

卷积神经网络的人脸超分辨率重建算法

杨洪臣 郝 韵

(中国刑事警察学院公安信息技术与情报学院 辽宁 沈阳 110035)

摘 要: 针对当前人脸超分辨率算法中存在的效率低和重建失真等问题, 提出一种改进的基于CNN的人脸超分辨率网络。网络的始端通过引入同态滤波器进行预处理来丰富人脸细节特征, 并且优化卷积神经网络的逻辑权重和运算过程, 解决了无效信息对重建的影响和重建失真的问题。该算法无论是在主观视觉质量, 还是在峰值信噪比和结构相似性等客观评价指标方面都明显优于现有其他算法, 具有一定的实用价值。

关键词: 超分辨率重建 卷积神经网络 同态滤波器

中图分类号: TP391.41

文献标识码: A

文章编号: 2095-7939 (2020) 05-0104-06

DOI: 10.14060/j.issn.2095-7939.2020.05.015

1 引言

人脸超分辨率 (Super-Resolution, SR) 又被称为“人脸幻想”, 是指由低分辨率 (Low-Resolution, LR) 人脸图像向高分辨率 (High-Resolution, HR) 人脸图像重建的过程。当前各领域均有与人脸有关的实战需求, 例如人脸识别、人脸对齐、表情识别和三维人脸重建等, 然而这些都是基于清晰的高分辨率人脸数据集实现的, 在处理低分辨率人脸图像时, 效果会大打折扣。因此, 人脸超分辨率重建技术在计算机视觉、生物识别、公安侦查领域显得尤其重要。

早期的人脸SR算法停留在对低等级特征研究阶段, Wang等^[1]提出了全局特征脸的概念, 利用主成分分析法从LR的人脸图像中提取信息特征进行特征转换来实现人脸SR重建。但这种方法适用环境单一, 无法处理多幅不同姿态下的模糊人脸图像。随着深度学习技术的兴起, 传统SR算法逐渐被取代, 基于自动学习的超分辨率图像重建算法被广泛应用。目前可查阅的重建方法主要有基于差

值、基于重建和基于学习3种。基于插值^[2]的图像重建方法早期出现且实时性很强, 但是重建后图像边缘特征丢失, 只能局限于对单帧图像的重建, 无法实现批量化图像处理; 基于重建^[3]的超分辨率算法是目前图像重建领域被广泛研究的一类算法, 重建效果比较理想, 但是计算量较大, 无法满足图像重建对实时性的要求; 基于学习^[4]的重建方法, 是近年来图像处理领域的主流算法, 可以自动学习到各级网络之间的层次化特征参数, 在输出图像和输入图像之间建立一一对应的关系, 实现自动学习重建的效果。在自动学习重建人脸图像发展的早期阶段, Dong等^[5]提出了SRCNN (Super-Resolution Convolutional Neural Network) 算法, 该算法可以满足重建对于清晰度的要求, 但因为对应向量级数过大降低了重建效率^[6]。Dong等^[7]在原有SRCNN的基础上提出了FSRCNN (Fast Super-Resolution Convolutional Neural Network)。FSRCNN的网络运行速度远超前SRCNN, 但是由于神经网络在收敛性和收敛速度之间无法实现平衡导致重建过程的精确度太低。

收稿日期: 2019-11-05

基金项目: 辽宁省网络安全执法协同创新中心开放课题 (编号: KFKT20190012)。

作者简介: 杨洪臣 (1964-), 男, 吉林梨树人, 中国刑事警察学院公安信息技术与情报学院教授, 主要从事图像检验与视频侦查技术研究。

针对上述问题, 本文提出了一种改进后的基于卷积神经网络的人脸图像超分辨率重建算法。网络的输入端使用分辨率较低的人脸图像, 在提取人脸特征之前经过同态滤波器预处理来丰富人脸细节特征, 再用池化和deconv1 (第一反卷积层) 和deconv2 (第二反卷积层) 共同实现非线性映射, 最后利用两个反卷积操作完成上取样重建获得清晰人脸图像。与其他算法相比, 本文方法同时满足了图像重建对精确度和重建效率的要求。

