

人脸图像去模糊和三维重建的联合优化方法

徐崇润, 刘利刚*

(中国科学技术大学数学科学学院 合肥 230026)
(lgliu@ustc.edu.cn)

摘要: 为了解决在模糊人脸图像的去模糊任务和人脸三维重建任务存在的信息缺失导致的效果不佳的问题, 提出一种对模糊人脸图像去模糊和三维重建任务进行联合优化的方法. 对于输入的模糊人脸图像, 交替重复执行以下步骤 3 次: 首先, 对人脸图像进行三维重建, 生成三维人脸模型; 然后, 使用 Res-UNet 对人脸图像进行去模糊; 并以三维重建所得信息作为辅助输入, 以原始模糊图像作为校正输入, 生成消除模糊的图像. 在 AFLW 数据集的清晰人脸图像, 以及约 20 000 幅 Giacomo Boracchi 采集的模糊核合成模糊人脸图像组成的数据集上, 分别以去模糊图像和人脸三维模型再渲染的图像对输入图像计算峰值信噪比(peak signal-to-noise ratio, PSNR)和结构相似度(structural similarity, SSIM), 以衡量去模糊和三维重建任务效果. 对比实验结果表明, 在测试集上的去模糊任务中, 相比 3DFaceDeblur 方法, 该方法的平均 PSNR 从 27.38 dB 提升到 29.23 dB, 平均 SSIM 从 0.833 提升到 0.855; 在三维重建任务中, 相比 Deep3DFace 方法, 该方法的平均 PSNR 从 22.78 dB 提升到 27.16 dB, 平均 SSIM 从 0.848 提升到 0.918.

关键词: 深度学习; 人脸图像; 去模糊; 三维重建; 联合优化
中图法分类号: TP391.41 **DOI:** 10.3724/SP.J.1089.2023-00334

Joint Optimization for Deblurring and 3D Reconstruction of Blurry Face Image

Xu Chongrun and Liu Ligang*

(School of Mathematical Sciences, University of Science and Technology of China, Hefei 230026)

Abstract: In order to address the issue of low performance in deblurring and 3D reconstruction on blurred face images due to lack of information, a joint optimization method for deblurring and 3D reconstruction of blurry face images is proposed. Given a blurry face image, the following steps are iteratively executed: apply 3D reconstruction on the face image to generate a 3D facial model; deblur the face image using Res-UNet, with 3D reconstruction information as assist input and the original blurry image as correction input, to generate a deblurred image; repeat three times. The experiment utilized approximately 20 000 blurred facial images synthesized from clear facial images from AFLW dataset and blur kernels collected by Giacomo Boracchi as dataset. Peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) were calculated from re-rendered images from the 3D face model and deblurred images to blurry input images respectively, as metrics of the performance of the deblurring and 3D reconstruction. For deblurring, in comparison to 3DFaceDeblur method, average PSNR improved from 27.38 dB to 29.23 dB, average SSIM improved from 0.833 to 0.855. For 3D reconstruction task, compared to Deep3DFace method, average PSNR improved from 22.78 dB to 27.16 dB, average SSIM improved from 0.848 to 0.918.

Key words: deep learning; face images; deblurring; 3D reconstruction; joint optimization

收稿日期: 2023-06-17; 修回日期: 2023-12-25. 基金项目: 国家自然科学基金(62025207). 徐崇润(1995—), 男, 硕士研究生, 主要研究方向为计算机图形学; 刘利刚(1975—), 男, 博士, 教授, 博士生导师, CCF 杰出会员, 论文通信作者, 主要研究方向为计算机图形学.

在计算机视觉中, 图像去模糊和人脸图像三维重建是 2 个不同的研究方向, 图像去模糊旨在消除拍摄过程中因运动、抖动或者失焦等产生的图像模糊, 应用于相机优化和影视后期处理等方面; 人脸图像三维重建旨在由二维人脸图像生成对应的三维模型, 在增强现实、动画制作等方面有着广泛的应用. 长期以来, 由于目的、手段都相差甚远, 这 2 项技术通常是独立发展的. 各种通用的图像去模糊方法能够广泛地处理多种模糊图像, 但是在处理人脸图像去模糊这一特定问题时表现不佳, 尤其是不能很好地处理五官等面部主要特征.

单帧图像的人脸三维重建问题上, 同样经常遇到不甚理想的拍摄条件. 人脸三维重建并不仅应用专业设备拍摄的清晰图像上, 也应用在各种日常场合, 包括非专业人员使用手机和一些廉价摄像头拍摄的图像, 这些图像由于人员操作和设备性能问题也会出现模糊. 现有的人脸三维重建方法处理模糊图像大多会得到不理想的效果.

Ren 等^[1]提出一种使用三维人脸重建获取模糊人脸图像的先验信息进行人脸图像去模糊的方法, 首先对模糊人脸图像使用参数化 3DMM(3D morphable models)方法进行人脸三维重建, 然后将三维重建得到的模型在原图像的相机参数上渲染, 最后将渲染的结果和 3DMM 参数加入图像去模糊的卷积神经网络输入中, 得到较清晰的人脸图像. 但该方法忽略了模糊图像的人脸重建结果可能是低质量的, 而低质量的三维重建结果会诱导低质量的去模糊结果. 实际上, 对于模糊图像, 三维重建和去模糊这 2 项任务是可以互相促进的, 既可以使用三维重建的结果优化去模糊任务, 也可以使用去模糊的结果优化三维重建任务.

本文提出一种人脸图像去模糊和三维重建的联合优化方法, 采用交替优化方法实现. 输入一幅模糊人脸图像, 交替重复以下步骤: (1) 对图像使用人脸三维重建方法进行重建, 更新三维人脸模型, 提取三维先验信息; (2) 将图像和三维先验信息输入去模糊神经网络消除模糊, 更新图像, 其中, 去模糊神经网络在每轮迭代中分别进行训练. 重复上述步骤合适次数后, 输出人脸图像和人脸三维模型. 实验结果表明, 本文方法能够输出清晰的人脸图像和良好的人脸三维模型, 其质量比现有方法得到了显著提升.

