DOI:10.16136/j.joel.2023.10.0441

基于先验信息与密集连接网络的人脸超分辨率重建方法

崔立尉*,高宏伟

(内蒙古农业大学 计算机技术与信息管理系,内蒙古 包头 014109)

摘要:图像超分辨率在医疗和安防等领域应用广泛,本文针对传统超分辨率重建(super-resolution reconstruction,SR)方法无法重建出边缘特征图像的不足,提出了一种先验信息与密集连接网络模型的重建方案,利用考虑输入统计信息的残差特征的不同组合,引入了多注意力模块,通过与主干网络结构协作,在不增加额外模块的情况下提高了网络性能。新提出的模型与现有复杂结构的技术(state-of-the-art,SOTA)模型相比,具有更好的性能。为了避免输入的身份特征会急剧漂移的问题,提出了一种基于先验信息引入注意力机制网络模块来分辨真实低分辨率(low resolution,LR)对应物的模型,这种模型在捕获运动噪声等方面具有优势。经实验验证得出,本文方法相比其他主流方法,在评价指标和主观可视化分析方面更具优势。

关键词:超分辨率重建(SR);密集连接网络;先验信息

中图分类号:TP277 **文献标识码**:A **文章编号**:1005-0086(2023)10-1097-08

Face super-resolution reconstruction based on prior information and dense connected network

CUI Liwei*, GAO Hongwei

(Department of Computer Technology and Information Management, Inner Mongolia Agricultural University, Baotou, Inner Mongolia 014109, China)

Abstract: Image super-resolution is widely used in medical and security fields. Aiming at the shortcomings that traditional super-resolution reconstruction (SR) methods cannot reconstruct edge feature images, this paper proposes a reconstruction scheme based on prior information and dense connected network model. By taking into account the different combinations of residual features of input statistical information, a multi attention module is introduced to improve the network performance without adding additional modules by cooperating with the backbone network structure. The proposed model has better performance than the existing state of the art (SOTA) model with complex structures. In order to avoid the sharp drift of the input identity features, a network module of attention mechanism based on prior information is proposed to estimate the real low resolution (LR) counterpart. This model has advantages in terms of capturing motion noise, etc. The experimental results show that this method has more advantages in evaluation indicators and subjective visual analysis than other mainstream methods.

Key words: super-resolution reconstruction (SR); dense connected network; prior information

0 引 言

随着信息科技的进步,人们的生活需求越来越高,公共场所人员密度较高的地点的安全问题备受重视。我们国家为此还专门提出了平安城市、天网工程等治安防控工程[1-3]。在诸多的个人信息中,人脸信息具有易获得、唯一性等优点而被广泛用于身份确认、目标追踪和视频侦缉等诸多

* E-mail:nmghw@163.com

收稿日期:2022-06-15 修订日期:2022-10-02

基金项目:国家自然科学基金(31560141)、内蒙古自治区高等学校科学研究项目(NJZY18066)、内蒙古自治区高等学校科学研究项目(NJZY20055)、内蒙古哲学社会科学规划项目(2020NDB009)和内蒙古农业大学职业技术学院科技创新团队智慧农牧业科技创新团队(TDE202308)资助项目

领域。但是，在很多实际的现场环境中，往往由于监控摄像头的安装角度、安装位置、自身的分辨率大小，以及现场环境中的天气变化、光照角度及强度大小等诸多不可测因素的影响，使得监控终端获得的人脸图像或视频资料存在残缺不全、影像模糊、图片噪声多过等问题[4]。

