融合生成隐编码库先验的人脸去模糊方法

梁远[1)], 姜雨馨[1)], 何盛烽[1,2)*]

[1)] (华南理工大学计算机科学与工程学院 广州 510006)
[2)] (School of Computing and Information Systems, Singapore Management University Singapore 178903)
(shengfenghe@smu.edu.sg)

摘 要: 对抗式神经网络的隐编码库包含了图像中不同尺度的细节特征, 为更好地结合隐编码库和图像的不同尺度特征来为图像去除运动模糊, 文中通过多层卷积网络提取模糊图像的多尺度特征, 并由隐编码库中不同尺度的隐向量指导产生多尺度的解码特征, 该解码特征在隐编码库的统一指导下生成去除模糊后的图像. 这种融合多尺度特征而非单一尺度特征的去模糊方法在更多细节上保证了人脸模糊图像的重建和复原. 与当前同样使用隐编码库进行人脸去模糊的方法比较, 文中方法取得了更好的效果, 在 PSNR 指标上提升了 30%, 在 SSIM 指标上提升了 23%.

关键词: 人脸去模糊; 隐编码库; 生成对抗网络
中图法分类号: TP391.41 **DOI:** 10.3724/SP.J.1089.2024.20147

Blind Face Deblurring via Latent Generative Priors

Liang Yuan[1)], Jiang Yuxin[1)], and He Shengfeng[1,2)*]

[1)] (*School of Computer Science and Engineering, South China University of Technology, Guangzhou* 510006)
[2)] (*School of Computing and Information Systems, Singapore Management University, Singapore* 178903)

Abstract: The latent code of generative adversarial networks contains different scales of detail features in images. To better combine the different scale characteristics of degraded images and the rich information from latent code to perform face image deblurring, we extract the multi-scale features of the blurred face images through a multi-layer convolution network, which is further guided by the latent code of generative adversarial networks to generate the decoded features, which is then further guided by the latent code to generate deblurred images. Through this multi-scale feature-guided scheme rather than single-scale generations, our method ensures the reconstruction and restoration of blurred images in more detail. Compared to the current methods, which also use the latent code of generative adversarial networks, our method achieves better performance, improving 30% on the metric of PSNR and 23% on the metric of SSIM.

Key words: face deblurring; latent code bank; generative adversarial networks

成像设备在采集图像时存在一定的曝光时间, 会导致曝光期间拍摄物体与成像设备可能产生相对位移, 根据相对位移的大小、运动轨迹将产生不同程度的图像退化, 即图像运动模糊. 运动模糊导致图像内容模糊不清, 不仅严重影响人类视觉观赏效果, 还将直接降低人脸识别、行为理解这类高阶视觉分析任务的运行结果.

在目前众多的主流图像去模糊方法中, 可以根据是否已知模糊核分为图像非盲去模糊与盲去模糊方法 2 大类. 图像非盲去模糊方法往往假定已

收稿日期: 2023-02-10; 修回日期: 2023-08-15. **基金项目**: 国家自然科学基金(61972162). **梁远**(1989—), 男, 博士研究生, 主要研究方向为计算机视觉、深度学习; **姜雨馨**(1998—), 女, 硕士研究生, 主要研究方向为计算机视觉、深度学习; **何盛烽**(1986—), 男, 博士, 副教授, 博士生导师, CCF 高级会员, 论文通信作者, 主要研究方向为计算机视觉、图像处理.

知图像模糊核，进而将主要精力集中于图像去模糊的复原部分研究，而对于真实情况未知运动模糊核或者复杂运动模糊核干扰的模糊图像往往处理效果不佳. 相对于非盲卷积方法，盲卷积方法需要额外预测未知运动轨迹加持下的运动模糊核. 早期的盲卷积方法利用基于贝叶斯的优化方法迭代地为每个图像块生成局部最优的模糊核[1]. 在这类盲去模糊方法中，预测模糊核的数量以及模糊核精度会很大程度地影响去模糊效果. 当预测模糊核与真实模糊核误差精度较大时，该类盲去模糊方法不能很好地区分应去除的抖动与应保留细节的特征差异，进而产生严重振铃现象，极大地降低了去模糊图像的视觉效果. 为了减少估计模糊核与真实核之间的差异，一些学者通过图像统计特性设计手工先验项的方式约束模糊核计算的解空间[2-6]. 然而，由于上述方法引入较复杂的数据保真项与先验项，且手工先验项设计复杂，导致了优化求解过程的复杂度更高，因此不利于对图像纹理等信息的复原.

受益于深度学习技术的发展，不断涌现出基于深度学习技术的图像去模糊方法，主要包括端到端的学习模式以及基于先验的深度图像去模糊方法 2 大类. 基于端到端的学习模式本质上是利用神经网络直接学习模糊图像与清晰图像的映射关系[7-10]，但其需要大量的训练数据，且神经网络模型的参数量较大，训练时间较长，容易出现过拟合问题. 为此，有学者通过将运动模糊的先验知识引入基于深度学习的图像去模糊方法中[11-15]，在一定程度上克服了网络模型容易过拟合的问题. 但是，传统的人工设计的先验过程复杂且不能完全准确地表达深度学习隐空间中的特征关系，因此不足以引导细节特征的复原.