1 相关工作

1.1 图像去模糊

1.1.1 图像模糊的数学模型与传统方法

图像模糊包括运动模糊、相机抖动模糊和失焦模糊等, 表现现象为被拍摄物体的各点没有成像到一个点上, 而是分布在点的一个邻域内. 可以用统一的数学模型对其进行描述: 假设存在一幅对应于模糊图像的潜在的清晰图像, 那么模糊图像可以视作清晰图像经过一个映射得到, 此映射将清晰图像的每个像素按照一定规律加权叠加到该像素的一个邻域上, 称为该点处的点扩散函数(point spread function, PSF)^[2], 整个图像各点的 PSF 构成 1 个 PSF 场.

因此模糊图像可以描述为一个卷积过程

$$g(x, y) = f(x, y) * k(x, y) + n(x, y) \quad (1)$$

其中, g 表示模糊图像; f 表示潜在的清晰图像; n 表示随机噪声; k 表示 PSF. 则去模糊问题表达为求解 f , 其中, g 已知, n 未知. k 已知时, 称为非盲去模糊; 否则, 称为盲去模糊. 无论 PSF 是否已知, 随机噪声总是未知的, 于是问题没有唯一解, 是一个病态问题.

维纳滤波^[3]是一种基本图像去模糊方法, 它在已知模糊核的情况下求解一个逆卷积核, 使其作用在模糊图像上与潜在清晰图像的均方误差最小.

非盲去模糊方法基于去噪^[4]和逆卷积^[5]; 而盲去模糊方法首先对模糊核进行估计, 然后进行逆卷积和去噪. Zhang 等^[6]提出一种模糊核估计方法, 使用贝叶斯后验估计进行筛选, 排除不良像素点. Chen 等^[7]提出一种噪声模型来拟合真实世界的图像噪声.

1.1.2 使用深度学习的通用图像去模糊方法

目前, 使用深度学习的方法解决去模糊问题取得了较好的效果, 主要分为基于数学模型的方法和端到端的方法.

基于数学模型的方法从上述模糊数学模型出发, 使用神经网络估计图像的模糊核和噪声, 减去噪声后作逆卷积进行去模糊. 已有很多去噪的深度学习方法, 关注点在于对模糊核的估计. Sun 等^[8]提出一种基于微元的模糊核估计, 即将图像分为若干小的 patch, 假定模糊在每个 patch 上是足够均匀的, 然后在每个 patch 上估计均匀的模糊核, 再在整个图片上拟合出模糊场; Gong 等^[9]提出一种直接在全局估计运动场还原出整个图像的 PSF 场的方法.

目前, 端到端的深度学习的方法取得了更好的效果, 这类方法不再关注模糊的数学模型, 而是直

接训练神经网络拟合从模糊图像到清晰图像的映射. 端到端的深度学习方法最基础的结构为编码-解码框架, 它使用若干卷积层将输入的模糊图像升维, 提取图像的高级语义信息, 即编码; 再使用若干卷积层进行降维, 从提取到的信息中恢复出清晰图像, 即解码. 该结构能够保持图像的整体结构基本不变, 并优化调整其局部细节, 适用于去模糊任务, 目前的去模糊方法多是在这一结构上进行重复、拓展和优化. Nah 等^[10]提出一种尺度递归网络, 从较粗尺度逐步恢复图像, 直到获得完整分辨率的清晰图像; 进一步, Tao 等^[11]提出尺度循环网络(scale-recurrent network, SRN), 通过跨尺度共享网络权值来去除模糊, 在从低尺度向高尺度递进还原图像的过程中, 使中间层共享权值, 达到通用去模糊中较优的效果.

1.1.3 人脸图像去模糊方法

人脸具有心理学意义上显著的几何特征, 如眼、耳、鼻等, 因此, 良好的人脸图像去模糊方法应当能够在结果上体现清晰的上述特征. 由于通用去模糊方法很难做到这一点, 因此其在人脸图像上表现不佳, 由此发展出多种针对性的人脸图像去模糊方法.

Chrysos 等^[12]提出了一个两步结构, 先恢复低频信息, 再恢复高频信息, 将输出图像约束到自然图像域上. Lai 等^[13]利用智能手机的多镜头拍摄功能, 在不同相机参数下同时拍摄图像来辅助人脸图像去模糊. Shen 等^[14]使用语义信息先验辅助人脸图像去模糊, 首先训练一个语义生成网络, 对输入图像进行区域划分并赋予语义; 然后将语义作为先验信息加入去模糊任务, 训练去模糊网络从输入图像和语义图像向清晰图像拟合. Ren 等^[1]使用三维先验辅助人脸图像去模糊, 首先通过一个人脸三维重建网络对模糊人脸图像进行三维重建, 然后再渲染重建得到的三维人脸模型, 最后将再渲染得到的图像和三维重建相关参数作为三维先验信息加入去模糊任务, 训练去模糊网络从输入图像和三维先验向清晰图像拟合. Jung 等^[15]使用深度先验辅助人脸图像去模糊, 提出一种包含人脸形状和纹理的深度先验信息, 并构造先验提取器提取先验信息, 以辅助去人脸图像模糊.

1.2 人脸三维重建

1.2.1 3DMM 方法

人脸三维重建指从二维图像中构造出相应的人脸三维模型及其纹理. Blanz 等^[16]提出一种受到广泛使用的 3DMM 方法, 核心思想是将所有的人

脸模型视作由若干正交基加权线性组合得到的, 进而可将人脸的三维形状与纹理参数化. 该方法首先预设一个标准人脸三维模型, 以固定的拓扑结构描述人脸; 然后对大量的人脸模型进行主成分分析降维, 得到占主要成分的分量以及人脸三维形状和纹理的各一组正交基. Gerig 等^[17]在 3DMM 方法中又增加了关于表情等维度的基. 这样任意人脸模型都可以由这些正交基上的系数描述, 人脸图像重建三维模型的问题就被转化为拟合这些正交基的系数, 以及拍摄图像的相机坐标和相机参数.