人脸超分辨率是计算机视觉领域的一个研究热点。现有的大多数超分辨率重建(super-resolution reconstruction，SR)方法都需要成对的高分辨率(high ressolution，HR)和低分辨率(low resolution，LR)图像作为训练样本来训练重建模型。然而，在实际场景中很难收集成对的 LR 和 HR 图像。模拟的低质量图像无法匹配复杂的退化，导致结果不令人满意。为了解决这一问题，基于先验信息和人脸关键点的人脸 SR 方法近年来受到越来越多的关注。DONG 等[5]率先将卷积神经网络引入图像超分辨率(super resolution convolutional neural networks，SRCNN)任务。CAO 等[6]利用深度强化学习，提出了一个注意感知人脸超分辨率框架。受深度卷积神经网络去噪器的启发，SAHARIA 等[7]提出了一种两步的面部增强方法。YU 等[8]开发了一种属性嵌入上采样网络，以减少人脸图像超分辨率中的模糊性。上述这些基于深度的方法没有事先考虑到高度结构化的面部，并且可能在带有噪声的面部超分辨率中失败。CAO 等[9]提出了一种级联双网络模型，其首先对人脸图像进行解析，再将其融合成一个密集对应域的形式，在人脸图像的不同尺度上对这个对应域进行训练，最终形成一幅重建的人脸图像。HOWARD 等[10]提出了一种基于解析图映射的多尺度网络，该网络由 Parsing Net 和 FishSRNet 两个子网络构成，前者从 LR 人脸图像中获得解析映射，然后将这种映射关系作为输入放到后一个子网络中，最终生成 SR 图像。KIM 等[11]在该方法的基础上，在设计解析图映射时将LR人脸图像与该映射相乘，产生一个掩码，并将其与 LR 人脸图像级联，将这个级联作为 FishSRNet 子网络的输入。从对比实验的结果来看，该方法的确能够解决上述问题，有一定的参考价值。而对于利用人脸关键点的算法而言，BULAT 等设计的 Super-FAN 网络是较为经典的方法之一[12]。主要用于将人脸特征中重建的人脸图像进行关键点检测，并将捕获到的人脸结构信息与真实的人脸图像进行比较，并将其作为一种损失约束[13]。注意机制可以更加关注关键特征，这有利于特征学习和模型训练。ZHANG 等[14]证明，通过考虑通道之间的相互依赖性并调整通道注意机制，可以重建高质量的图像。HE 等[15]提出了一种面部空间注意机制，它使用沙漏结构形成注意机制，因此卷积层可以自适应地提取与关键面部结构相关的局部特征。在提取图像的全局表示方面有很大的优势，但仅依靠图像级别的自我关注仍会导致局部细粒度细节的丢失。因此，如何有效地结合图像的全局信息和局部特征对高质量的图像进行重建至关重要，这也是本文工作的研究目标之一。

为了避免输入的身份特征急剧漂移，更好地重建出人脸边缘和纹理的细节特征，提出了一种基于先验信息引入注意力机制网络来估计真实的 LR 对应物，这种模型在捕获运动噪声等方面具有优势。

1　本文方法

1.1　网络模型架构

如图 1 所示，本文通过提出一种基于先验信息引入注意力机制网络来估计真实的 LR 对应物，并且融入了注意力机制，为了获得可靠的细节信息，我们设计 MSB 来混合 LRB 功能，$F_M^{\mathrm{LR}}F^{\mathrm{Col}}$ 横向标高 $0 \leq m \leq M$，如图 2 所示。MSB 包括旋转层、剩余残差块、上/下采样和混合块。每个混合块接受多级上/下采样的特征，使用卷积层处理特征，连接结果，并