鉴于上述问题，本文提出了一种融合多尺度特征生成隐编码库先验的人脸去模糊方法. 首先，针对图像去模糊的细节缺失问题，通过编码器提取模糊图像的多分辨率特征，这些特征包含了不同尺度下的重要图像细节线索以及模糊图像的空间结构. 针对不同尺度的特征层，编码器使用残差嵌套模块(residual-in-residual dense block, RRDB)[16]生成隐编码张量，用于指导隐编码空间的逻辑层次结构和解码器中多尺度去模糊图像的生成. 也即，编码器特征在输入到隐编码库中协助生成图像不同细节的隐编码的同时，还作为解码器的输入产生多分辨率的解码特征. 此外，针对运动模糊图像先验缺失的问题，本文采用预训练的 StyleGAN[17]作为人脸先验的隐编码库. 该隐编码库在大型人脸数据集上训练，编码了不同尺度的诸多人脸细节(例如眉毛、发丝等). 为进一步提高该预训练编码库对运动模糊的适应能力，本文将编码库在合成的运动模糊图像数据集上进行了微调；并改进了 StyleGAN 的图像生成方式，使其更专注于图像的纹理和细节特征生成.

本文的主要贡献为：(1) 优化了隐编码在图像去模糊过程中表达能力不足的问题，提出了多尺度特征图融合的方法，加强了模糊图像的细节特征表达能力，也提高了生成图像特征与原图的关联性. (2) 提出了一种结构简单的网络模型，借助在大型人脸数据集上预训练的隐编码库来提取丰富的图像先验纹理知识，以辅助图像去模糊. (3) 验证了隐编码结构本身良好的解耦性设计，并通过实验可视化了图像层次表达特征的过程.

1 相关工作

由于图像盲去模糊方法无法获取原始清晰图像以及退化图像的退化模式，因此这类方法所处理的是典型的 ill-posed 问题. 早期传统的盲去模糊方法对图像退化过程的建模较为简单，即假定图像的模糊是均匀线性、相机曝光轨迹可预测或是模糊局部线性等，便于建模和优化求解，在早期硬件算力有限的情况下其具有更高的可用性. 然而处理真实情况下拍摄的具有运动前景的图像时，图像模糊的程度是不均匀的，曝光时运动轨迹也是不可预测的. 所以传统方式在处理真实情况中非均匀、非线性、轨迹不可预测的模糊图像时效果普遍不好.

随着近年来训练数据量的大幅增长、硬件算力的不断提升，以及深度神经网络技术的发展，基于深度学习的图像处理技术得以发展. 图像去模糊研究也随之向基于深度学习的图像去模糊方法演进.

1.1 基于深度学习的盲去模糊方法

这一类方法往往采用端到端的卷积神经网络训练模式，深度卷积神经网络可通过海量的参数对退化图像特征进行编码，再通过解码器对提取到的特征反卷积得到复原图像. 相比传统方法，基于深度学习的盲去模糊方法不需要提前假设先验模型，本质上是利用卷积神经网络学习退化图像与清晰图像之间的映射关系. 基于崇元等[1]提出的预测模糊图像中模糊核的方法，Levina 等[2]利用贝叶斯最小均方差采样算法，通过反卷积消除图像

模糊. Nah 等[18]首次将多尺度机制引入端到端卷积神经网络进行动态场景下的盲图像去模糊. 尽管端到端的训练模式较为简洁，但是由于其训练过程中卷积神经网络需要将大量模糊核共同作用于训练图像，从而极大地增加计算开销和训练时间，存在训练时间较长，容易过拟合等问题. 此外，由于神经网络往往假设模糊核满足高斯分布，对高斯核的计算误差会影响图像生成质量，导致复原结果出现振铃现象.

1.2　基于先验的深度盲去模糊方法

基于端到端训练模式的方法存在模型参数多，不易训练，容易产生伪影的弊端，引起这些问题的一个重要原因在于此类网络模型未考虑模糊图像产生退化的因素. 为此，有学者尝试将已训练好的模型或者生成模型作为先验项引入深度盲去模糊方法中. 例如, Li 等[19]通过一个已训练好的二分类器作为鉴别输入图像模糊与否的先验模型；而 Tao 等[20]聚焦于减少参数数量和简化网络结构，通过在不同尺度、不同层级上共享权重，以减少参数数量和网络复杂度. 随着生成器模型的广泛应用，图像去模糊方法可以使用生成器模型获得一定程度的先验知识，从而获取模糊图像多尺度的、从粗糙到精细的特征细节，如 DeblurGAN[21]，SRN-DeblurNet[22]，PSS-NSC[23]和 PSP[24]等.

1.3　基于深度学习的人脸去模糊

相比于自然图像去模糊，人脸去模糊更加专注于人脸特征的重建和恢复，所需要关注的特征纹理细节数量级大大下降，但更专注于对人脸风格的特征提取与细节重建. 在解码器部分可以直接使用解码器重建，也可以使用隐编码库提供的先验知识针对人脸特征进行重建. 例如，Pan 等[14]使用预训练的神经网络模型获取人脸的边缘特征，并将特征作为惩罚项用于辅助人脸去模糊模型的训练. 然而人脸图像去模糊有其内在的独特性，需要对人脸图像独特风格的细节进行复原，上述方法仅靠隐编码库约束并不足以实现对人脸精细特征的复原.