1.2.2 使用深度学习求解 3DMM 的方法

Tran 等^[18]首先提出使用深度学习方法求解 3DMM 系数, 其中使用残差神经网络(residual neural network, ResNet)拟合人脸图像到 3DMM 系数的映射. 之后, 使用深度学习求解 3DMM 的方法得到了广泛的发展. Deep3DFace 方法^[19]是一种弱监督求解 3DMM 的方法, 其将重建得到的三维模型在输入图像的相机参数和相机坐标下进行可微渲染, 并计算再渲染图像和输入图像之间的误差, 将其作为训练神经网络的损失函数.

由于 3DMM 模型本身是低维的, 对于人脸细部特征表达能力有限, 因此又发展出了基于 3DMM 的精细重建方法. 蒋罗^[20]提出一种双线性人脸模型, 扩大了参数空间和表达能力, 并加入对于人脸较小尺度变形进行描述的校正场, 能够重建出具有更多细节的人脸模型. 郭玉东^[21]提出一种基于 3DMM 的精细重建方法, 通过对三维人脸模型添加像素级的位移量, 能够有效地重建诸如皱纹和酒窝等精细人脸结构. Wood 等^[22]使用稠密人脸特征点改善参数的拟合效果, 通过识别和定位大量面部特征点, 为求解 3DMM 参数提供更多约束, 增强了拟合的稳定性, 能够得到良好的效果. Lei 等^[23]提出一种分层表示网络, 将人脸解耦为低频几何形状、中频细节和像素级的高频细节, 由粗到细地重建人脸三维模型.

针对低质量图像重建, Tiwari 等^[24]提出一种抗遮挡神经网络求解有缺损的人脸图像的三维重建问题, 使用基于上下文的神经网络弥补人脸图像的缺损部分, 并重建出完整三维人脸模型.

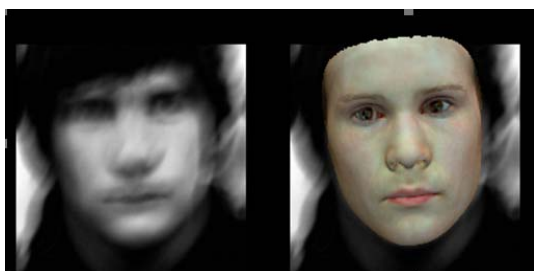
2 本文方法

2.1 联合优化思想

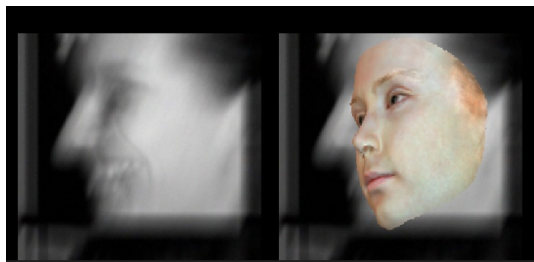
2.1.1 三维先验的有效性和缺陷

与其他去模糊方法相比, Ren 等^[1]首次提出将

人脸三维重建引入到去模糊任务中, 在人脸图像上取得了显著进展, 这是因为其所使用的 3DMM 人脸三维重建方法引入对去模糊任务来说缺失的重要信息, 对去模糊任务施加了强约束. 在任何模糊或者不模糊的人脸图像上, 重建得到的三维人脸模型具有相同的拓扑结构, 因而在五官等主要特征处具有明确的边缘信息, 并且具有清晰的纹理. 因此, 当重建得到的模型被再渲染为图像时, 一定能够得到清晰的人脸图像, 并且人脸图像在五官等主要特征处有明显的尖锐边缘, 如图 1a 所示. 即使这样重建得到的模型相对于潜在的清晰图像存在误差, 但仍然提供了模糊图像相对清晰图像所缺失的关于图形边缘和图像梯度的重要信息.



a. 模糊程度低的人脸图像



b. 模糊程度高的人脸图像

图 1 模糊图像上重建得到的模型

但是, 仅使用单步人脸三维重建作为先验信息进行去模糊是不足的. 大部分情况下, 该方法能够在模糊图像上重建得到三维人脸模型, 并且其位置大致相符, 纹理大致相似; 但细节处相对于清晰图像有误差, 而且并不总能得到良好的三维重建结果. 如图 1b 所示, 会出现人脸位置出现较大位移, 人脸结构不能吻合等情况. 这是由于人脸三维重建方法通常并不针对模糊图像进行优化, 其数据集通常并不包含模糊图像, 模型结构也没有对模糊图像进行特殊优化等造成的. 在这种情况下, 低质量的三维先验会诱导低质量、甚至错误的去模糊结果.

2.1.2 三维重建和去模糊任务的联合优化

为了弥补单步三维先验的不足, 需要提升对

于模糊人脸图像进行三维重建的质量. 因为模糊图像存在信息的缺失, 尤其是五官等重要结构在模糊后失去了明显的结构特征, 所以直接设计算法对模糊人脸图像进行三维重建是困难的. 特别地, 作为大多数 3DMM 参数拟合方法的前置, 人脸特征点提取就很难在模糊人脸图像上进行. 但是, 如果先对输入图像进行一定程度的模糊消除, 就能够得到良好的人脸三维重建效果; 并且消除模糊的效果越好, 得到的人脸三维重建效果也就越好.

本文发现, 人脸图像去模糊需要人脸三维重建引入信息, 而模糊图像上的人脸三维重建又需要人脸图像去模糊的预处理; 换言之, 人脸三维重建能对人脸图像去模糊进行优化, 人脸图像去模糊也能对模糊图像的人脸三维重建进行优化. 这 2 项任务能够互相为对方提供进一步优化所需的信息, 构成了一个正反馈的闭环. 因此, 三维重建任务和去模糊任务是高度耦合的, 不应当分别独立进行, 而应当联合起来进行优化; 其输入为模糊的人脸图像, 输出同时包括消除模糊的人脸图像和重建得到的人脸三维模型. 优化目标需要同时包含去模糊误差和三维重建误差.