2 理论基础

2.1 FSRCNN

FSRCNN和SRCNN同为两种基于深度学习的SR网络, FSRCNN是对SRCNN的改进算法。首先, SRCNN需要使用bicubic法放大图像尺寸, 算法复杂且耗时较长, 而FSRCNN经卷积操作即可实现目标图像的放大效果。其次, FSRCNN使用的卷积核更小, 降低网络的特征维数的同时提高了运算速度。另外, FSRCNN的映射层部分可以共同使用, 在训练不同采样倍数的模型时, 改变最后的反卷积层对低分辨率图像进行上采样重建即可, 节省了训练时间。改进后的网络虽然运行速度几乎达到了SRCNN的17倍, 但是由于将原始低分辨率图像直接输入到网络当中导致冗余信息和无效信息过多, 人脸图像

重建效果较差, 无法同时满足重建对精确度和实时性的要求。

2.2 ESPCN

ESPCN的核心概念是亚像素卷积层^[8]。使用亚像素卷积神经网络的SR算法直接在LR空间进行特征提取, 在该网络中, 机器自动学习差值函数以便实现对图像的放大操作, 卷积运算过程均在分辨率较低的小尺寸图像上进行。该方法的学习效率得到提升满足了重建对实时性的要求, 但降低了ESPCN的重建性能。

3 本文方法

在上述两种重建方法的基础上, 提出一种利用卷积神经网络来进行人脸超分辨率重建提升人脸分辨率的方法, 将deconv1和deconv2与池化层组合实现非线性映射, 使用反卷积操作来代替ESPCN当中的亚像素卷积层, 同时提升了整个网络的超分辨率重建质量和运算速度。在网络始端加入同态滤波器, 解决了FSRCNN中由于直接输入低分辨率图像导致人脸细节特征不足重建质量较差的问题。网络结构的搭建由同态滤波器预处理阶段、特征提取阶段、收缩阶段、非线性映射阶段、扩张阶段、反卷积重建阶段6个部分组成。图1是本文的网络结构图。

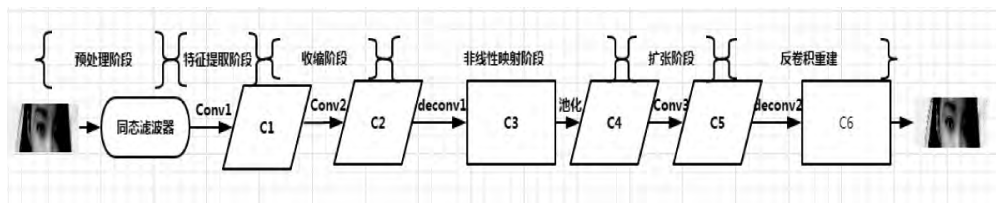


图1 本文的网络结构图

3.1 同态滤波器预处理

根据人脸图像成像的原理, 可以将人脸图像的物理特性用经过镜头处理过的“入射—反射”模型来表示。我们所得到的图像是由入射分量和反射分量共同作用产生的, 即:

$$H(x, y) = Q(x, y) \times L(x, y) \quad (1)$$

式中, 入射分量 $Q(x, y)$ 涵盖了人脸上绝大部分可以被使用的细节特征, 而反射分量 $L(x, y)$ 表示的是低频且对人脸图像质量几乎没有影响的反射部分。因此需要将低频的反射分量剥离出来, 采用同态滤波器降低 $L(x, y)$ 的动态范围。假设压缩后的 $L(x, y)$ 为 $L'(x, y)$, 那么处理后的图像变为:

$$H'(x, y) = Q(x, y) \times P'(x, y) \quad (2)$$

设 $z(x, y) = \ln H(x, y) = \ln P(x, y) + \ln Q(x, y)$, 引入傅里叶变换可以得到:

$$F(z(x, y)) = F(\ln P(x, y)) + F(\ln Q(x, y)) \quad (3)$$

$$\text{即: } Z(u, v) = F_P(u, v) + F_Q(u, v) \quad (4)$$

采用同态滤波函数对 $z(u, v)$ 进行处理可得:

$$S(u, v) = H(u, v) z(u, v) = H(u, v) F_P(u, v) + H(u, v) F_Q(u, v) \quad (5)$$

对经过同态滤波器处理后的人脸图像进行反变换可以得到:

$$S(x, y) = F^{-1}S(u, v) = F^{-1}(H(u, v) F_P(u, v) + H(u, v) F_Q(u, v))$$

$$(u,v)) + F^{-1}(H(u,v)F_0(u,v)) \quad (6)$$

以上就是使用同态滤波器对人脸图像进行预处理的过程, 处理后的人脸图像模糊部分被去除, 脸部的细节信息得到有效增强, 解决了冗余和无效信息给网络造成的影响。接下来用卷积神经网络对人脸图像的进行重建和校正, 使得人脸的清晰度进一步增强。

3.2 提取特征部分

将经过预处理的人脸图像从网络的输入端输入, 第一个conv1的目的是对输入的预处理后的图像进行一次特征提取, 提取过程是用一组卷积核在预处理后的图像上沿着水平和竖直方向依次滑动得到一组特征图。实现特征提取功能的卷积操作如公式(7)所示:

$$F_1(Y) = \sigma(P_1 \times Y + b_1) \quad (7)$$

式中, P_1 表示 a_1 个尺寸为 $m \times z_1 \times z_1$ 的卷积核(其中 m 表示输入图像的通道数, z_1 表示卷积核的尺寸), Y 表示经过同态滤波器预处理的人脸图像, \times 表示卷积操作, b_1 表示 a_1 维的偏差, 激活函数为 $\sigma(X)$, $F_1(Y)$ 表示经过第一个卷积层conv1之后得到的特征图像(特征图像的数量为 a_1)。

在用CNN来预测特征图像时, 图像每经过一次卷积运算, 它的尺寸就会变小一次。例如, 将一个大小为 $(y+1) \times (y+1)$ 的图像在输入端输入, 它的感受野为 $y \times y$, 经过 y 个卷积层后, 输出的特征图尺寸则变为 1×1 。经过对其他超分辨率重建的算法的研究发现, 绝大多数算法都是根据周围的像素情况来估计中心像素, 因此周围像素的情况决定着最终重建图像效果的好坏。由于经过卷积神经网络特征图会变小, 利用CNN的超分辨重建算法会对最终生成的特征图进行裁剪, 导致所得到的特征效果图尺寸减小很多。因此需要在每个卷积层进行卷积操作之前先进性填零操作(padding), 本文将设padding = $(z_1 - 1) / 2$, 以保证所有经过卷积操作之后的特征图的尺寸相同, 这样就有效防止图像随着特征提取的过程而逐渐变小的现象出现。填零操作的示意图如图2所示。

3.3 网络的收缩阶段和损失函数

为了减少网络参数降低运算复杂度, 采用10个大小为 1×1 的小卷积核对经卷积层生成的特征图像进行降低维度处理, 随着特征图像维度的降低, 网络计算的复杂程度也随之下降从而提升了网络的训练速度。另外在损失函数方面, 本文使用样本方差