目前仍少有工作探究基于人脸图像的隐编码库和多尺度先验的去模糊工作，也缺乏多尺度层级特征下的由粗糙到精细上的结构性解耦. 因此，在人脸图像去模糊任务中，如何构建人脸去模糊图像的隐编码库，多尺度的特征细节对模糊图像的重建有什么样的作用，依旧是一个有趣而又值得挑战的课题. 本文针对以上考虑展开研究.

2　基于生成对抗网络的人脸去模糊方法

当前，基于生成对抗网络(generative adversarial networks, GAN)的方法在图像盲去模糊方法中能够取得较好的效果[22,24-25]，这些方法使用生成器和判别器模型为去模糊提供先验知识. 但是，这些方法没有探究不同尺度特征对去模糊过程的影响程度，而在人脸去模糊任务中，人脸细节的复原尤为重要.

StyleGAN[17]是近年来提出的一种基于风格的 GAN. StyleGAN 的创新之处在于提出了一种基于风格的图像生成架构，与传统生成器[13,16,26]相比，该生成器可以控制图像不同层次、不同尺度的特征；其主要结构为映射网络和合成网络. 其中映射网络用于根据特征尺度生成和层次化解耦相关的特征信息，这些特征作为合成网络不同层的特征输入. 本文使用基于 StyleGAN 的网络结构来提取多尺度的先验知识.

本文的主要创新在于使用现有的编码器-解码器架构来对退化的人脸图像构建隐编码库，基于 GAN 的隐编码库能够解耦原图中不同尺度的特征纹理信息，进而将解耦后的多尺度先验知识应用在图像去模糊过程中，以获得从结构到各级纹理更加合理、清晰的去模糊图像. 本文选取 StyleGAN 结构来生成人脸图像隐编码库，其优势在于生成器结构中每个尺度/分辨率上都有独立的特征提取和应用模块，在保证了多尺度特征提取和应用独立性的同时，也更好地解耦了不同尺度下的特征信息. 传统基于 GAN 提取先验知识进行图像重建的方法都将隐编码库作为生成器的唯一输入，在复原的图像上缺少细节信息. 针对此问题，本文创新地在编码器结构中通过迭代下采样来进行多尺度的隐编码库的生成和多分辨率特征图的提取. 此外，为提高在生成细节上的表达能力，本文增加了多尺度特征图作为隐编码库 StyleGAN 在生成器部分的输入，以协助隐编码库进行图像重建.

2.1　网络结构介绍

本文的网络结构如图 1 所示，其输入的模糊图像及输出的去模糊图像大小均为 $1\,024\times1\,024$ 像素. 其中，$E_i(i=0,1,2,\cdots,7)$ 和 $S_i(i=0,1,2,\cdots,7)$ 分别表示编码器块和隐编码库，$D_i(i=0,1,2,3)$ 表示解码器块. 本文中编码器读取特征信息隐编码，所以暂不关注映射网络的实现细节. 但是与映射网络保持一致，本文在使用 StyleGAN 的过程中，保证每个