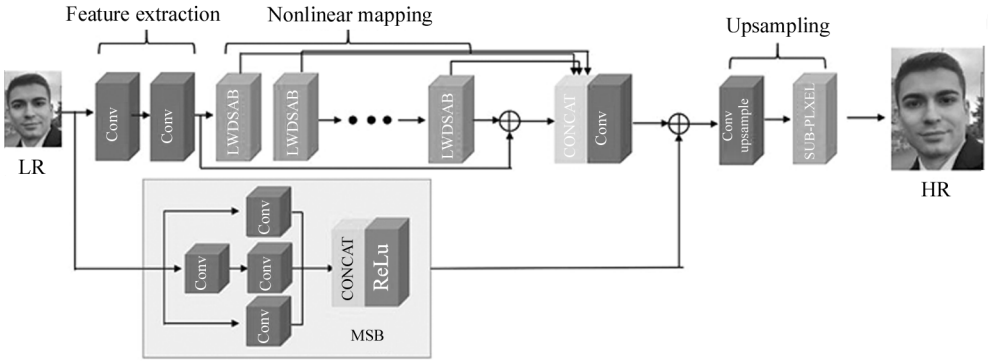

图 1　本文网络架构

Fig. 1　Network architecture in this paper

进行性能调整,以实现更强大的特征表示并提高重建性能。最后,从各个级别匹配特征的解决方案,连接并使用上采样层来创建图像 I^{SR}。

图 2 组件将成为集合特征金字塔和输入特征,从而提供更强大的特征表示。平均池特征 F^{avg} 和最大池功能 F^{max} 从模块输入中提取,用于共享网络生成。$\hat{F}^{avg}$ 和 $\hat{F}^{max}$ 分别地共享网络由两个完全连接的 FC 层组成:前一层由一个 Actorof 减少通道尺寸 1/16。然后,$\hat{F}^{avg}$ 和 $\hat{F}^{max}$ 由元素相加合并,通过 sigmoid 函数,生成通道注意图 M^{ch}。

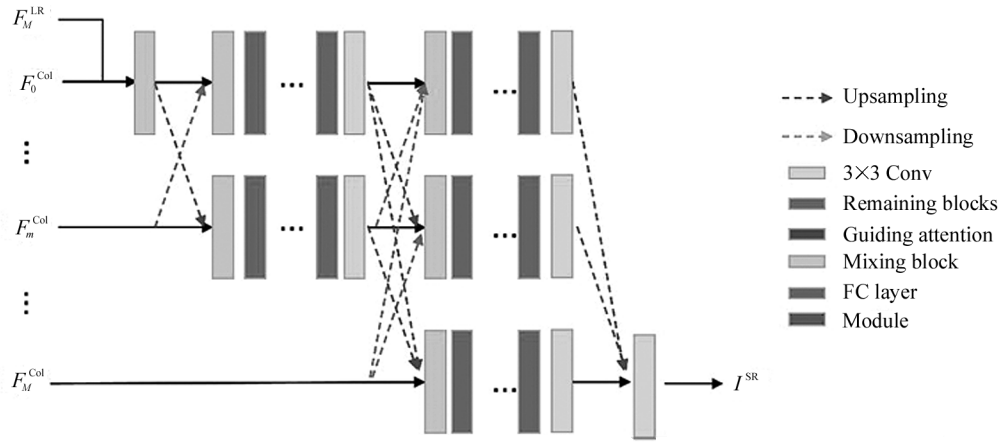

图 2　MSB 流程图

Fig. 2　MSB flowchart

1.2　先验信息人脸模块

通常情况下,图像超分辨率是以提高一般模糊图像的分辨率为主要目的,而人脸图像的超分辨率属于其中比较特殊的一种,这是由于人脸高度结构化,具有许多自身所独有的信息[16]。现将人脸图像所具有的先验信息总结如下。

1) 解析图先验信息。这种方法将人脸图像的各个部分分割出来,如图 3 所示。该方法为每个部分分配一个像素级(不同颜色)的标签,SR 时对不同的标签进行单独的训练,将训练的结果融合后就形成了最终的 HR 图像。

(a)　(b)

图 3　人脸图像(a) 原图及(b) 解析图

Fig. 3　(a) Original face image and (b) analysis diagram

2) 热图先验信息。同关键点先验信息的作用类似,该方法将人脸图像的关键部分以概率的方式表示出来,如图 4 所示,其也能够提供人脸图像的位置信息和面部轮廓信息,这些信息都可以为人脸图像的 SR 提供有效的先验信息。

(a)　(b)

图 4　人脸图像(a) 原图及(b) 热图

Fig. 4　(a) Original face image and (b) heat map

通过实验发现,将不同类型的先验信息加入到 SR 工作中可以减少重建时间、提高重建速度。而且,根据一些人脸图像及 SR 算法的特点,还可以从不同的阶段,例如将 LR 人脸图像作为输入的阶段,HR 的人脸图像重建作为完成阶段,以及这两个阶段中的一些中间阶段等等,来提取人脸图像的先验信息,也能不同程度地提高重建速度。