2.2 交替优化方法

2.2.1 概述

为了实现联合优化的思想, 需要引入一种机制在三维重建任务和去模糊任务之间传递信息. 由于三维重建使用的神经网络与去模糊使用的神经网络在结构上非常不同, 很难设计一个大规模神经网络来同时执行这 2 项任务.

受到交替方向乘子法^[25]的启发, 本文提出交替优化方法实现上述联合优化, 即轮流执行 2 项任务, 分别降低每个方向上的误差. 使用人脸三维重建作为先验辅助人脸图像去模糊, 其效果优于没有引入先验时的结果; 使用经过去模糊预处理的人脸图像进行三维重建, 其效果优于没有经过预处理时的结果. 这保证了分别执行 2 项任务时总误差一定下降, 因此能够稳定收敛.

交替优化方法将整个优化过程分解为若干轮迭代, 每轮迭代中分别执行 2 项任务. 图 2 所示为该方法的框架, 其中, 左下角为输入的模糊人脸图像, 上方虚线框内为三维重建分支, 每步经过一个 3DMM 参数拟合神经网络; 下方虚线框内为去模糊分支, 每步经过一个去模糊神经网络. 每轮迭代中, 输入图像首先进入三维重建分支, 得到三维人脸模型, 获取三维先验信息; 然后进入去模糊分支, 由模糊图像和三维先验得到消除模糊的图像.

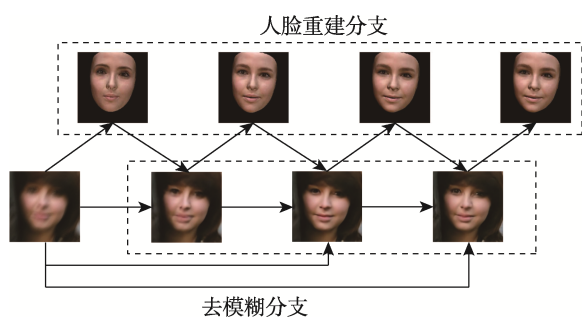


图2 交替优化方法框架

之后进入下一轮迭代, 直至得到足够良好的效果。

2.2.2 去模糊分支

去模糊网络的结构如图3所示, 选用 Res-UNet^[26]结构, 该结构被证明在去模糊任务是显著有效的^[1]。本文使用深度为4层, 每层2个残差块的Res-UNet; 所有卷积使用 5×5 卷积核; 损失函数为输出图像和目标图像的均方误差; 使用 Adam

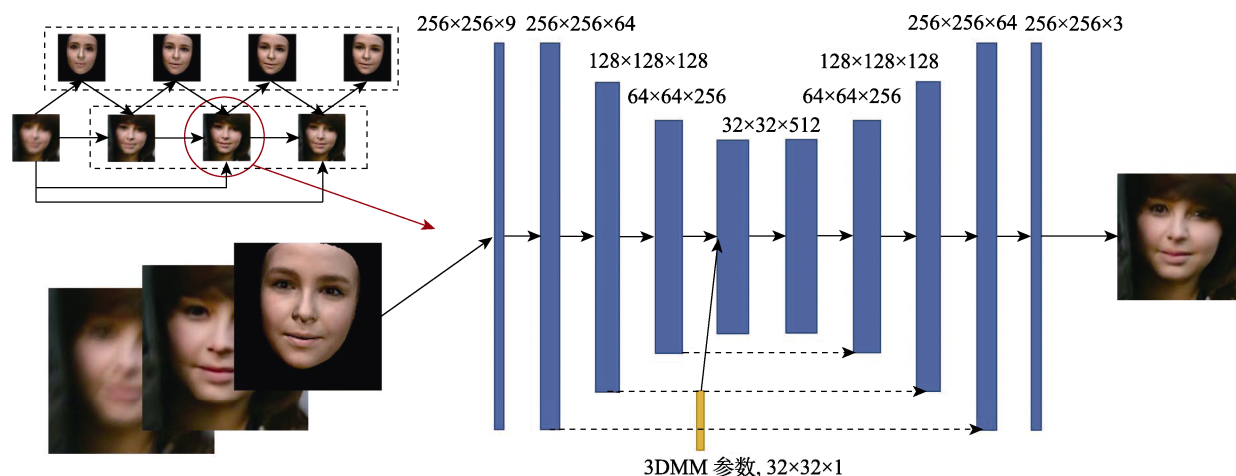


图3 图像去模糊分支的结构

3 实验与结果分析

3.1 数据集

目前尚无由模糊人脸图像组成的成熟数据。由于模糊图像由清晰图像卷积模糊核后叠加噪声形成, 因此可用这种方法构造所需数据集^[14]。本文实验数据集为 AFLW 数据集^[27]的 21 080 幅人脸图像, 使用 MTCNN(multi-task cascaded convolutional neural networks)^[28]进行人脸识别并截取正方形范围的人脸区域, 缩放至 256×256 像素大小。根据 Boracchi 等^[29]和 Shen 等^[14]的方法, 使用从相机运动轨迹生成的 20 000 个由模糊核来生成模糊图像。共生成 1 000 000 幅图像作为基本训练集, 约 50 000 幅图像作为基本验证-测试集。

实验中, 基于设备性能限制, 随机选取基本训

练集训练, 设学习率为 0.000 1, 学习率为指数衰减, 衰减率为 0.98。

迭代过程中, 本文将原始模糊图像与迭代中的去模糊图像同时输入去模糊网络, 避免迭代过程中每步引入的信息的误差积累之后偏离原始图像。以原始图像作为校正输入, 可以将去模糊结果约束在原始图像附近, 避免出现误差的积累。