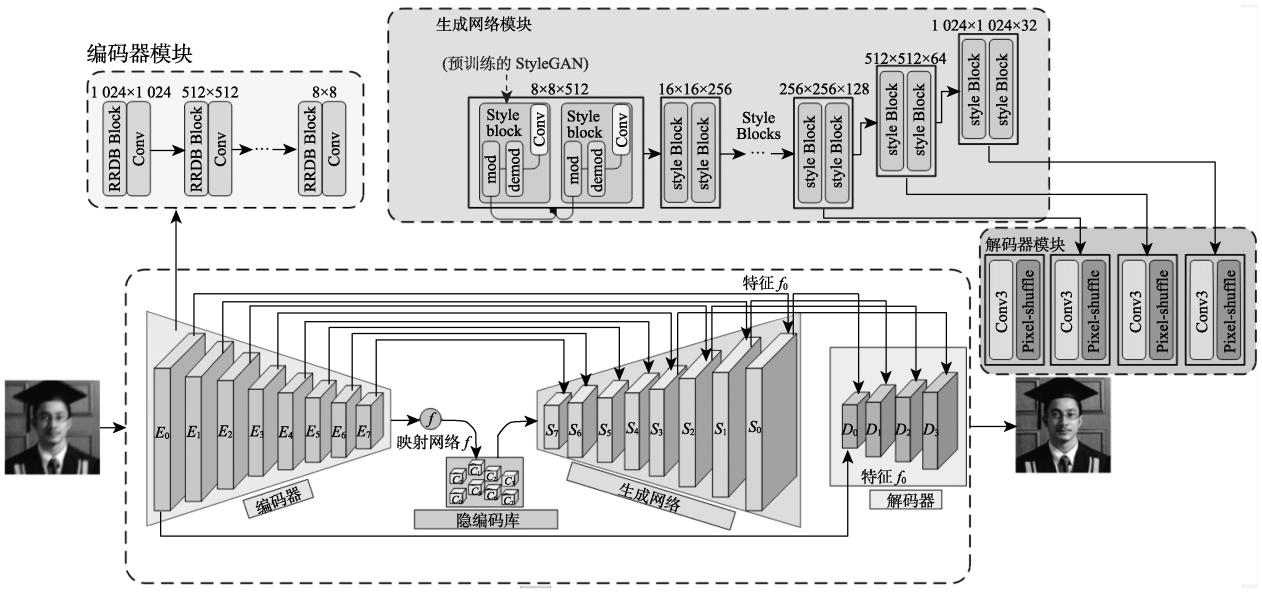


图 1 网络结构图

层级输入了对应尺度的特征；图像分辨率从小到大，图像特征的应用尺度也从大到小，以适应面部姿势、脸部轮廓到随机性高的皮肤纹理和毛发姿态等.

在生成网络中，每个层级输入保持了其隐编码特征的尺度；在应用过程中，随着分辨率增大，生成器通过加入随机噪声来保持小尺度特征的随机性和拟真性，减少了人工生成图像痕迹. 噪声的加入在人脸姿势等大尺度的复原上并不明显，在一些细小特征如毛发姿态、面部纹理等会受随机噪声的加入的较大影响，这样保证了图像生成模式具备随机性，减少失真感. 在合成网络中，每个模块都用了 2 个自适应实例归一化(adaptive instance normalization, AdaIN)[27]模块. AdaIN 本质上是一种图像风格归一化方法，用于将一幅图像的风格、纹理迁移到另一幅图像中，同时保留图像的主体结构特征. AdaIN 基于前向神经网络，具有很快的生成速度. 合成网络主要将风格应用在上采样后的生成图像，同时加入随机噪声保证图像的拟真性.

2.2 编码器

编码器采用 RRDB 网络来提取模糊图像的多尺度特征. 将 RRDB 网络和提取到的特征约定为 E 和 f，多尺度特征提取过程表达为

$$f_i = E_i\left(f_{i-1}\right),\ i \in \{1,2,\cdots,N\}.$$

其中，$E_i(i=0,1,2,\cdots,N)$ 表示采用不同步长卷积的编码操作. 随后，通过不同步长的卷积神经网络连续提取不同尺度下的图像特征，再使用卷积神经网络和全连接层来生成隐编码矩阵 $\boldsymbol{C}=E_{N+1}(f_N)$.

矩阵 $\boldsymbol{C}$ 的列表示 StyleGAN 需要的隐编码矩阵，它包含了图像的隐表示，为隐编码库(StyleGAN)提供高维度信息，以协助人脸去模糊中的纹理等细节信息.

2.3 隐编码库

本文将隐编码库加入到编码器-解码器架构中，使得模型可以获得更多的图像先验知识，从而使编码器可以更加专注于多尺度特征的提取. 在训练隐编码库的过程中，将人脸图像的训练数据集转化为丰富的人脸先验知识，即从训练集得到的分好类别和尺度的人脸特征纹理等细节信息. 尽管这些信息在图像中的存在形式是离散的，但是由于大数据集的训练加持，这些信息转化为隐空间中带有坐标信息的特征矩阵. 在隐编码库的生成器网络结构内，通过 AdaIN 对这些特征进行了特征归一化，即将编码器输入的特征进行多尺度的特征解耦，以实现对人脸的不同局部进行特征编码.

编码器的每一层都会生成成对的特征图和隐编码库，即每层一个隐编码矩阵 $\boldsymbol{C}$，其每一列 $\boldsymbol{C}_i$ 代表了一个尺度的特征隐编码库，其目的在于使特征融合的过程中某一尺度的特征不会干扰其他特征的生成. 例如，计算混合眼球样式特征(颜色、瞳孔大小等)时，不会影响眼角姿态等. 编码器同时将多尺度的特征输入隐编码库中，每一个编码器模块在生成网络中额外使用了一个附加卷积来进行特征融合，第 i 个附加卷积模块的输出特征

$$\boldsymbol{g}_i = \begin{cases} S_0(\boldsymbol{C}_0, \boldsymbol{f}_N), & i=0 \\ S_i(\boldsymbol{C}_i, \boldsymbol{f}_{N-i}), & i=1,2,3,\cdots \end{cases}.$$

其中，S_i 表示生成网络的第 i 个附加卷积模块；$\boldsymbol{C}$

和 $\boldsymbol{f}$ 分别对应该模块的隐编码矩阵和来自编码器的输入特征.

2.4　解码器

解码器采用了一个简单的卷积神经网络，采用 conv3 +pixel-shuffle 结构，由粗到细聚合起不同尺度的特征，并且每层使用了不同隐编码进行特征融合. 该做法不同于传统的基于隐编码进行图像复原的方法[21-24]，后者往往会忽略不同尺度的隐编码，只将隐编码作用于单一尺度的特征，随后进行统一的卷积上采样操作. 本文的解码器实现了不同层级的隐编码指导人脸去模糊的操作.

