将各所述身份标签系数的平均系数作为伪标签监督系数;

计算所述第一人脸图像的身份系数与所述伪标签监督系数之间的均方误差,并根据所述均方误差得到身份约束损失函数。

8、根据权利要求1所述的方法,其中,所述根据所述第一人脸图像、相应的二维重建图像和所述人脸重建系数构建系数回归网络的身份约束损失函数包括:

根据各所述系数回归网络输出的身份系数,构建伪标签监督系数;

根据所述第一人脸图像的身份系数与所述伪标签监督系数得到所述第一人脸图像的身份约束损失函数。

9、根据权利要求8所述的方法,其中,所述根据各所述系数回归网络输出的身份系数,构建伪标签监督系数包括:

计算各所述系数回归网络输出的身份系数的平均系数,并将所述平均系数作为伪标签监督系数;

所述根据所述第一人脸图像的身份系数与所述伪标签监督系数得到所述第一人脸图像的身份约束损失函数包括:

计算所述第一人脸图像的身份系数与所述伪标签监督系数之间的均方误差,并根据所述均方误差得到身份约束损失函数。

10、根据权利要求2所述的方法,其中,所述无监督训练损失函数包括:所述第一人脸图像的图像重构损失函数,

根据所述第一人脸图像、相应的二维重建图像和所述人脸重建系数构建无监督训练损失函数包括:

利用L2范数计算第一人脸图像和相应二维重建图像间的像素误差,并根据所述像素误差得到所述第一人脸图像的图像重构损失函数;

所述无监督训练损失函数包括:所述第一人脸图像的图像感知损失函数,

根据所述第一人脸图像、相应的二维重建图像和所述人脸重建系数构建无监督训练损失函数包括:

利用人脸识别网络分别提取所述第一人脸图像的第一图像特征和对应二维重建图像的第二图像特征;

计算所述第一图像特征和所述第二图像特征之间的相似度,并根据所述相似度得到所述第一人脸图像的图像感知损失函数;

所述无监督训练损失函数包括:所述第一人脸图像的关键点损失函数,

根据所述第一人脸图像、相应的二维重建图像和所述人脸重建系数构建无监督训练损失函数包括:

获取第一人脸图像对应的第一人脸关键点和相应二维重建图像的第二人脸关键点;

计算所述第一人脸关键点和所述第二人脸关键点间位置的均方误差,并根据所述均方误差得到所述第一人脸图像的关键点损失函数;

所述无监督训练损失函数包括:所述第一人脸图像的正则损失,

根据所述第一人脸图像、相应的二维重建图像和所述人脸重建系数构建无监督训练损失函数包括:

计算所述第一人脸图像的身份系数的第一范数平方以及纹理系数的第二范数平方,并根据所述第一范数平方和所述第二范数平方得到所述第一人脸图像的身份及纹理约束损失函数;

使用L1范数对所述第一人脸图像的表情系数进行约束,以得到所述第一人脸图像的表情约束损失函数;

根据所述身份及纹理约束损失函数和所述表情约束损失函数得到所述第一人脸图像的正则损失。

11、根据权利要求 10 所述的方法，其中，所述获取第一人脸图像对应的第一人脸关键点包括：

使用头部姿态估计算法处理所述第一人脸图像，以得到对应的偏航角初始位置；

所述偏航角初始位置不超过角度阈值时，使用 Dilb 检测所述第一人脸图像中位于人脸内部的第一人脸关键点，并使用 FAN 检测所述第一人脸图像中位于人脸轮廓的第一人脸关键点；

所述偏航角起始值超过角度阈值时，使用 FAN 检测所述第一人脸图像中的第一人脸关键点；

所述获取第一人脸图像相应二维重建图像的第二人脸关键点包括：

使用头部姿态估计算法处理所述第一人脸图像相应的二维重建图像，以得到对应的偏航角初始位置；

所述偏航角初始位置不超过所述角度阈值时，使用 Dilb 检测所述二维重建图像中位于人脸内部的第二人脸关键点，并使用 FAN 检测所述二维重建图像中位于人脸脸颊的第二人脸关键点；

所述偏航角起始值超过角度阈值时，使用 FAN 检测所述二维重建图像中的第二人脸关键点。

12、根据权利要求 10 所述的方法，其中，所述图像重构损失函数的表达式为：

$$L_{pho} = \frac{\sum_{i \in \mathcal{M}} A_i \cdot \|I_i - \hat{I}_i\|_2}{\sum_{i \in \mathcal{M}} A_i}$$

其中， L_{pho} 表示图像重构损失函数， $\mathcal{M}$ 表示人脸所在的区域， i 表示像素索引， $i \in \mathcal{M}$ ， A_i 表示人脸掩码中像素索引为 i 的像素的掩码值， I_i 表示第一人脸图像中像素索引为 i 的像素的像素值， $\hat{I}_i$ 表示二维重建图像中像素索引为 i 的像素的像素值；

所述图像感知损失函数的表达式为：

$$L_{per} = 1 - \frac{(f(I_i), f(\hat{I}_i))}{\|f(I_i)\| \cdot \|f(\hat{I}_i)\|}$$

其中， L_{per} 表示图像感知损失函数， $f(I_i)$ 表示第一人脸图像的第一图像特征， $f(\hat{I}_i)$ 表示二维重建图像的第二图像特征；

所述关键点损失函数的表达式为：

$$L_{km} = \sum_{i=1}^n \omega_i \cdot \|Q_i - \hat{Q}_i\|^2$$

其中， L_{km} 表示关键点损失函数， n 表示人脸关键点的总数量， i 表示当前的关键点索引， ω_i 表示关键点索引为 i 的人脸关键点的权重， Q_i 表示关键点索引为 i 的第一人脸关键点在第一人脸图像的位置， $\hat{Q}_i$ 表示关键点索引为 i 的第二人脸关键点在二维重建图像的位置；