1.3　网络特征提取模块

本文提出的模型架构包括 3 个模块:特征提取模块、具有多个动态剩余保持组的非线性映射模块和重构模块。

1）特征提取模块：特征提取模块是具有 3 个卷积核的第一个卷积层。输入图像是第一个卷积层，可见下式：

$$x_0 = F_{\text{ext}}(I_{\text{LR}}), \tag{2}$$

式中，F_{ext} 表示特征提取操作，I_{LR} 表示 LR 输入图像。第一个卷积 Allayer 形成浅层特征提取，输出特征 x_0 进入非线性映射模块和构造模块。

2）非线性映射模块：非线性映射模块由多个面阵组成，将信息提取到构造边缘。下面的功能 x_k 是已通过处理的：

$$x_k = F^k_{\text{DRAG}}(x_{k-1}), k = 1,2,\cdots,K, \tag{3}$$

式中，F^k_{DRAG} 致力于三面体操作，表示 F^k_{DRAG} 中的拖动数。拟定阻力由多个剩余块(RB)组成，多个残差块、动态冗余块(DRM)的提案如图 5 所示。此外，图中还描述了雷达部分的信号流。图 5 中显示了将控制剩余连接抽出以自适应地组合特征，从而允许抽出以适当的重量组合剩余块的特征，以处理各种输入图像的属性。图中详细说明了 DRSA 开发注意机制：DRA(网格框的上部)和 RSA(下部)。

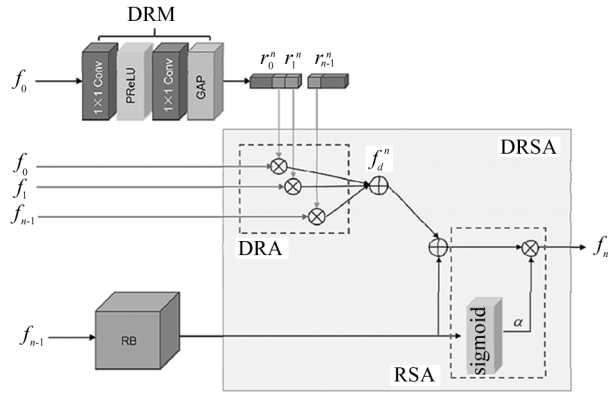

图 5　动态残余注意组分支提取的特征信息流程图

Fig. 5　Flowchart of feature information extraction from dynamic residual attention component branch

3）重构模块：提取深层特征 x_k，然后通过上采样网络进行放大：

$$x_{\text{up}} = G_{\text{up}}(F_{3\times3}(x_k) + x_0), \tag{4}$$

式中，G_{up} 表示上采样函数，它由一个卷积层和像素随机层组成。遵循全球剩余学习模式，浅层特征在混合操作之前被添加为特征。像素混洗层通过深度到空间操作转换输入功能的形状，放大后，网络按如下方式重建超分辨率图像：

$$I_{\text{SR}} = F_{\text{rec}}(x_{\text{up}}), \tag{5}$$

式中，I_{SR} 是重建的 SR 图像，F_{rec} 是核大小为 3 的最后一个卷积层的运算。动态残余注意组作为非线性映射模块的构建块，DRA 提取残余特征以重建精确的 HR 图像。图 5 显示了动态残余注意组分支提取的特征信息模块，有两个进化层，它控制着 DRA 和拖动中剩余块之间的剩余连接。使用 sigmoid 函数作为激活函数，利用其作为宽范围内的残余系数来实现特征多样性。

这表明了剩余故障的所有特征，这些特征来自于潜在连接的剩余故障块，从而允许剩余故障的牵引力输出。在计算了动态残余故障后，需要重新进行自我关注。剩余的自我注意系数 α 计算式为：

$$\alpha = \sigma(F^n_{\text{res}}(f_{n-1})), \tag{6}$$

式中，α 表示激活函数。通过采用 3D 注意力机制，它可以自适应地缩放特征通道和空间。使用与(Bulat 和 Tzimiropoulos)相同的网络架构来设计 f_{sr}。它有 3 个剩余单元，分别包含 12 个、3 个、3 个共卷块，前两个单元的末端有 2 个放大块。为了避免棋盘效应，使用了最近邻插值进行放大，而不是转置卷积。图 6 显示了该网络的示意图。网络的深化保证了高性能的高参数和计算资源。上述，这些分模型的计算成本具有一定程度的差异，这给实时场景的应用带来了困难。因此，各种轻量级网络也被提出用于实际应用。例如，深度递归神经网络(deeply-recursive convolutional network，DRCN)、DRRN (deep recursive residual network)和 DRFN(deep recurrent fusion network)利用轻量化结构来减少参数的数量。图 7 显示了注意力机制的内部构件，平均池特征 F^{avg}