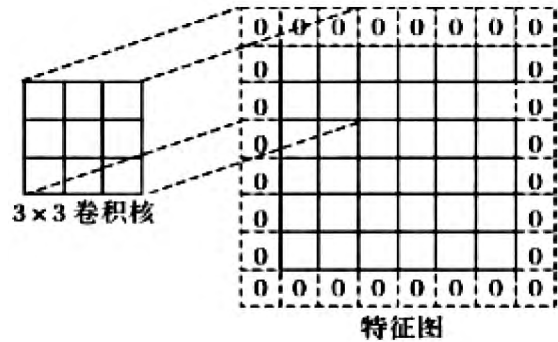


图2 填零示意图

作为损失函数, 如公式(8)所示:

$$L = \frac{1}{N-1} \sum_{m=1}^N \|Y - X_m\|^2 \quad (8)$$

式中, Y 表示经过神经网络预测得到的人脸图像, X_m 表示真实的人脸图像样本。先求出网络预测图像与真实获取图像的离差的平方, 再对此变量取平均数作为样本方差。样本方差用来表示一组数据的变化程度, 在这里代表了整个网络操作的误差范围。

3.4 非线性映射与网络扩张

非线性映射阶段需要将反卷积操作和池化过程结合起来, 对经过第一次卷积生成的特征图像进行进一步的放大从而提取到更加有利于重建的特征信息, 这样可以满足非线性映射部分所需要的条件。经处理后的特征图会按照一定比例的放大, 如公式(9)所示:

$$F_2(Y) = \sigma(p_2 \odot F_1(Y) + b_2) \quad (9)$$

式中, P_2 表示 a_2 个尺寸为 $n_1 \times z_2 \times z_2$ 的反卷积核(z_2 为deconv1的大小), $F_1(Y)$ 表示经过第一次特征提取所得到的特征图片, \odot 表示反卷积操作, b_2 表示 a_2 维的偏差, 激活函数为 $\sigma(X)$, $F_2(Y)$ 表示经过第二次反卷积操作之后得到的成比例放大的特征图像(特征图像的数量是 a_2)。

另外池化过程与卷积操作类似但是它不受反向传播的影响, 它可以将相邻的多个特征点用一个特征来代替, 因此可以降低特征图的维数, 压缩数据和参数的数量, 从而提升训练效率。该操作如公式(10)所示:

$$F_3(Y) = P_3 \odot F_2(Y) \quad (10)$$

式中, P_3 表示 a_2 个尺寸为 $n_2 \times z_3 \times z_3$ 的滤波器(z_3 为所用滤波器的尺寸大小), $F_2(Y)$ 表示经过第二次反卷积后得到的同比例放大的特征图像, \odot 表示池化, $F_3(Y)$ 表示经池化处理后得到的 a_2 张特征图。假设输入的预处理后的图像的尺寸大小为

$M \times N$ ，经反卷积后的特征图尺寸如公式（11）所示，式中， q 代表零填充的过程， s 指步长， z_2 代表反卷积核的尺寸， f 代表放大因子。

$$[s(H-1)-2q+z_2] \times [s(W-1)-2q+z_2] = fM \times fN \quad (11)$$

经过非线性映射后的网络扩张阶段同样采用50个小卷积核（ 1×1 ），对池化层处理后得到的特征图进行扩张维度处理，可以进一步改善人脸图像的重建效果。网络结构最终获得的高分辨率图像是由低分辨率图像上采样得来的，对采样得到的高分辨率图像进行重组完成反卷积重建最终获得清晰人脸图像。

4 仿真测试

仿真实验选用如下配置：软件部分在系统中搭建了caffe^[9]开源框架，在对算法测试时采用MatlabR2016a；硬件部分选择了服务器配置Intel Core i5型号的处理器和GTX1080Ti型号的显卡，以及50GB的内存。

4.1 数据准备

在理论数据验证层面，本文使用CASPEAL-R1的正脸图像，91images和Face Recognition Technology人脸库中的正脸图像作为训练集，采用的测试数据集为set14^[10]、set5^[11]、BSD200^[12]。为了提升训练效果，将数据集91images当中每一张经过训练的图像分别旋转45°、90°、135°、180°来扩大数据集。实际应用验证层面，本文分别随机截取了一段视频当中的模糊人脸图像、人脸眼部细节特征图像和视频中的模糊车牌图像作为训练对象，将训练对象放大3倍后进行超分辨率重建，对比重建效果。