所述身份及纹理约束损失函数的表达式为：

$$L_{reg} = \sum_{n=1}^M (\lambda_a \| \alpha_n \|^2 + \lambda_s \| \delta_n \|^2)$$

其中， L_{reg} 表示身份及纹理约束损失函数， M 为身份基、纹理基的总数量， n 表示纹理基的索引和身

份基的索引, α_n 表示索引为 n 的身份基在身份系数中对应的系数, δ_n 表示索引为 n 的纹理基在纹理系数中对应的系数, λ_α 和 λ_δ 为超参数;

所述表情约束损失函数的表达式为:

$$L_{sp} = \sum_{i=1}^m \lambda_\alpha |\beta_i|$$

其中, L_{sp} 表示表情约束损失函数, m 为表情基的总数量, i 表示表情基的索引, β 表示索引为 i 的表情基在表情系数中对应的系数, λ_α 为超参数。

13、根据权利要求 1 所述的方法, 其中, 所述损失函数还包括各帧第一人脸图像的表情互斥约束损失函数,

根据所述第一人脸图像、相应的二维重建图像和所述人脸重建系数构建表情互斥约束损失函数包括:
获取预先构建的全部互斥表情对, 每个互斥表情对包含至少两个互斥的表情基;
查找各所述互斥表情对包含的表情基在所述第一人脸图像中的表情系数;
根据查找到的表情系数, 利用指数函数得到第一人脸图像的表情互斥约束损失函数。

14、根据权利要求 13 所述的方法, 其中, 所述表情互斥约束损失函数的表达式为:

$$L_{exp} = \sum_{(i,j) \in O} \|\mathbb{I}[\beta_i > \beta_j] \cdot \beta_j\|^2 + \|\mathbb{I}[\beta_j > \beta_i] \cdot \beta_i\|^2$$

其中, L_{exp} 表示表情互斥约束损失函数, O 表示各互斥表情对组成的集合, (i, j) 为 O 中的一个互斥表情对且所述互斥表情对包含第 i 个表情基和第 j 个表情基, β_i 和 β_j 分别表示第 i 个表情基和第 j 个表情基在所述第一人脸图像中的表情系数, $\mathbb{I}\{\cdot\}$ 表示指数函数, $\|\cdot\|^2$ 表示二范数的平方。

15、根据权利要求 13 所述的方法, 其中, 所述损失函数中所述身份约束损失函数的超参数大于所述表情互斥约束损失函数的超参数。

16、根据权利要求 1 所述的方法, 其中, 所述系数回归网络训练完成后, 还包括:

获取第二目标对象的第二人脸图像;

将所述第二人脸图像输入至所述系数回归网络, 以得到所述第二人脸图像的人脸迁移系数, 所述人脸迁移系数包括表情系数和位姿系数;

根据所述表情系数和所述位姿系数生成与所述第二人脸图像对应的虚拟画像。

17、一种用于人脸重建的网络训练方法, 其中, 包括:

获取第一训练集, 所述第一训练集包括多个人脸样本图像;

通过第一参数估计网络模型得到所述人脸样本图像对应的第一训练 3DMM 参数, 所述第一训练 3DMM 参数包括身份系数、表情系数、纹理系数、位姿系数和光照系数;

通过预设第一损失函数训练所述第一参数估计网络模型, 所述预设第一损失函数包括图像重构损失函数、图像感知损失函数、关键点重构损失函数和正则损失函数, 所述预设第一损失函数通过所述人脸样本图像、所述人脸样本图像对应的渲染图像和所述第一训练 3DMM 参数得到, 所述渲染图像是将所述第一训练 3DMM 参数引入预设的 3DMM 模型后渲染到图像平面而得到;

获取第二训练集, 所述第二训练集包括多个人脸样本图像和每个人脸样本图像对应的粗糙样本 3D 表情模板, 所述粗糙样本 3D 表情模板通过训练好的所述第一参数估计网络模型输出的所述人脸样本图像的第一训练 3DMM 参数构建;

通过第二参数估计网络模型得到所述人脸样本图像的第二训练 3DMM 参数, 所述第二训练 3DMM 参数包括表情系数、纹理系数、位姿系数和光照系数;

通过预设第二损失函数训练所述第二参数估计网络模型, 所述预设第二损失函数包括图像重构损失函数、图像感知损失函数、关键点重构损失函数、正则损失函数和表情梯度损失函数, 所述预设第二损失函数通过所述人脸样本图像、所述人脸样本图像对应的渲染图像、所述人脸样本图像对应的粗糙样本 3D 表情模板和所述粗糙样本 3D 表情模板对应的精细样本 3D 表情模板得到, 所述渲染图像是将基于所述第二训练 3DMM 参数得到的 3D 人脸渲染到图像平面后得到, 所述精细样本 3D 表情模板通过所述粗糙样本 3D 表情模板得到。

18、根据权利要求 17 所述的方法, 其中, 还包括:

通过映射网络得到所述粗糙样本 3D 表情模板对应的三维表情样本形变;

根据所述三维表情样本形变和所述粗糙样本 3D 表情模板得到精细样本 3D 表情模板。

19、根据权利要求 18 所述的方法, 其中, 所述通过预设第二损失函数训练所述第二参数估计网络模型时, 还包括:

通过所述预设第二损失函数训练所述映射网络。

20、根据权利要求 18 所述的方法, 其中, 所述通过映射网络得到所述粗糙样本 3D 表情模板对应的三维表情样本形变包括:

将所述粗糙样本 3D 表情模板映射到 UV 空间, 得到二维样本 UV 图;

将所述二维样本 UV 图输入至映射网络, 得到二维表情样本形变;

将所述二维表情样本形变映射到粗糙样本 3D 表情模板所在的三维空间, 得到三维表情样本形变。

21、根据权利要求 20 所述的方法, 其中, 还包括:

计算粗糙样本 3D 表情模板对于预设标准人脸的形变中每一个顶点的欧式距离, 所述预设标准人脸通过所述身份系数得到;