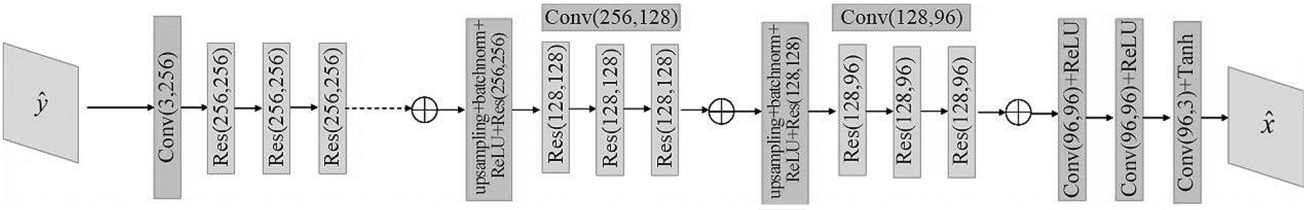

图 6　注意力机制缩放模块

Fig. 6　Attention mechanism scaling module

和最大池功能 $F^{\max}$ 从模块输入中提取，$\hat{F}^{\text{avg}}$ 和 $\hat{F}^{\max}$ 分别地共享网络由两个全连接的 FC 层组成:前一层由一个 Actorof 减少通道尺寸 1/16。然后，$\hat{F}^{\text{avg}}$ 和 $\hat{F}^{\max}$ 元素相加合并，通过 sigmoid 函数，生成通道注意图。

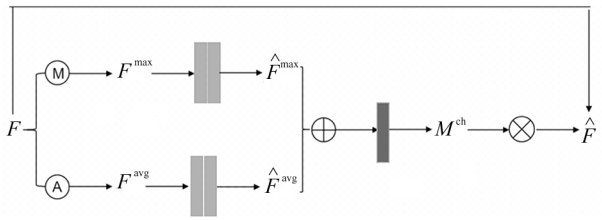

图 7　Attention 的模块流程图

Fig. 7　Flowchart of attention module

利用像素级 MSE 损耗($L_{\text{sr}}^{\text{MSE}}$)的组合来训练这个网络 sr 和对抗性损失($L_{\text{sr}}^{\text{Adv}}$)，定义如下：

$$L_{\text{sr}}^{\text{MSE}} = \| x - \hat{x} \|_2 , \tag{7}$$

$$L_{\text{sr}}^{\text{Adv}} = - E_{x \in X} D_s(f_{\text{sr}}(f_{\text{deg}}(x, z))) 。 \tag{8}$$

2　实　验

2.1　数据集

本实验训练数据集采用了 ILSVRC2015_VID 数据集[17]，该数据集包含 6 000 多个被逐帧标注的视频序列和约 140 万个人工标注数据，目标类别数为 40。此外，为进一步验证算法的有效性，在训练后选用了数据集 OTB50[18] 做进一步的测试，该数据集包含 60 个人工标注的视频帧。

2.2　实验分析

实验结果如图 8 所示。首先，如表 1 所示，两个基准都能够确定更好的模型。然而，这些基线并不能预测某些图像对人眼的影响要比它们的对应图像大得多。这一观察结果非常直观，因为两个基线模型都经过了训练，以估计单个图像的质量，虽然两个图像在一个人看来可能相当不错，但在配对比较任务中，一个图像明显更适合。相比，提出的 NeuralSBS 模型准确地预测了分布的形状，如图 8 所示。

NeuralSBS 可以作为一种研究工具，在通用基准上对 SR 模型进行充分的无参考评估。本实验组织如下。对于数据集中的每幅图像，从 9 种可用的 SR 方法中选择所有可能，并用 NeuralSBS 和每个基线对它们进行排序。获得的预测数据集中在 PARNAC 上，与其他数据相差较大。结果如表 2 和表 3 所示，NeuralSBS 模型在这项任务上显著优于其他方法。