3　实验与分析

本节通过实验对所提出的网络框架结构进行定性及定量分析. 首先，介绍实验规模、数据集、配置和评估指标，并与已有的方法进行分析比较；最后，对隐编码的耦合性进行对比实验，以分析隐编码的风格解耦性和对人脸去模糊效果的特征影响程度.

3.1　实验数据准备

3.1.1　数据集

人脸网络相册数据集(Flickr-Faces-HQ, FFHQ)[17]，最初是以 GAN 为基准而创建的高清人脸数据集，它提供了分辨率为 1 024×1 024 像素的不同年龄段、不同种族和风格各异的 70 000 幅高清人脸 PNG 格式图像. FFHQ 数据集在人脸属性上拥有非常多的变化，可用于实验中增强对不同人脸风格的学习. 此外，该数据集主要用于人脸图像隐编码库的学习，以预先学习人脸先验知识，进而指导后续人脸去模糊的训练.

GoPro 数据集是近期基于深度学习的图像去模糊方法中使用最为广泛的数据集，其包括通过高帧率摄像设备捕获近似清晰的视频，并截取视频中某些帧作为清晰图像，再将其与相邻的清晰帧求平均值后得到合成的运动模糊图像. 原理上，是用最接近真实拍照动作的曝光延长来产生运动模糊图像.

相较于 GoPro 数据集，FFHQ 数据集提供的是清晰人脸图像. 为产生人脸运动模糊图像，本文使用模拟复杂运动的模糊核方法[28]来产生接近真实运动模糊的人脸模糊图像. 该方法遵循随机轨迹生成的概念，将亚像素插值法应用到轨迹矢量中. 每个轨迹矢量都用来模拟在二维平面内物体随机运动的离散位置. 该轨迹通过马尔可夫过程产生，即通过迭代地计算前一个点的速度、位置及其高斯摄动和脉冲摄动来随机生成下一个点的位置. 针对 FFHQ 数据集的 70 000 幅图像，本文使用模糊核方法进行了人脸运动模糊图像生成，将其制作成与 GoPro 数据集格式统一的 blurr-sharp 数据集对，作为本文的训练输入数据.

3.1.2　超参数配置

训练过程使用 4 个 GeForce RTX 2080 Ti GPU，设定 batch size 为 8，进行 50 000 次迭代. FFHQ 数据集训练过程中数据会被缩放到 32×32 作为 StyleGAN 的输入矩阵. 本文方法实现过程中使用了 Open-MMlab-mmedit 框架作为统一的网络验证框架.

3.2　评价指标

本文主要采用的评价标准是 Fréchet 距离(Fréchet inception distance, FID)[13]、结构相似性(structural similarity index measure, SSIM)和峰值信噪比(peak signal-to-noise ratio, PSNR).

FID 用于计算真实样本和生成样本在特征空间之间的距离；其聚焦于特征提取，而不关心特征的距离关系. 在图像去模糊实验中，用 FID 对去模糊后图像特征保留的完整性进行校验，其计算公式为

$$\mathrm{FID}(\boldsymbol{F}_x,\boldsymbol{F}_y)=\left\|\boldsymbol{\mu}_x-\boldsymbol{\mu}_y\right\|_2^2+\mathrm{tr}\left(\boldsymbol{\sigma}_x+\boldsymbol{\sigma}_y-2(\boldsymbol{\sigma}_x\boldsymbol{\sigma}_y)\right)^{1/2}.$$

其中，$\boldsymbol{F}_x$ 和 $\boldsymbol{F}_y$ 矩阵是由真实高清图像和由本文方法得到的去模糊图像经 Inception Net-V3 模型处理后得到的输出结果；$\boldsymbol{\mu}_x$ 和 $\boldsymbol{\mu}_y$，$\boldsymbol{\sigma}_x$ 和 $\boldsymbol{\sigma}_y$，分别对应 $\boldsymbol{F}_x$ 和 $\boldsymbol{F}_y$ 的均值和方差.

SSIM 用于提取人眼所看到的图像中的结构化关系，其将图像失真建模为亮度、对比度和结构 3 个不同因素的组合. 用均值作为亮度的估计，标准差作为对比度的估计，协方差作为结构相似程度的度量. 使用 SSIM 在宏观角度对图像去模糊实验质量进行评估，其计算公式为

$$\mathrm{SSIM}(\boldsymbol{I}_x,\boldsymbol{I}_y)=\frac{(2\boldsymbol{\mu}_x\boldsymbol{\mu}_y+\boldsymbol{c}_1)(2\boldsymbol{\sigma}_{xy}+\boldsymbol{c}_2)}{(\boldsymbol{\mu}_x^2+\boldsymbol{\mu}_y^2+\boldsymbol{c}_1)(\boldsymbol{\sigma}_x^2+\boldsymbol{\sigma}_y^2+\boldsymbol{c}_2)}.$$

其中，$\boldsymbol{I}_x$ 和 $\boldsymbol{I}_y$ 分别是真实高清图像和由本文方法得到的去模糊图像；$\boldsymbol{\mu}_x$ 和 $\boldsymbol{\mu}_y$，$\boldsymbol{\sigma}_x$ 和 $\boldsymbol{\sigma}_y$，分别对应 $\boldsymbol{I}_x$ 和 $\boldsymbol{I}_y$ 的均值和方差；$\boldsymbol{c}_1$ 和 $\boldsymbol{c}_2$ 分别是 2 个常数. 当 SSIM 值越接近 1 时，表示图像还原度越高.