根据所述欧式距离, 形成注意力掩码, 并将所述注意力掩码应用于所述二维表情样本形变, 所述注意力掩码大于等于 0 且小于等于 1。

22、根据权利要求 21 所述的方法, 其中, 所述根据所述欧式距离, 形成注意力掩码包括:

若所述欧式距离大于或等于设定的阈值, 则保留所述欧式距离, 若所述欧式距离小于设定的阈值, 则将所述欧式距离置为零;

将各所述顶点的欧式距离进行归一化, 形成注意力掩码。

23、根据权利要求 21 所述的方法, 其中, 所述精细样本 3D 表情模板表示为:

$$\mathbf{B}_i' = \mathbf{B}_i + F(\mathbf{A}_i \Delta_i)$$

其中, $\mathbf{B}_i$ 表示第 i 个粗糙样本 3D 表情模板, $\mathbf{B}_i'$ 表示 $\mathbf{B}_i$ 对应的精细样本 3D 表情模板, Δ_i 表示 $\mathbf{B}_i$ 对应的二维表情样本形状, $\mathbf{A}_i$ 表示 Δ_i 对应的注意力掩码, $F(\mathbf{A}_i \Delta_i)$ 表示三维表情样本形变。

24、根据权利要求 17 所述的方法，其中，所述获取第二训练集包括：

获取多个人脸样本图像；

将每个所述人脸样本图像分别输入至训练好的所述第一参数估计网络模型，得到每个所述人脸样本图像对应的第一训练 3DMM 参数；

基于所述第一训练 3DMM 参数和预设的 3DMM 模型确定所述人脸样本图像对应的粗糙样本 3D 表情模板。

25、根据权利要求 24 所述的方法，其中，所述基于所述第一训练 3DMM 参数和预设的 3DMM 模型确定所述人脸样本图像对应的粗糙样本 3D 表情模板包括：

将所述第一训练 3DMM 参数中的身份系数应用于表情基底，并结合 3DMM 模型的平均人脸的形状，得到所述人脸样本图像重建的预设标准人脸；

根据预设标准人脸和表情基底的叠加结果，得到粗糙样本 3D 表情模板。

26、根据权利要求 17 所述的方法，其中，所述预设第一损失函数为：

$$L_{\text{com}} = \lambda_{\text{pho}} L_{\text{pho}} + \lambda_{\text{per}} L_{\text{per}} + \lambda_{\text{im}} L_{\text{im}} + \lambda_{\text{reg}} L_{\text{reg}} + \lambda_{\text{sp}} L_{\text{sp}}$$

其中， L_{pho} 、 L_{per} 、 L_{im} 和 $(L_{\text{reg}}, L_{\text{sp}})$ 分别为通过图像重构损失函数、图像感知损失函数、关键点重构损失函数和正则损失函数计算的损失值； λ_{pho} 、 λ_{per} 、 λ_{im} 、 λ_{reg} 、 λ_{sp} 均大于 0，分别为相应损失函数的超参数；

所述图像重构损失函数的表达式为：

$$L_{\text{pho}} = \frac{\sum_{i \in \mathcal{M}} A_i \cdot \|I_i - \hat{I}_i\|_2}{\sum_{i \in \mathcal{M}} A_i}$$

其中， $\mathcal{M}$ 表示人脸所在的区域， i 表示像素索引， $i \in \mathcal{M}$ ， A_i 表示人脸掩码中像素索引为 i 的像素的掩码值， I_i 表示人脸样本图像中像素索引为 i 的像素的像素值， $\hat{I}_i$ 表示重建后的渲染图像中像素索引为 i 的像素的像素值；

所述图像感知损失函数的表达式为：

$$L_{\text{per}} = 1 - \frac{\langle f(I_i), f(\hat{I}_i) \rangle}{\|f(I_i)\| \cdot \|f(\hat{I}_i)\|}$$

其中， $f(I_i)$ 表示人脸样本图像的图像特征， $f(\hat{I}_i)$ 表示渲染图像的图像特征；

所述关键点损失函数的表达式为：

$$L_{\text{im}} = \sum_{i=1}^n \omega_i \cdot \|q_i - \hat{q}_i\|^2$$

其中， n 表示人脸关键点的总数量， i 表示关键点索引， ω_i 表示关键点权重， q_i 表示关键点索引为 i 的

人脸关键点在人脸样本图像的位置, $\tilde{Q}_i$ 表示关键点索引为 i 的人脸关键点在渲染图像的位置;

所述正则损失函数的表达式包括:

$$L_{reg} = \sum_{n=1}^M (\lambda_\alpha \|\alpha_n\|^2 + \lambda_\delta \|\delta_n\|^2)$$

其中, M 为身份基底、纹理基底对应的总数量, n 表示索引, α_n 表示索引为 n 的身份系数, δ_n 表示索引为 n 的纹理系数, λ_α 和 λ_δ 为超参数;

所述正则损失函数的表达式还包括:

$$L_{sp} = \sum_{i=1}^m \lambda_\beta |\beta_i|$$

其中, m 为表情模板板数, i 表示表情模板的索引, β_i 表示索引为 i 的表情系数, λ_β 为超参数。

27、根据权利要求 17 所述的方法, 其中, 所述预设第二损失函数为:

$$L = L_{com} + \lambda_{gra} L_{gra}$$

其中, L_{com} 为所述预设第一损失函数, L_{gra} 为表情梯度损失函数, λ_{gra} 为表情梯度损失函数的超参数,

所述表情梯度损失函数为:

$$L_{gra} = \|G_{x_a \rightarrow (x_i + \Delta i)} - G_{x_a \rightarrow x_i}\|^2$$

其中, $G_{a \rightarrow b}$ 表示变形后的三维人脸图像 b 对于原三维人脸图像 a 的梯度。

28、根据权利要求 17 所述的方法, 其中, 通过所述第一训练 3DMM 参数得到所述渲染图像的过程包括:

将第一训练 3DMM 参数引入预设的 3DMM 模型, 所述 3DMM 模型包括人脸形状和人脸纹理;

将 3DMM 模型中的人脸纹理经过球谐光照模型处理, 所述球谐光照模型处理过程中结合光照系数, 以建模环境光照;

将经过球谐光照模型处理后的 3DMM 模型经过透视投影相机模型, 渲染到图像平面上, 以得到渲染图像, 所述透视投影相机模型处理过程中结合位姿系数。

29、根据权利要求 19 所述的方法, 其中, 还包括:

通过训练好的第一参数估计网络, 计算输入的二维人脸图像对应的第一 3DMM 参数;

基于第一 3DMM 参数和预设的 3DMM 模型, 得到二维人脸图像对应的粗糙三维人脸形状;

根据粗糙三维人脸形状, 通过训练好的映射网络, 得到三维表情形变;

根据所述粗糙三维人脸形状和三维表情形变得到所述二维人脸图像对应的精细三维人脸形状。

30、根据权利要求 17 所述的方法, 其特征在于, 所述第二参数估计网络模型的训练过程中使用第一参数估计网络模型输出的身份系数。

31、一种用于人脸重建的网络训练装置，其中，包括：

图像获取模块，用于获取第一目标对象的多帧第一人脸图像；

图像输入模块，用于将每帧所述第一人脸图像分别输入至对应的系数回归网络，以得到每帧所述第一人脸图像的人脸重建系数，各所述系数回归网络具有相同的网络参数，所述人脸重建系数包括身份系数、表情系数、纹理系数、位姿系数和光照系数；

图像重建模块，用于根据所述人脸重建系数得到对应帧所述第一人脸图像的二维重建图像；

损失构建模块，用于根据所述第一人脸图像、相应的二维重建图像和所述人脸重建系数构建所述系数回归网络的损失函数，所述损失函数包括各帧第一人脸图像的无监督训练损失函数和身份约束损失函数，所述无监督训练损失函数包括：图像重构损失函数、图像感知损失函数、关键点损失函数和正则损失函数；

参数更新模块，用于根据所述损失函数更新所述系数回归网络的网络参数，以训练所述系数回归网络。

32、一种用于人脸重建的网络训练装置，其中，包括：

第一训练模块，用于获取第一训练集，所述第一训练集包括多个人脸样本图像；通过第一参数估计网络模型得到所述人脸样本图像对应的第一训练 3DMM 参数，所述第一训练 3DMM 参数包括身份系数、表情系数、纹理系数、位姿系数和光照系数；通过预设第一损失函数训练第一参数估计网络模型，所述预设第一损失函数包括图像重构损失函数、图像感知损失函数、关键点重构损失函数和正则损失函数，所述预设第一损失函数通过所述人脸样本图像、所述人脸样本图像对应的渲染图像和所述第一训练 3DMM 参数得到，所述渲染图像是将所述第一训练 3DMM 参数引入预设的 3DMM 模型后渲染到图像平面而得到；

第二训练模块，用于获取第二训练集，所述第二训练集包括多个人脸样本图像和每个人脸样本图像对应的粗糙样本 3D 表情模板，所述粗糙样本 3D 表情模板通过训练好的所述第一参数估计网络模型输出的所述人脸样本图像的第一训练 3DMM 参数构建；通过第二参数估计网络模型得到所述人脸样本图像的第二训练 3DMM 参数，所述第二训练 3DMM 参数包括表情系数、纹理系数、位姿系数和光照系数；通过预设第二损失函数训练所述第二参数估计网络模型，所述预设第二损失函数包括图像重构损失函数、图像感知损失函数、关键点重构损失函数、正则损失函数和表情梯度损失函数，所述预设第二损失函数通过所述人脸样本图像、所述人脸样本图像对应的渲染图像、所述人脸样本图像对应的粗糙样本 3D 表情模板和所述粗糙样本 3D 表情模板对应的精细样本 3D 表情模板得到，所述渲染图像是将基于所述第二训练 3DMM 参数得到的 3D 人脸渲染到图像平面后得到，所述精细样本 3D 表情模板通过所述粗糙样本 3D 表情模板得到。

33、一种用于人脸重建的网络训练设备，其中，包括：

一个或多个处理器；

存储器，用于存储一个或多个程序；

当所述一个或多个程序被所述一个或多个处理器执行，使得所述一个或多个处理器实现如权利要求 1-16 中任一所述的用于人脸重建的网络训练方法或如权利要求 17-30 中任一所述的用于人脸重建的网络训练方法。

34、一种计算机可读存储介质，其上存储有计算机程序，其中，该程序被处理器执行时实现如权利要求 1-16 中任一所述的用于人脸重建的网络训练方法或如权利要求 17-30 中任一所述的用于人脸重建的网络训练方法。