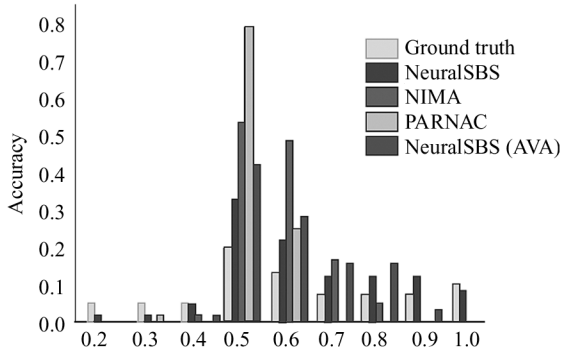

图 8　ILSVRC2015_VID 测试集的 Ground truth Score 和各种算法生成的分数分布

Fig. 8　Distribution of Ground truth Score and scores generated by various algorithms in ILSVRC2015_VID test set

表 1　ILSVRC2015_VID 测试集评估准确性对比

Tab. 1　Comparison of ILSVRC2015_VID test set evaluation accuracy

	NeuralSBS	MOS	NIMA	PARNAC	NIQE
Accuracy%	80.6	73.1	79.4	76.7	41.6

表 2　1ontheMa 测试集评估准确性对比

Tab. 2　Comparison of the evaluation accuracy of the 1ontheMa test set

	NeuralSBS	MOS	NIMA	PARNAC	NIQE
Accuracy%	65.3	45.9	46.8	49.9	55.6

表 3　ILSVRC2015_VID 数据集上各种重建模型评估结果

Tab. 3　Evaluation results of various reconstruction models on ILSVRC2015_VID datasets

Method	Scene metric					
	NIMA	PARNAC	NeuralSBS	NIQE	MOS	LPIPS
Bicubic	4.563	0.252	0.103	20.94	1.85	0.319
MSRResNet	5.09	0.305	0.488	21.24	3.25	0.164
SRFBN	5.094	0.306	0.499	21.12	—	0.16
MSRGAN	5.314	0.347	0.575	19.18	2.17	0.025

续表 3　Continued Tab. 3

Method	Scene metric					
	NIMA	PARNAC	NeuralSBS	NIQE	MOS	LPIPS
ESRGAN (GAN)	5.343	0.344	0.606	18.38	—	0.013
Bicubic	4.215	0.34	0.13	18.84	—	0.353
MSRResNet	4.602	0.393	0.488	19.12	—	0.107
SRFBN	4.609	0.394	0.502	18.91	—	0.094
MSRGAN	4.632	0.395	0.518	17.69	—	0.023
ESRGAN (GAN)	4.637	0.396	0.54	16.9	—	0.003
Bicubic	4.105	0.31	0.128	18.81	1.36	0.333
MSRResNet	4.492	0.363	0.486	19.09	2.18	0.087
SRFBN	4.499	0.364	0.5	18.88	—	0.074
MSRGAN	4.522	0.365	0.516	17.66	3.45	0.003
ESRGAN (GAN)	4.527	0.366	0.538	16.87	—	0.017

2.3　消融实验

在本文中，介绍了 NeuralSBS 这种新的参考重整测量超分辨率工具，仅挑战不同 SR 模型及其超参数之间的比较。通过大量的实验，神经系统表现出不存在近似人类对超龄偏好的参考测量值。NeuralsBSB 设计的动机是一个现实的实际场景。