PSNR 一般用于评价图像还原后的视觉品质，单位为 dB；其可对图像重建后质量进行量化评估，PSNR 计算公式为

$$\mathrm{PSNR}=10\times\lg\left(\frac{\mathrm{MAX}_I^2}{\mathrm{MSE}}\right).$$

其中，MSE 是真实高清图像 I 和去模糊后的图像 K 之间的均方误差；MAX_I 为图像 I 的最大像素值. MSE 的计算公式为

$$\mathrm{MSE}=\frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}\left(\boldsymbol{I}(i,j)-\boldsymbol{K}(i,j)\right)^2 .$$

其中，m 和 n 是图像 I 和图像 K 的像素总数.

3.3 定性与定量分析

3.3.1 与主流方法的量化比较

将本文方法与主流的盲去模糊方法[12,19,24,29-33]进行对比，并分析其优缺点. 从表 1 可以看到，相比于其他方法，本文方法在 PSNR, SSIM 和 FID 这 3 个评价指标上均具有优势. 同时，从定性分析的角度对比不同方法，对比结果如图 2 所示，可以观察到，本文方法去模糊后的图像细节最好，同时人工痕迹如振铃现象也最少.

表 1 本文方法与主流方法对比图

去模糊方法	PSNR/dB	SSIM	FID
L_0 SparsityMotionDeblurring[29]	15.81	0.278	426.21
motion-psf-estimation[30]	13.37	0.292	337.91
sharp-by-non-blind-deconv[12]	11.94	0.200	337.91
two-phase-kernel[19]	12.97	0.218	337.91
Pulse[31]	19.53	0.552	228.75
ALAE[32]-celeba 训练	14.91	0.421	205.91
ALAE[32]-FFHQ 训练	13.28	0.510	166.70
Pixel2Style2Pixel[24]	18.17	0.520	168.03
CodeFormer[33]	19.93	0.590	77.37
本文	**24.89**	**0.680**	**67.94**

注. 加粗表示最优值.



a. 原图 b. 模糊图像 c. 文献[12] d. 文献[19] e. 文献[24] f. 文献[29] g. 文献[30] h. 文献[31] i. 文献[32] j. 本文

图 2 人脸图像去模糊效果对比

3.3.2 隐编码库解耦实验分析

本节对所提出的隐编码库部分进行解耦分析. 首先分别对每一层编码进行单独定性评价，从表 2 可以发现，不同层的隐编码库带来的影响大小相近，说明每层编码库都具备了一定的细节提取能力. 其中，隐编码库仅应用在单层时 PSNR 值维持在 4 dB, SSIM 维持在 0.1 以内, FID 为 380+. 这也表明，单一层次的隐编码库应用并未对图像重建过程产生决定性影响.

表 2 单层隐编码库输入的评估结果

隐编码库应用层级	PSNR/dB	SSIM	FID
1	5.07	0.059	441.81
2	4.68	0.328	435.34
3	5.12	0.082	369.07
4	4.56	0.392	372.53
5	4.89	0.143	432.46
6	5.66	0.056	383.64
7	5.29	0.132	393.13
8	5.73	0.087	385.16

然后，本文分层提取隐编码库并将其应用于图像重建，以验证不同层次隐编码库对图像重建的影响. 作为与隐编码库指导重建方法的对比，本文将解码结构中的卷积、风格应用和噪声添加等操作替换为普通的多层卷积来进行上采样操作. 与 StyleGAN 中从低分辨率图像到高分辨率图像复原的图像重建过程类似，本文将网络设计为从 32×32 大小到 1024×1024 大小一共 8 层的上采样卷积，而对照实验组用上采样操作代替单层的图像复原步骤. 通过图 3 的可视化比较可以看出，随着高层特征被逐层提取，颜色信息首先开始失真，而直至剩余最后 2 层仍可以保留轮廓信息. 这说明隐编码库能够对图像的不同结构进行分层编码，并且上层隐编码库更关注人脸的形状等结构信息，下层隐编码库更关注图像的色彩和纹理等细节.

最后，本文定量地分析了不同分层的隐编码库组合对图像重建的影响，以验证所提分层模型的解耦性能，表 3 展示了不同分层的隐编码库组合对图像重建的影响. 可以看出，随着多层隐编码库的不断组合，PSNR 和 SSIM 的评估值逐渐增加，说明不同尺度的特征和隐编码库包含了不同尺度的解耦信息.

结合表 2 中单层隐编码库的性能影响，这 2 个实验验证了隐编码库在风格应用层面的解耦完备性.