2.2.3 三维重建分支

目前, Deep3DFace 方法^[19]在基于 3DMM 进行人脸三维重建的方法中表现最优, 本文采用其作为人脸三维重建分支使用的方法。随着输入图像逐步清晰, Deep3DFace 方法^[19]可以给出足够良好的三维重建结果, 因此无须对该方法做进一步改进。三维重建分支使用 Deep3DFace^[19]所述的 ResNet50 网络。其中, 最后一层全连接层修改为 239 个通道以对应 3DMM 系数。

练集中 10 000 幅图像作为实验训练集, 基本验证-测试集中 1 000 幅图像作为验证集, 10 000 幅图像作为测试集。

3.2 神经网络训练过程中的一些观察

实验过程中, 观察到如下现象: (1) 迭代 3 次以后获得良好效果, 并且提升幅度出现较大衰减。(2) 随着训练的进行, 在一定进度后, 损失函数在验证集和训练集上都仍在下降, 但在验证集上显著高于训练集。这是潜在的过拟合现象。通常神经网络训练以损失函数不再在验证集上下降为过拟合的标志, 对于迭代方法, 情况有所不同。在进入下一次迭代时, 训练集和验证集应当具有一致的“去模糊程度”。如果神经网络在训练集和验证集上的表现差异过大, 那么在进入下一次迭代时训练集就不再具有代表性, 神经网络也不可能具有

泛化能力.

表 1 所示为 3 次迭代步骤中不同步骤的训练集和验证集上损失函数的对比, 显示了迭代中训练集和验证集是如何分化的. 可以看出, 在 3 次迭代中, 训练集和验证集上可以分别使损失函数下降较多的 epoch, 但只能在较少的 epoch 内保持损失函数具有较小的差距.

表 1 训练集和验证集损失函数对比

训练进度		损失函数	
迭代步骤	epoch	训练集	验证集
第 1 次	15 epoch	0.009 161	0.010 246
	25 epoch	0.006 497	0.008 917
第 2 次	20 epoch	0.007 455	0.008 650
	60 epoch	0.004 051	0.007 376
第 3 次	100 epoch	0.001 614	0.006 650

因此对实验做出如下调整: (1) 迭代次数固定为 3 次. (2) 前 2 次迭代为避免潜在的过拟合, 在训练到损失函数在训练集和验证集上出现分化时停止; 第 3 次迭代因不需再考虑后续效应, 训练到验证集上损失函数不再下降时停止.

3.3 对比实验

3.3.1 实验设计

SRN^[11]是目前较优且较常用的通用去模糊方法, 本文以该方法作为通用去模糊对比方法.

目前, 3DFaceDeblur 方法^[1]达到人脸图像去模糊的最优效果, 以该方法作为人脸图像去模糊对比方法. 3DFaceDeblur 方法^[1]的开源代码有少量缺失, 本文以其开源代码的结构为基础进行复现, 使用与本文方法实验相同的训练集、验证集和测试集, 使用开源代码提供的训练方法进行训练.

使用常用的图像相似度指标峰值信噪比(peak signal-to-noise ratio, PSNR)和结构相似度(structural similarity, SSIM)评价图像去模糊的结果. PSNR 直接衡量 2 幅图像的逐点误差, 其值越高, 说明图像相似度越高. Wang 等^[30]提出 SSIM 作为更好地衡量 2 幅图像的相似程度的指标, 其比 PSNR 更接近人类对图像相似程度的感知.

3.3.2 迭代效果对比

在训练集、验证集和测试集上, 每次迭代的去模糊效果的平均 PSNR 和 SSIM 如表 2 所示. 其中, 衡量每次迭代的效果时, 使用该次迭代中训练到验证集损失函数不再下降时的参数作为代表, 表示该步骤在不考虑潜在过拟合的情况下能得到的

表 2 去模糊的迭代分步结果对比

数据集	训练进度	PSNR/dB	SSIM
输入图像		24.88	0.731
训练集	第 1 次迭代(25 epoch)	28.33	0.828
	第 2 次迭代(60 epoch)	33.83	0.910
	第 3 次迭代(100 epoch)	34.42	0.915
验证集	第 1 次迭代(25 epoch)	27.82	0.820
	第 2 次迭代(60 epoch)	29.08	0.853
	第 3 次迭代(100 epoch)	29.25	0.856
测试集	第 1 次迭代(25 epoch)	27.77	0.818
	第 2 次迭代(60 epoch)	29.01	0.851
	第 3 次迭代(100 epoch)	29.23	0.855

最好效果.

从表 2 可以看出: (1) 每次迭代在训练到验证集上的损失函数不再下降时, 训练集上的效果和验证集上的效果已经产生了明显差异, 这是一种潜在的过拟合现象, 验证了表 1 所示的训练集和验证集产生分化的问题; 如果以此结果继续迭代, 则会失去泛化能力. 因此, 本文方法中, 以保持验证集和训练集损失函数一致的参数进入下一次迭代, 是合理且有效的. (2) 第 1 次迭代相对输入图像有较大提升, 实际上验证了 3DFaceDeblur 方法^[1]的有效性; 第 2 次迭代仍有显著提升, 说明本文方法能够达到更佳的效果; 第 3 次迭代仍有较小提升, 但相对前 2 次迭代有较大衰减, 说明此时迭代已经基本收敛, 故采用 3 次迭代是合适的.

图 4 所示为 4 幅模糊图像在每次迭代下的去模糊效果, 其中, 前 3 行为训练集中的合成图像, 第 4 行为现实中拍摄的图像. 可以看出, 从人类视觉上看, 去模糊效果符合实验数据; 每次迭代得到的图像质量较上一次迭代均有提升, 至第 3 次时已经接近清晰图像. 实验结果还表明, 本文方法虽然在人工合成的图像上训练, 对现实拍摄的图像上仍然有效.

3.3.3 去模糊效果对比

在测试集上, 将本文方法与 SRN 方法^[11]和 3DFaceDeblur 方法^[1]进行实验, 结果如表 3 所示. 可以看出, SRN^[11]作为通用去模糊方法, 在人脸去模糊问题上效果较差, 说明人脸图像去模糊这一任务需要进行特殊优化; 3DFaceDeblur 方法^[1]取得了比较好的效果; 而本文方法效果最佳.