2.4　可视化评价

本文提出的方法可以显著提高基于深度学习的 SR 结果，与当今主流的方法对整脸和局部脸可视化结果对比如图 9 和图 10 所示。本文的重建模型能够

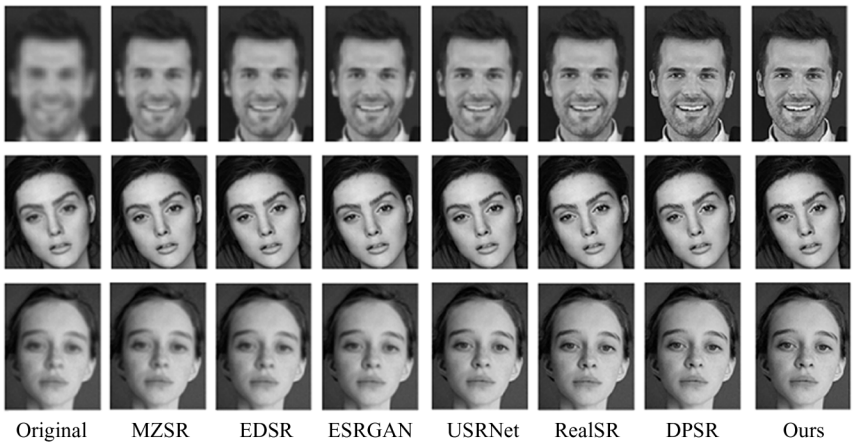

图 9　真实人脸整脸细节重建效果可视化

Fig. 9　Visualization of full face detail reconstruction of real faces

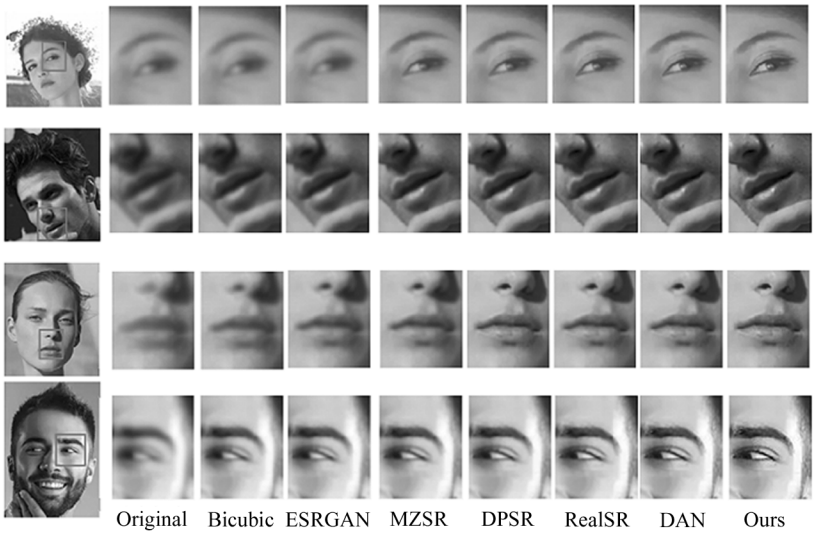

图 10　真实人脸局部细节重建效果可视化

Fig. 10　Visualization of partial detail reconstruction of real faces

适应真实 LR 图像中出现的图像特定退化,并能够重建尖锐的物体边界和无噪声的图像。将分割损失引导与域自适应引导相结合,重建出逼真的纹理并保证颜色的一致性。在真实和合成的 LR 图像上的实验结果表明,与传统的方法相比,该方法有显著的改进,得到了更少的噪声和更好的视觉质量。这得到了人类对超分辨率图像的排名的支持,在这里,本文方法大大优于其他方法。

3　结　论

本文针对传统 SR 方法无法重建出边缘特征图像,提出了一种先验信息与密集连接网络模型重建方案,利用了考虑输入统计信息的残差特征的不同组合。此外,提出模型还引入了注意力机制,避免输入的身份特征急剧漂移。通过与剩余结构协作,在不增加额外模块的情况下提高了网络性能。该网络设计方案具有很好的鲁棒性。通过与主干网络结构协作,在不增加额外模块的情况下提高了网络性能。在基准数据集上的实验结果表明,本文提出方法与复杂结构的传统模型相比,具有更好的或可比的性能。

参考文献:

[1] SUN C W,CHEN X. Multiscale feature fusion back-projection network for image super-resolution [J]. Acta Automatica Sinica,2021,47(7):1689-1700.
孙超文,陈晓.基于多尺度特征融合反投影网络的图像超分辨率重建[J].自动化学报, 2021, 47(7):1689-1700.