图 3　将隐编码库逐渐应用到解码器不同层中后图像去模糊的效果展示

表 3　不同分层隐编码库组合的评估结果

隐编码库应用层级	PSNR/dB	SSIM	FID
1	5.07	0.059	441.81
1, 2	5.30	0.098	483.62
1, 2, 3	5.77	0.089	488.10
1, 2, 3, 4	7.37	0.050	512.97
1, 2, 3, 4, 5	8.56	0.255	352.64
1, 2, 3, 4, 5, 6	15.21	0.474	229.37
1, 2, 3, 4, 5, 6, 7	16.56	0.533	135.13
1, 2, 3, 4, 5, 6, 7, 8	24.89	0.681	67.94

3.3.3　消融实验

为了验证本文模型各部分的有效性，本节选用 GoPro 数据集中相机拍摄的真实运动模糊图像，进行一系列消融实验，分别验证编码器的有效性、验证隐编码库的有效性和验证解码器的有效性. 表 4 中，消融实验 1~3 分别去除了相应模块后再进行相应训练和测试，以验证该模块带来的影响. 可以看出，3 个模块对最终的去模糊效果均有一定影响，尤其是隐编码库和编码器对 FID 和 PSNR 产生了显著的影响. 下面结合图 4 中的对比进行逐项分析.

表 4　隐编码库从高层减少输入的评估结果

消融实验	去除模块	PSNR/dB	SSIM	FID
实验 1	编码器	17.67	0.410	170.81
实验 2	隐编码库	17.35	0.514	203.01
实验 3	解码器	19.10	0.468	108.55
本文		24.89	0.681	67.94



a. 运动模糊图像　b. 消融实验 1　c. 消融实验 2　d. 消融实验 3　e. 本文

图 4　消融实验对比图

实验 1 中移除编码器的特征输入，改为 StyleGAN 的生成器直接产生输出图像. 当没有编码器特征作为输入，直接复原图像的方法类似于典型的 GAN 反转，即只学习隐编码库. 这种做法能够在给定隐编码库的情况下合成较为真实的图像，然而其只由低维向量引导，缺乏空间信息，因而只恢复了全局属性，仍然缺乏内部细节. 如图 4b 所示，实验结果能够复原头发颜色和姿势等特征，但未能复原更精细的细节.

实验 2 是将隐编码库替换为常用的批归一化

模块(batch normalization, BN). BN 可以对编码器特征进行归一化, 使特征符合正态分布, 但缺乏对不同细节的特殊处理. 实验 2 的结果表明, 去掉隐编码库对生成图像的质量有着较大影响, 尤其对 FID 和 PSNR 这 2 项指标. 如图 4c 所示, 去掉隐编码库后生成图像的整体色彩效果有较大损坏, 人脸细节也有较大缺失.

实验 3 中移除解码器, 将隐编码库的输出作为最终输出, 如图 4d 所示, 在没有解码器的情况下, 尽管从整体上看图像质量尚可, 但在图像放大后能够观察到部分伪影. 本文提出的解码器网络以从粗到细的方式融合多尺度特征信息, 有助于生成更自然的特征细节. 此外, 编码器和解码器之间的多尺度连接, 加强了解码器捕获特征的空间信息的能力, 进而弥补了隐编码库蕴含丰富语义但缺乏空间位置信息的不足.

如图 4e 所示, 本文方法同时采用隐编码库和多尺度融合, 无论在视觉效果和局部细节上都呈现了更好的生成效果.

3.3.4 针对散焦模糊图像的实验

为进一步验证本文方法在人脸去模糊上的广泛适应性, 将本文训练好的模型在人脸散焦图上进行了测试, 如图 5 所示. 可以看到, 本文方法能够有效地去除人脸散焦模糊, 同时较好地保持了原图背景. 然而本文方法也存在一些不足之处, 例如, 在有前景遮挡物(第 1 行)和亮度较低(第 3 行)的场景下, 人脸散焦模糊的去除效果相对较差.



a. 散焦模糊图 b. 本文方法去模糊结果

图 5 本文方法在散焦人脸图像上的效果

4 结 语

本文提出了使用编码器-隐编码库-解码器的框架, 用于解决图像去模糊问题. 在生成器部分, 有针对性地通过下采样分层提取特征并以隐编码库指导的方式来保证图像先验知识应用到去模糊过程中; 这也是本文与传统编码器-解码器框架去模糊方法的主要区别, 该改进主要提高了图像去模糊任务中基础特征的稳定性和图像生成过程中不同局部特征的解耦性, 同时减轻了 StyleGAN 在图像生成过程中因随机噪声而带来的负面影响, 从整体到细节上提高了图像输出的稳定性. 本文方法采用了多层次特征融合的策略, 因此存在模型参数量较大、图像处理速度较慢的局限性. 未来将在进一步提升图像去模糊能力的基础上, 采用更加高效的网络结构, 以提高图像去模糊的处理效率和广泛适用性.