将图 4 中的 4 幅模糊图像在多种去模糊方法下进行实验, 结果如图 5 所示. 可以看出, 从人类视觉上看, SRN 方法^[11]的结果具有较大瑕疵; 3DFace-

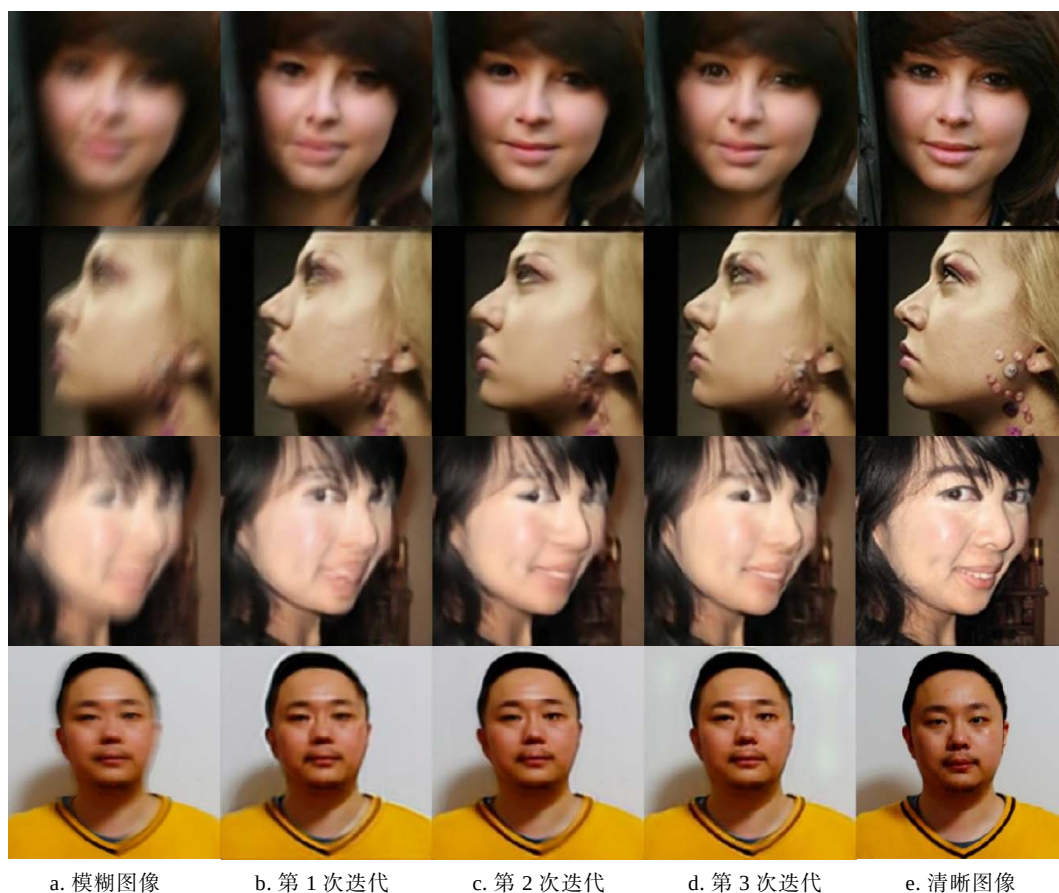


图 4 人脸图像去模糊的迭代分步效果

表 3 3 种去模糊方法结果对比

方法	PSNR/dB	SSIM
输入模糊图像	24.88	0.731
SRN ^[11]	24.49	0.737
3DFaceDeblur ^[11]	27.38	0.833
本文	29.23	0.855

Deblur 方法^[11]的结果具有较好的心理学观感; 本文方法的结果与清晰图像最接近, 能更好地还原人脸细节特征和表情.

3.3.4 三维重建效果对比

为在测试集上, 每步迭代的三维重建效果的平均 PSNR 和 SSIM 对比如表 4 所示. 由于目前尚未有专门针对模糊图像进行优化的三维重建方法, 因此本文使用目前基于 3DMM 的人脸三维重建方法中较优的 Deep3DFace 方法^[19]作为对比, 其中, 输入模糊图像的三维重建效果也代表了 Deep3DFace 方法^[19]的效果. 可以看出, 第 1 次和第 2 次迭代都取得了显著提升, 而第 3 次迭代提升有限, 说明 3 次迭代是合适的; 与目前的常用方法相比, 本文方法的效果取得了大幅提升.

表 4 三维重建效果的迭代分步比较

迭代步骤	PSNR/dB	SSIM
输入模糊图像(直接应用 Deep3DFace ^[19])	22.78	0.848
第 1 次迭代(20epoch)	25.64	0.899
第 2 次迭代(60epoch)	26.99	0.917
第 2 次迭代(100epoch)	27.16	0.918

图 6 给出了上述将图 4 中的 4 幅模糊图像在每步迭代下的三维重建效果. 可以看出, 从人类视觉上看, 本文方法较对比方法能更好地还原人脸细节特征和表情; 随着迭代步骤的增加, 本文方法的结果接近清晰图像上的重建结果.