[2] WU J A,GUO R,LIU R Z,et al. Edge area constraint guided filter depth image super-resolution reconstruction algorithm[J]. Infrared and Laser Engineering, 2021, 50(1):20200081.
武军安,郭锐,刘荣忠,等.边缘区域约束的导向滤波深度像超分辨率重建算法[J].红外与激光工程,2021,50(1):20200081.

[3] CAI T J,PENG X Y,SHI Y P,et al. Channel attention and residual concatenation network for image super-resolution [J]. Optics and Precision Engineering,2021,29(1):142-151.
蔡体健,彭潇雨,石亚鹏,等.通道注意力与残差级联的图像超分辨率重建[J].光学精密工程,2021,29(1):142-151.

[4] YANG S C,WANG H F,WANG Y H,et al. Super-resolution reconstruction algorithm based on deep learning mechanism and wavelet fusion [J]. Journal of Beijing University of Aeronautics and Astronautics,2020,46(1):189-197.
杨思晨,王华锋,王月海,等.深度学习机制与小波融合的超分辨率重建算法[J].北京航空航天大学学报,2020,46(1):189-197.

[5] DONG C,LOY C C,HE K,et al. Image super-resolution using deep convolutional networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 38(2):295-307.

[6] MA N N,ZHANG X Y,ZHENG H T,et al. ShuffleNet V2: Practical guidelines for efficient CNN architecture design [C]//European Conference on Computer Vision (ECCV), September 8-14, 2018, Munich, Germany. Berlin: Springer-Verlag,2018:122-138.

[7] SAHARIA C,HO J,CHAN W,et al. Image super-resolution via iterative refinement[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(4): 4713-4726.

[8] YU Y,QIAO Y F,ZHONG X,et al. Super-resolution reconstruction of staring imaging degraded model[J]. Acta Optica Sinica,2017,37(8):116-125.
余烨,乔彦峰,钟兴,等.凝视成像降质模型的超分辨率重建[J].光学学报,2017,37(8):116-125.

[9] CAO M M,GAN Z I,CUI Z G,et al. Novel neighbor embedding face hallucination based on non-negative weights and 2D-PCA feature[J]. Journal of Electronics & Information Technology,2015,37(4):777-783.
曹明明,干宗良,崔子冠,等.基于 2D-PCA 特征描述的非负权重邻域嵌入人脸超分辨率重建算法[J].电子与信息学报,2015,37(4):777-783.

[10] HOWARD A, SANDLER M, CHU G, et al. Searching for mobilenetv3[C]//IEEE/CVF International Conference on Computer Vision, October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE, 2019:1314-1324.

[11] KIM H,KIM J,WON S,et al. Unsupervised deep learning for super-resolution reconstruction of turbulence[J]. Journal of Fluid Mechanics,2021,910:A29.

[12] BULAT A,PEREZ,RUA J M,Sudhakaran S,et al. Space-time mixing attention for video transformer[J]. Advances in Neural Information Processing Systems, 2021, 34: 19594-19607.

[13] CHEN J,LI B,XUE X. Scene text telescope:Text-focused scene image super-resolution[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition,June 20-25,2021,Nashville,TN,USA. New York:IEEE,2021:12021-12030.

[14] ZHANG X,ZHOU X,LIN M,et al. Shufflenet:An extremely efficient convolutional neural network for mobile devices[C]//IEEE Conference on Computer Vision and Pattern Recognition,June 18-23,2018,Salt Lake City,UT,USA. New York:IEEE,2018:6848-6856.

[15] HE K,ZHANG X,REN S,et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence,2014,37(9):1904-1916.

[16] QIU D,ZHENG L,ZHU J,et al. Multiple improved residual networks for medical image super-resolution[J]. Future Generation Computer Systems,2021,116:200-208.

[17] CHEN Y,LIU L,PHONEVILAY V,et al. Image super-resolution reconstruction based on feature map attention mechanism[J]. Applied Intelligence,2021,51(7):4367-4380.

[18] SURYANARAYANA G,CHANDRAN K,KHALAF O I,et al. Accurate magnetic resonance image super-resolution using deep networks and Gaussian filtering in the stationary wavelet domain[J]. IEEE Access,2021,9:71406-71417.

作者简介：

崔立尉 (1985—),女,硕士,讲师,主要从事软件开发、图像处理方向的研究.