参考文献(References):

[1] Chong Yuan, Xu Xiaogang. New method of motion deblurring from a single imag[J]. Computer Engineering and Design, 2012, 33(11): 4305-4308(in Chinese)
(崇元, 徐晓刚. 单幅图象去运动模糊的新方法[J]. 计算机工程与设计, 2012, 33(11): 4305-4308)

[2] Levin A, Weiss Y, Durand F, *et al*. Efficient marginal likelihood optimization in blind deconvolution[C] //Proceedings of the CVPR 2011. Los Alamitos: IEEE Computer Society Press, 2011: 2657-2664

[3] Hu Z, Huang J B, Yang M H. Single image deblurring with adaptive dictionary learning[C] //Proceedings of the IEEE International Conference on Image Processing. Los Alamitos: IEEE Computer Society Press, 2010: 1169-1172

[4] Pan J S, Sun D Q, Pfister H, *et al*. Deblurring images via dark channel prior[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(10): 2315-2328

[5] Shan Q, Jia J Y, Agarwala A. High-quality motion deblurring from a single image[J]. ACM Transactions on Graphics (TOG), 2008, 27(3): 1-10

[6] Chen L, Fang F M, Wang T T, *et al*. Blind image deblurring with local maximum gradient prior[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2019: 1742-1750

[7] Brock A, Donahue J, Simonyan K. Large scale GAN training for high fidelity natural image synthesis[OL]. [2023-02-10]. https://arxiv.org/abs/1809.11096

[8] Schmidhuber J. Learning factorial codes by predictability minimization[J]. Neural Computation, 1992, 4(6): 863-879

[9] Ridgeway K. A survey of inductive biases for factorial representation-learning[OL]. [2023-02-10]. https://arxiv.org/abs/1612.

05299

[10] Achille A, Soatto S. Emergence of invariance and disentanglement in deep representations[J]. The Journal of Machine Learning Research, 2018, 19(1): 1947-1980

[11] Gao H Y, Tao X, Shen X Y, *et al*. Dynamic scene deblurring with parameter selective sharing and nested skip connections[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2019: 3843-3851

[12] Dumoulin V, Perez E, Schucher N, *et al*. Feature-wise transformations[EB/OL]. [2023-02-10]. https://distill.pub/2018/feature-wise-transformations/

[13] Karras T, Aila T, Laine S, *et al*. Progressive growing of GANs for improved quality, stability, and variation[OL]. [2023-02-10]. https://arxiv.org/abs/1710.10196

[14] Pan J S, Ren W Q, Hu Z, *et al*. Learning to deblur images with exemplars[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 41(6): 1412-1425

[15] Li L R H, Pan J S, Lai W S, *et al*. Blind image deblurring via deep discriminative priors[J]. International Journal of Computer Vision, 2019, 127(8): 1025-1043

[16] Wang X T, Yu K, Wu S X, *et al*. ESRGAN: enhanced super-resolution generative adversarial networks[C] //Proceedings of the European Conference on Computer Vision. Heidelberg: Springer, 2018: 63-79

[17] Karras T, Laine S, Aila T. A style-based generator architecture for generative adversarial networks[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2019: 4396-4405

[18] Nah S J, 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

[19] Li X, Jia J Y. Two-phase kernel estimation for robust motion deblurring[C] //Proceedings of the 11th European Conference on Computer Vision. Heidelberg: Springer, 2010: 157-170

[20] 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

[21] Krishnan D, Fergus R. Fast image deconvolution using hyper-Laplacian priors[C] //Proceedings of the 22nd International Conference on Neural Information Processing Systems. Red Hook: Curran Associates Inc., 2009, 22: 1033-1041

[22] Kupyn O, Budzan V, Mykhailych M, *et al*. DeblurGAN: blind motion deblurring using conditional adversarial networks[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2018: 8183-8192

[23] Sun J, Cao W F, Xu Z, *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 Alamito: IEEE Computer Society Press, 2015: 769-777

[24] Richardson E, Alaluf Y, Patashnik O, *et al*. Encoding in style: a StyleGAN encoder for image-to-image translation[C]. //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2021: 2287-2296

[25] Laine S. Feature-based metrics for exploring the latent space of generative models[EB/OL]. [2023-02-10]. https://openreview.net/pdf?id=BJslDBkwG

[26] Heusel M, Ramsauer H, Unterthiner T, *et al*. GANs trained by a two time-scale update rule converge to a local Nash equilibrium[C] //Proceedings of the 31st International Conference on Neural Information Processing Systems. Red Hook: Curran Associates Inc., 2017: 6629-6640

[27] Huang X, Belongie S. Arbitrary style transfer in real-time with adaptive instance normalization[C] //Proceedings of the IEEE International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2017: 1510-1519

[28] Boracchi G, Foi A. Modeling the performance of image restoration from motion blur[J]. IEEE Transactions on Image Processing, 2012, 21(8): 3502-3517

[29] Xu L, Zheng S C, Jia J Y. Unnatural L_0 sparse representation for natural image deblurring[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2013: 1107-1114

[30] Deshpande A M, Patnaik S. Single image motion deblurring: an accurate PSF estimation and ringing reduction[J]. Optik, 2014, 125(14): 3612-3618

[31] Menon S, Damian A, Hu S J, *et al*. PULSE: self-supervised photo upsampling via latent space exploration of generative models[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2020: 2434-2442

[32] Pidhorskyi S, Adjeroh D A, Doretto G. Adversarial latent autoencoders[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2020: 14092-14101

[33] Zhou S C, Chan K C K, Li C Y, *et al*. Towards robust blind face restoration with codebook lookup transformer[C] //Proceedings of the 36th International Conference on Neural Information Processing Systems. Red Hook: Curran Associates Inc., 2022: 30599-30611