4 结 语

本文提出一种人脸图像去模糊和三维重建的联合优化方法, 对人脸三维重建和人脸图像去模糊 2 项任务交替进行优化, 并互相为对方提供信息. 实验结果表明, 迭代过程中, 图像和模型质量都有提升, 迭代方法是有效的; 与常用方法相比, 最终输出的人脸图像和三维模型的质量都取得显

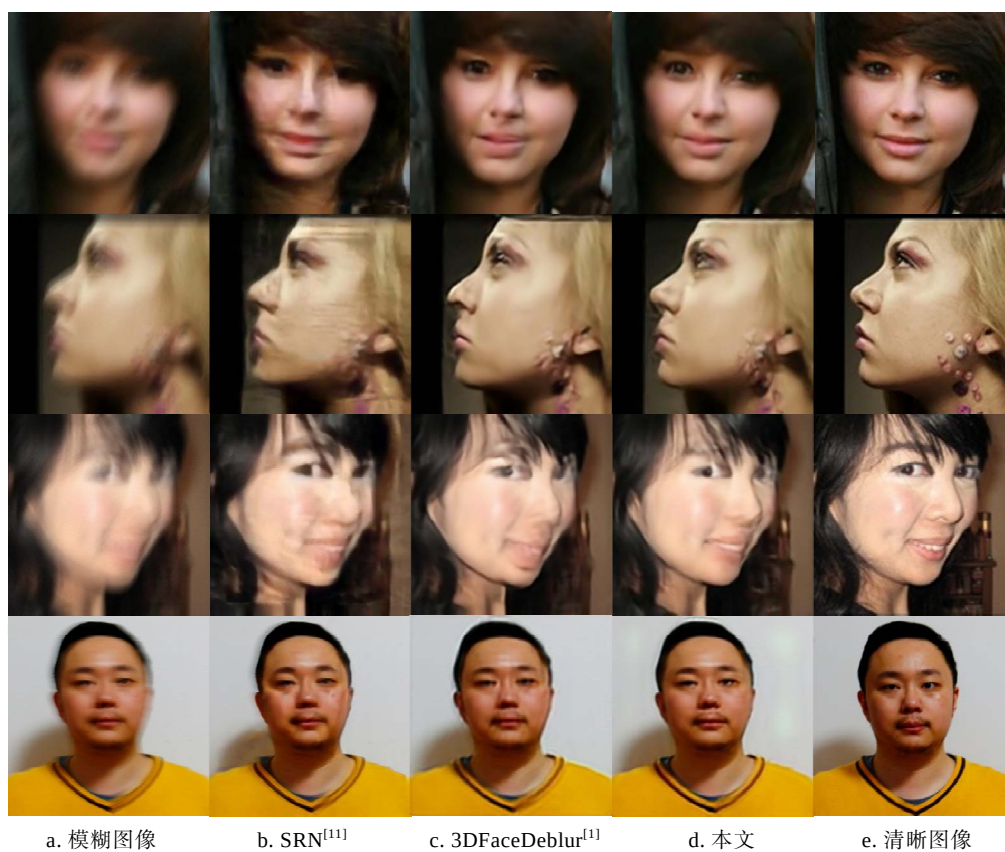


图 5 多种人脸图像去模糊方法的效果对比

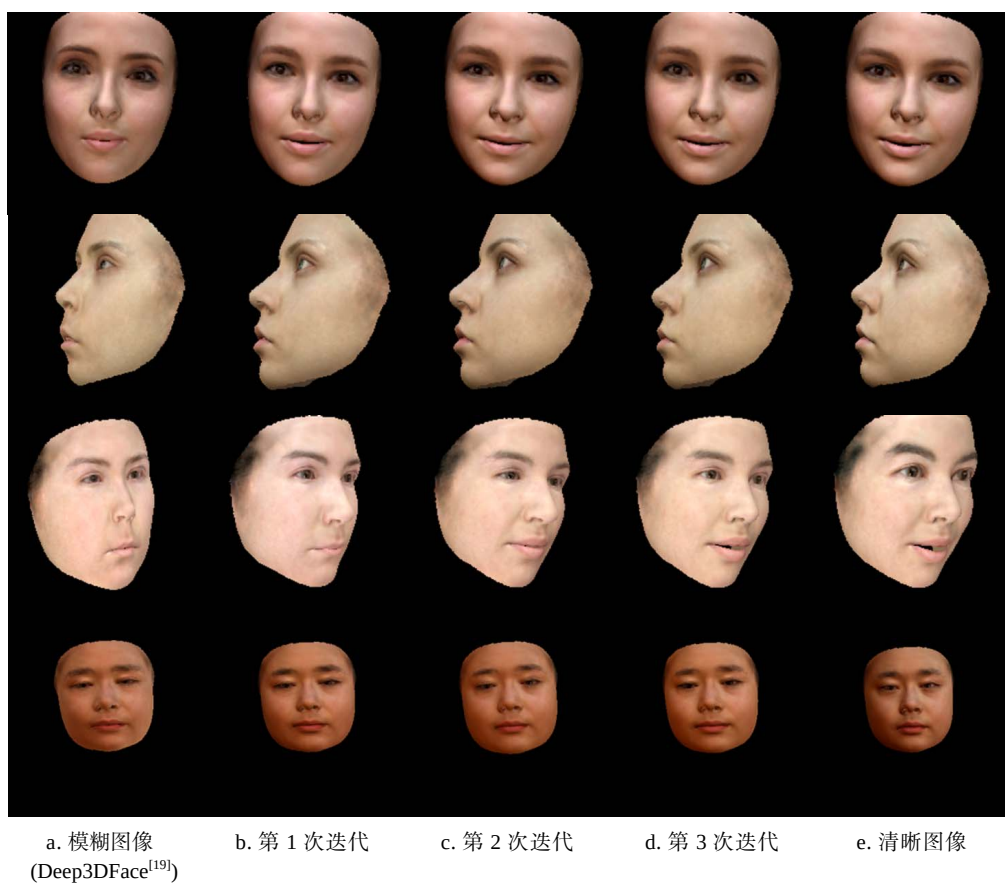


图 6 人脸三维重建的迭代分步效果

著提升. 表明本文方法同时在2项任务上取得了更好的效果.

目前, 本文实验基于图像仅包含单张人脸的假设而设计, 且对输入图像的分辨率作强制缩放. 未来将会对本文方法进行扩展, 适配多种分辨率和复杂内容图像的输入.

因为本文所基于的三维先验方法已经推广应用于人脸超分辨率和人脸图像复原等方面, 所以本文提出的联合优化方法在这些方面也会有类似的推广应用.