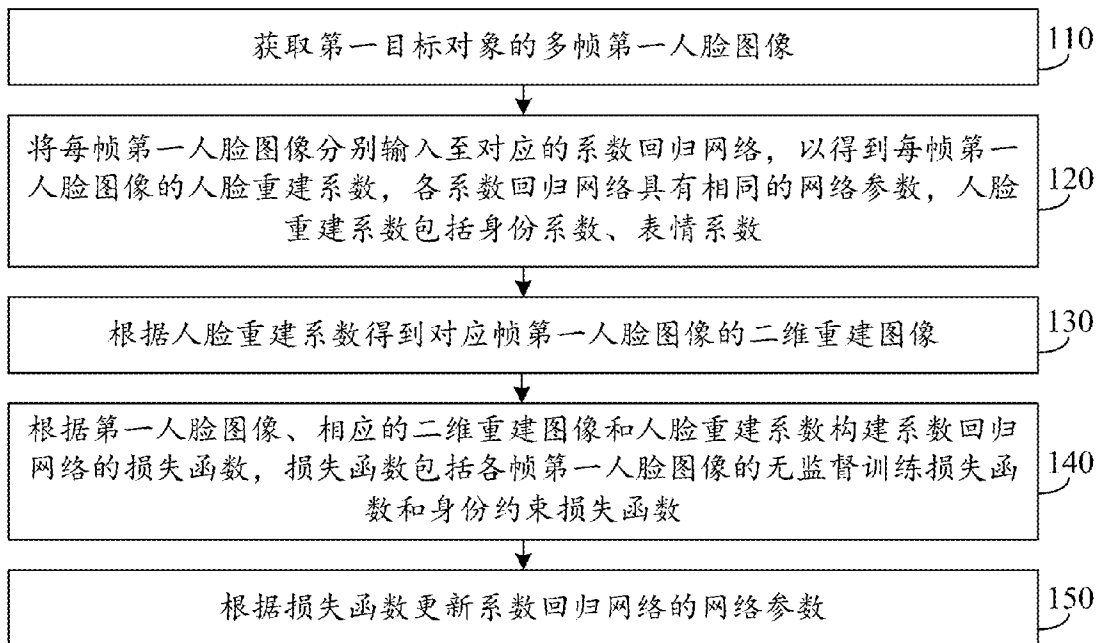


图 1

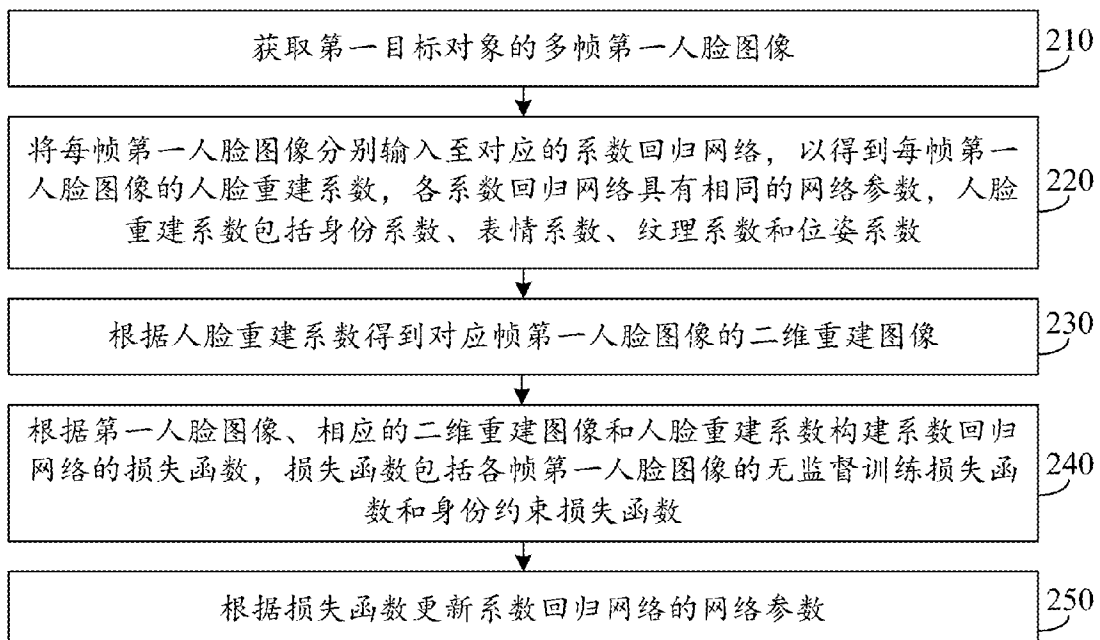


图 2

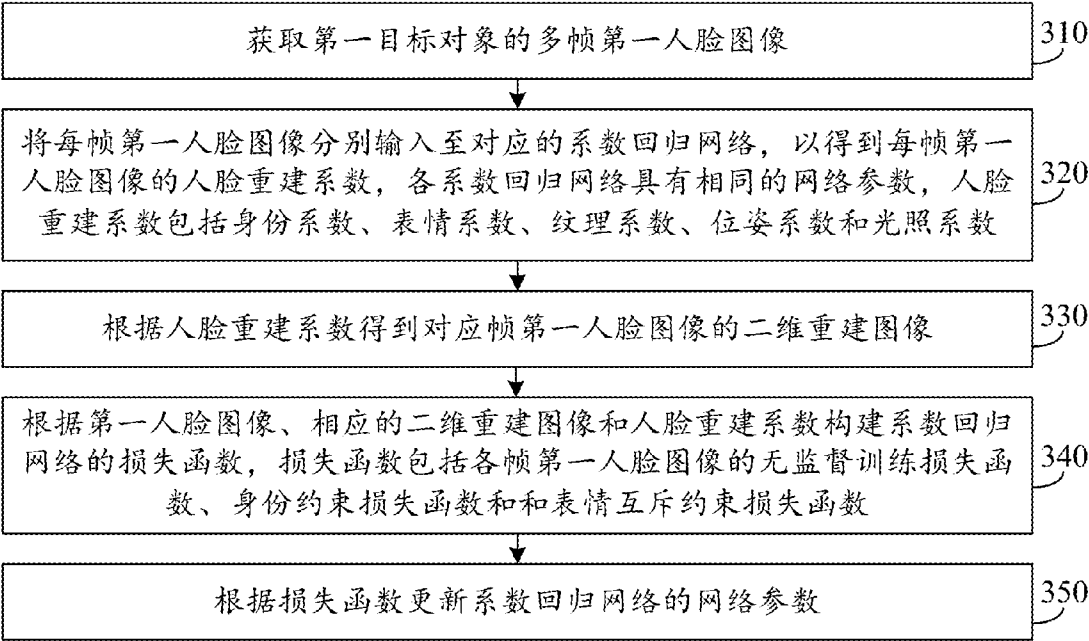


图 3

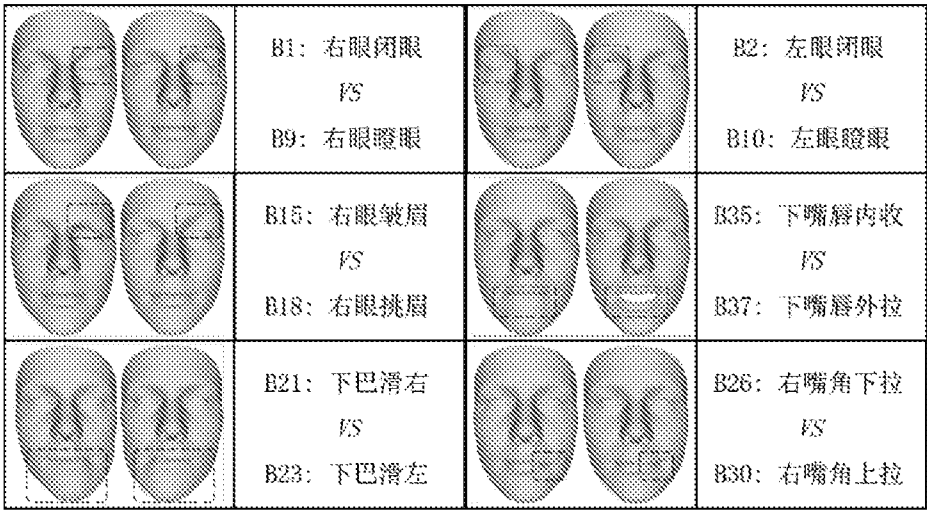


图 4

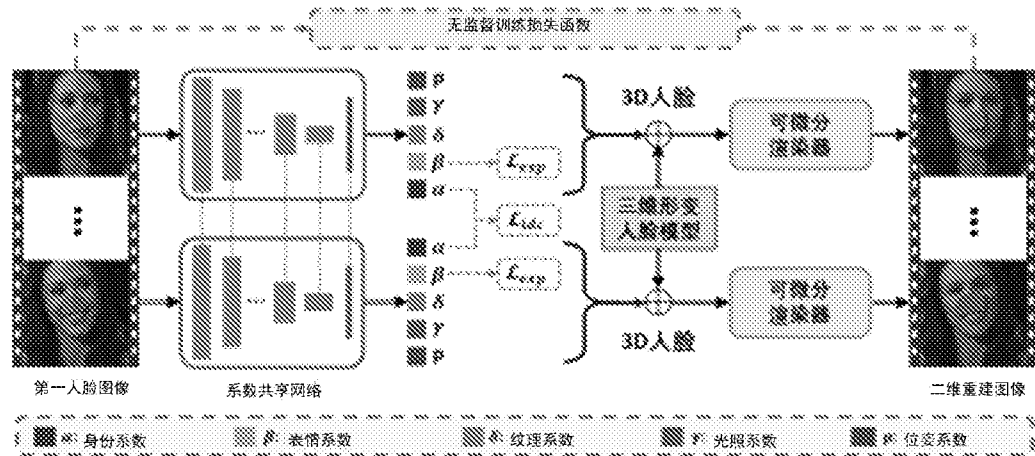


图 5

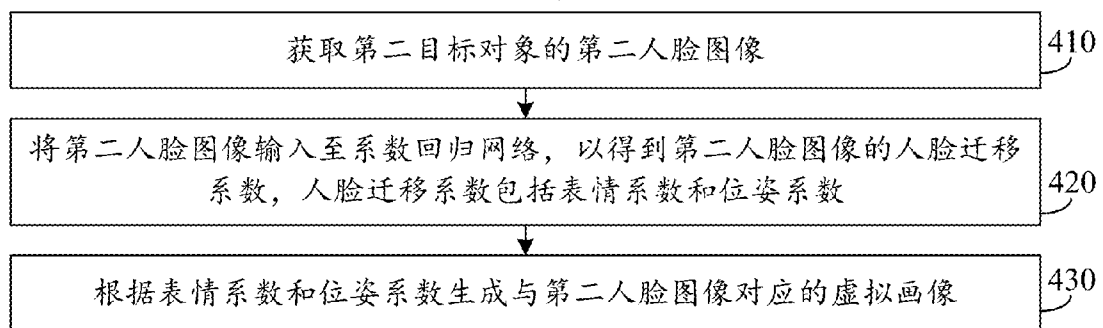


图 6

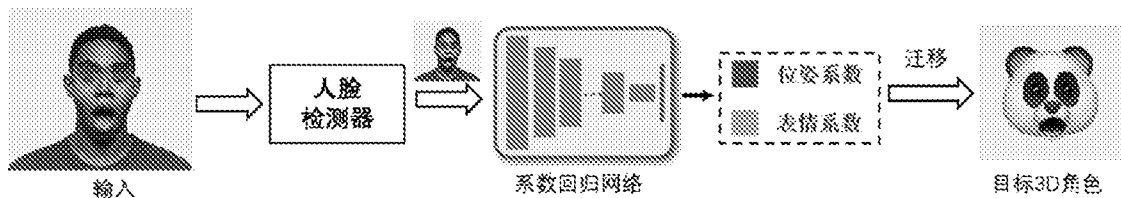


图 7

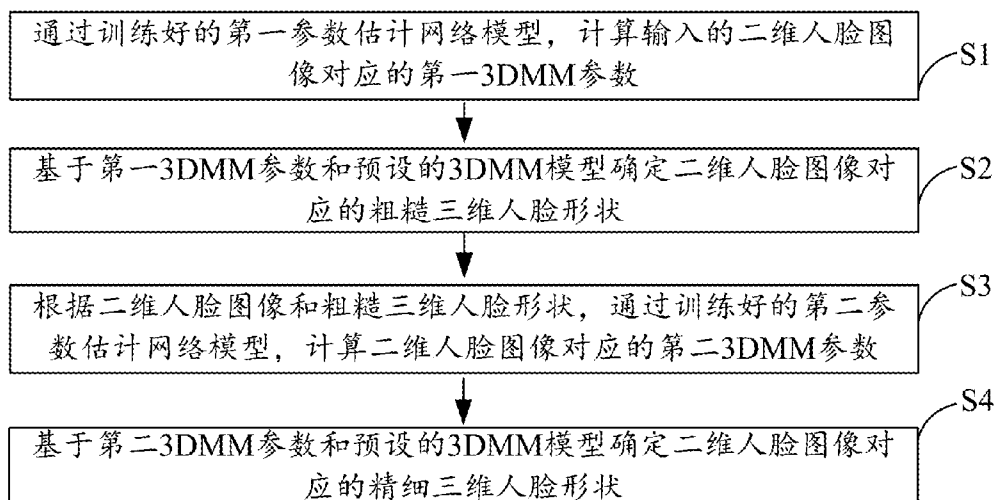


图 8

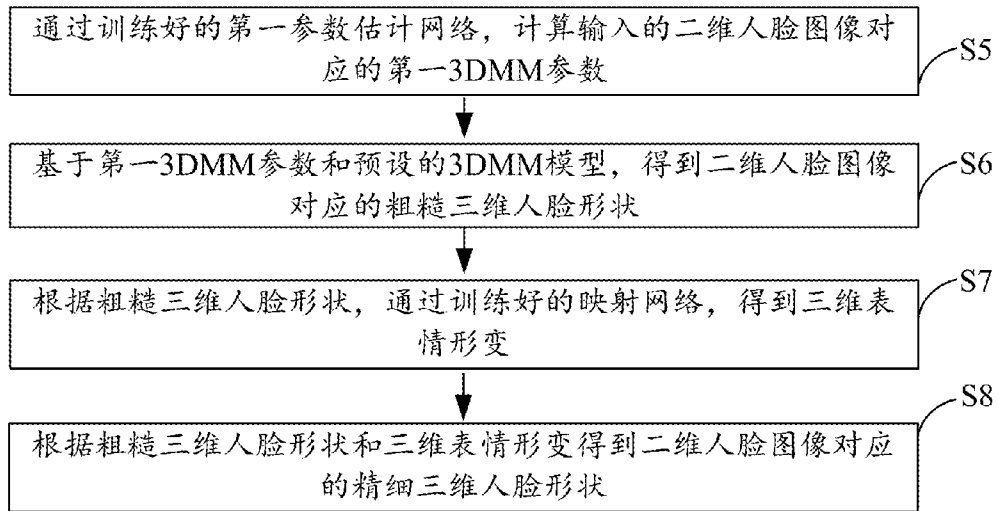


图 9

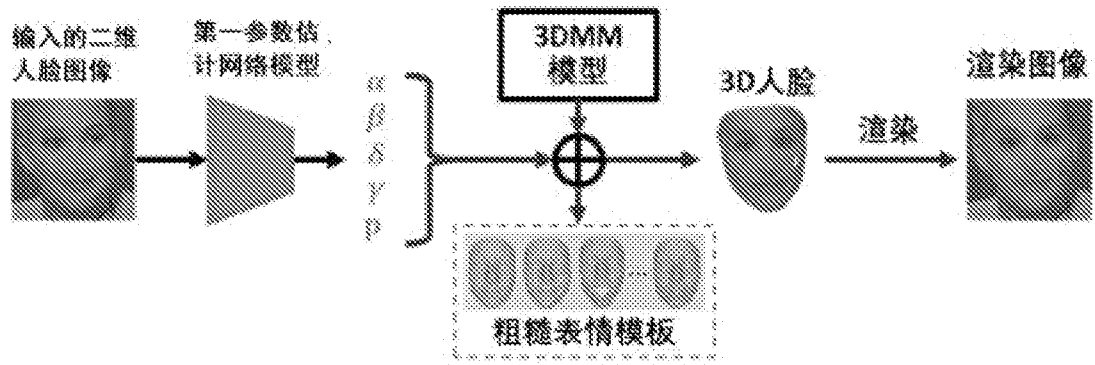


图 10

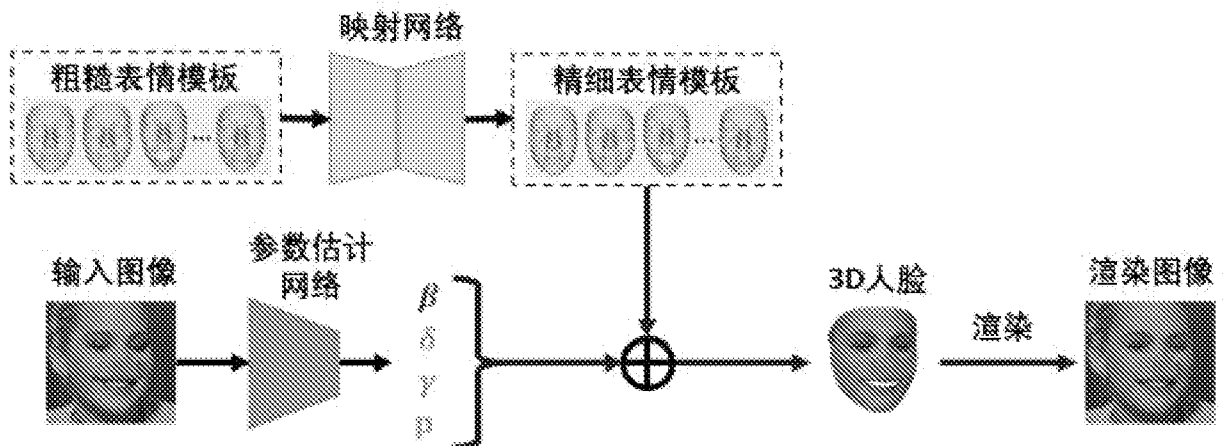


图 11

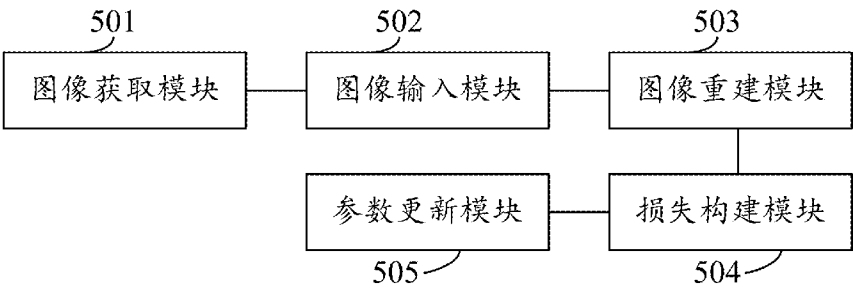


图 12

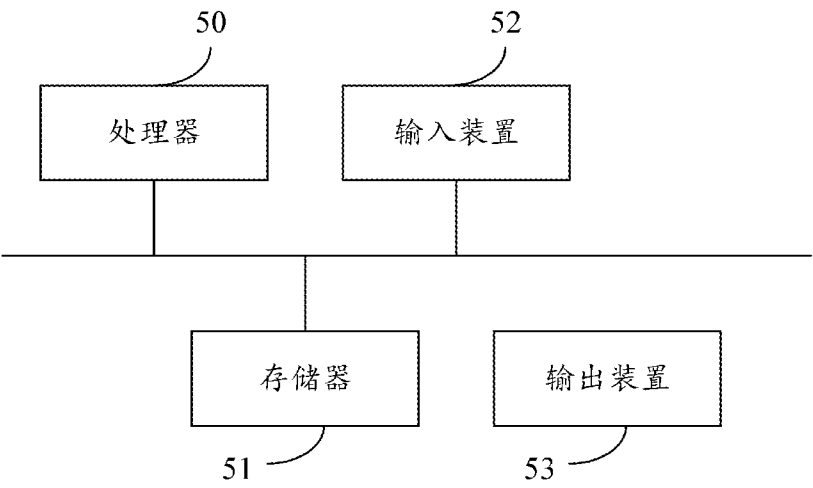


图 13

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2022/108304

A. CLASSIFICATION OF SUBJECT MATTER

G06T 13/40(2011.01)i; G06T 7/11(2017.01)i; G06T 7/194(2017.01)i; G06N 3/04(2006.01)i; G06N 3/08(2006.01)i;
G06K 9/62(2022.01)i; G06V 10/774(2022.01)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)

G06T G06N G06K G06V