参考文献(References):

- [1] Ren W Q, Yang J L, Deng S Y, *et al.* Face video deblurring using 3D facial priors[C] //Proceedings of the IEEE/CVF International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2019: 9387-9396
- [2] Gonzalez R C, Woods R E. Digital image processing[M]. Ruan Qiuqi, Ruan Yuzhi, trans. 2nd ed. Beijing: Publishing House of Electronics Industry, 2003(in Chinese)
(Gonzalez R C, Woods R E. 数字图像处理[M]. 阮秋琦, 阮宇智, 译. 2 版. 北京: 电子工业出版社, 2003)
- [3] Szeliski R. Computer vision: algorithms and applications[M]. 2nd ed. Cham: Springer, 2022
- [4] Krishnan D, Fergus R. Fast image deconvolution using hyper-Laplacian priors[C] //Proceedings of the 22nd International Conference on Neural Information Processing Systems. New York: Curran Associates Inc, 2009: 1033-1041
- [5] Zoran D, Weiss Y. From learning models of natural image patches to whole image restoration[C] //Proceedings of International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2011: 479-486
- [6] Zhang M N, Fang Y Y, Ni G X, *et al.* Pixel screening based intermediate correction for blind deblurring[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2022: 01-09
- [7] Chen L, Fang F M, Lei S, *et al.* Enhanced sparse model for blind deblurring[C] //Proceedings of the 16th European Conference on Computer Vision. Heidelberg: Springer, 2020: 631-646
- [8] Sun J, Cao W F, Xu Z B, *et al.* Learning a convolutional neural network for non-uniform motion blur removal[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2015: 769-777
- [9] Gong D, Yang J, Liu L Q, *et al.* From motion blur to motion flow: A deep learning solution for removing heterogeneous motion blur[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2017: 3806-3815
- [10] Nah S, Kim T H, Lee K M. Deep multi-scale convolutional neural network for dynamic scene deblurring[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2017: 257-265
- [11] Tao X, Gao H Y, Shen X Y, *et al.* Scale-recurrent network for deep image deblurring[C] //Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Los Alamitos: IEEE Computer Society Press, 2018: 8174-8182
- [12] Chrysos G G, Favaro P, Zafeiriou S. Motion deblurring of faces[J]. International Journal of Computer Vision, 2019, 127 (6-7): 801-823
- [13] Lai W S, Shih Y, Chu L C, *et al.* Face deblurring using dual camera fusion on mobile phones[J]. ACM Transactions on Graphics, 2022, 41(4): Article No.148
- [14] Shen Z Y, Lai W S, Xu T F, *et al.* Deep semantic face deblurring[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2018: 8260-8269
- [15] Jung S H, Lee T B, Heo Y S. Deep feature prior guided face deblurring[C] //Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. Los Alamitos: IEEE Computer Society Press, 2022: 884-893
- [16] Blanz V, Vetter T. A morphable model for the synthesis of 3D faces[C] //Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques. New York: ACM Press, 1999: 187-194
- [17] Gerig T, Morel-Forster A, Blumer C, *et al.* Morphable face models-an open framework[C] //Proceedings of the 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018). Los Alamitos: IEEE Computer Society Press, 2018: 75-82
- [18] Tran A T, Hassner T, Masi I, *et al.* Regressing robust and discriminative 3D morphable models with a very deep neural network[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2017: 1493-1502
- [19] Deng Y, Yang J L, Xu S C, *et al.* Accurate 3D face reconstruction with weakly-supervised learning: From single image to image set[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Los Alamitos: IEEE Computer Society Press, 2019: 285-295
- [20] Jiang Luo. 3D Face reconstruction and face recognition[D]. Hefei: University of Science and Technology of China, 2019(in Chinese)
(蒋罗. 三维人脸重建与人脸识别[D]. 合肥: 中国科学技术大学, 2019)
- [21] Guo Yudong. 3D Face reconstruction and its applications[D]. Hefei: University of Science and Technology of China, 2021(in Chinese)
(郭玉东. 三维人脸重建及其应用[D]. 合肥: 中国科学技术大学, 2021)
- [22] Wood E, Baltrušaitis T, Hewitt C, *et al.* 3D face reconstruction with dense landmarks[C] //Proceedings of the 17th European Conference on Computer Vision. Heidelberg: Springer, 2022: 160-177
- [23] Lei B W, Ren J Q, Feng M Y, *et al.* A hierarchical representation network for accurate and detailed face reconstruction from

- in-the-wild images[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2023: 394-403
- [24] Tiwari H, Kurmi V K, Venkatesh K S, *et al.* Occlusion resistant network for 3D face reconstruction[C] //Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. Los Alamitos: IEEE Computer Society Press, 2022: 297-306
- [25] Boyd S, Parikh N, Chu E, *et al.* Distributed optimization and statistical learning via the alternating direction method of multipliers[J]. Foundations and Trends in Machine Learning, 2011, 3(1): 1-122
- [26] Xiao X, Lian S, Luo Z M, *et al.* Weighted Res-UNet for high-quality retina vessel segmentation[C] //Proceedings of 9th International Conference on Information Technology in Medicine and Education (ITME). Los Alamitos: IEEE Computer Society Press, 2018: 327-331
- [27] Köstinger M, Wohlhart P, Roth P M, *et al.* Annotated facial landmarks in the wild: a large-scale, real-world database for facial landmark localization[C] //Proceedings of IEEE International Conference on Computer Vision Workshops (ICCV workshops). Los Alamitos: IEEE Computer Society Press, 2011: 2144-2151
- [28] Zhang K P, Zhang Z P, Li Z F, *et al.* Joint face detection and alignment using multitask cascaded convolutional networks[J]. IEEE Signal Processing Letters, 2016, 23(10): 1499-1503
- [29] Boracchi G, Foi A. Modeling the performance of image restoration from motion blur[J]. IEEE Transactions on Image Processing, 2012, 21(8): 3502-3517
- [30] Wang Z, Bovik A C, Sheikh H R, *et al.* Image quality assessment: from error visibility to structural similarity[J]. IEEE Transactions on Image Processing, 2004, 13(4): 600-612