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNABS; CNTXT; CNKI; VEN; USTXT; WOTXT; EPTXT; IEEE: 人脸, 重建, 回归, 系数, 身份, 表情, 纹理, 位姿, 光照, 损失, 感知, 关键点, 正则, 互斥, 梯度, face, reconstruction, regression, coefficient, identity, expression, texture, pose, illumination, loss, perception, keypoint, regulation, mutual, gradient, coarse, fine

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DENG, Yu et al. "Accurate 3D Face Reconstruction with Weakly-Supervised Learning: From Single Image to Image Set" <i>arXiv</i> , 20 March 2019 (2019-03-20), pp. 1-11	1-13, 15, 16, 31, 33, 34
A	DENG, Yu et al. "Accurate 3D Face Reconstruction with Weakly-Supervised Learning: From Single Image to Image Set" <i>arXiv</i> , 20 March 2019 (2019-03-20), pp. 1-11	14, 17-30, 32
A	CN 112288851 A (WUHAN UNIVERSITY) 29 January 2021 (2021-01-29) entire document	1-34
A	CN 112529999 A (BAIGUOYUAN TECHNOLOGY (SINGAPORE) PTE LTD.) 19 March 2021 (2021-03-19) entire document	1-34



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

31 August 2022

Date of mailing of the international search report

28 September 2022

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**

Authorized officer

Facsimile No. (86-10)62019451

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2022/108304

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	112288851	A	29 January 2021	None			
CN	112529999	A	19 March 2021	WO	2022095721	A1	12 May 2022

国际检索报告

国际申请号

PCT/CN2022/108304

A. 主题的分类

G06T 13/40(2011.01)i; G06T 7/11(2017.01)i; G06T 7/194(2017.01)i; G06N 3/04(2006.01)i; G06N 3/08(2006.01)i; G06K 9/62(2022.01)i; G06V 10/774(2022.01)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)

G06T G06N G06K G06V

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))

CNABS;CNTXT;CNKI;VEN;USTXT;WOTXT;EPTXT;IEEE: 人脸, 重建, 回归, 系数, 身份, 表情, 纹理, 位姿, 光照, 损失, 感知, 关键点, 正则, 互斥, 梯度, face, reconstruction, regression, coefficient, identity, expression, texture, pose, illumination, loss, perception, keypoint, regulation, mutual, gradient, coarse, fine

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
X	Yu Deng 等. "Accurate 3D Face Reconstruction with Weakly-Supervised Learning: From Single Image to Image Set" arXiv, 2019年3月20日 (2019 - 03 - 20), 第1-11页	1-13、15、 16、31、33、34
A	Yu Deng 等. "Accurate 3D Face Reconstruction with Weakly-Supervised Learning: From Single Image to Image Set" arXiv, 2019年3月20日 (2019 - 03 - 20), 第1-11页	14、17-30、32
A	CN 112288851 A (武汉大学) 2021年1月29日 (2021 - 01 - 29) 全文	1-34
A	CN 112529999 A (百果园技术新加坡有限公司) 2021年3月19日 (2021 - 03 - 19) 全文	1-34

☐ 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

2022年8月31日

国际检索报告邮寄日期

2022年9月28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

徐菲

电话号码 (86-512) 88997299

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2022/108304

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	112288851	A	2021年1月29日	无			
CN	112529999	A	2021年3月19日	W0	2022095721	A1	2022年5月12日