4.2 仿真结果及分析

4.2.1 客观仿真数据比较结果

在图像处理领域，两个最常被使用的评价指标分别是PSNR（Peak Signal to Noise Ratio）峰值信噪比和SSIM（Structural Similarity）结构相似指数。其中PSNR常常用简单的均方误差来定义，它表示的是两张图片对应像素点之间的误差，其值越大表明误差越低，图像的重建效果越好。SSIM表示两张图像之间的相似度，相似度越接近于1，表明输出越接近于高分辨图像，重建效果越好。

分别在2倍、3倍、4倍放大倍率下，通过对比PSNR和SSIM值来将本文算法和Bicubic^[13]、FSRCNN、ESPCN算法进行比较。当网络测试中基准训练数据集是Set5，测试集是CASPEAL-R1的正

脸图像，测试集和训练集的PSNR值对比如表1所示。当训练集是CASPEAL-R1的正脸图像和Face Recognition Technology人脸库中的正脸图像时，测试集和训练集的PSNR值对比如表2所示。可以看出，本文方法在Set5、Set14和BSD200上PSNR的效果要普遍的优于其他算法。

表1 各测试集经不同方法处理后的PSNR值

| Data set | Factor | Bicubic | FSRCNN | ESPCN | 本文方法 |
|----------|------------|---------|--------|-------|-------|
| Set5 | $\times 2$ | 33.66 | 36.34 | 36.55 | 36.81 |
| | $\times 3$ | 30.39 | 32.39 | 32.59 | 32.71 |
| | $\times 4$ | 28.42 | 30.09 | 30.28 | 30.29 |
| Set14 | $\times 2$ | 30.23 | 32.18 | 32.28 | 32.29 |
| | $\times 3$ | 27.54 | 29.00 | 29.13 | 29.14 |
| | $\times 4$ | 26.00 | 27.20 | 27.32 | 27.31 |

表2 CASPEAL-R1中的正脸图像经不同方法处理后的PSNR值

| Data set | Factor | Bicubic | FSRCNN | ESPCN | 本文方法 |
|----------|------------|---------|--------|-------|-------|
| Set5 | $\times 2$ | 33.66 | 36.34 | 36.55 | 36.78 |
| | $\times 3$ | 30.39 | 32.39 | 32.59 | 32.87 |
| | $\times 4$ | 28.42 | 30.09 | 30.28 | 30.51 |
| Set14 | $\times 2$ | 30.23 | 32.18 | 32.28 | 32.32 |
| | $\times 3$ | 27.54 | 29.00 | 29.13 | 29.19 |
| | $\times 4$ | 26.00 | 27.20 | 27.32 | 27.39 |
| BSDS | $\times 2$ | 29.70 | 31.38 | 31.44 | 31.62 |
| | $\times 3$ | 27.26 | 28.28 | 28.36 | 28.93 |
| | $\times 4$ | 25.97 | 26.73 | 26.83 | 26.92 |

经过不同放大倍率，每一个测试集用不同重建方法处理后SSIM的平均值如表3所示。可见除了本文算法在3倍放大倍率下略低于ESPCN以外，其他方面本文算法均优于其他算法。

表3 各测试集经不同方法处理后的SSIM值

| Data set | Factor | Bicubic | FSRCNN | ESPCN | 本文方法 |
|----------|------------|---------|--------|--------|--------|
| Set5 | $\times 2$ | 0.9299 | 0.9521 | 0.9544 | 0.9546 |
| | $\times 3$ | 0.8675 | 0.9033 | 0.9088 | 0.9067 |
| | $\times 4$ | 0.8104 | 0.8530 | 0.8603 | 0.8623 |
| Set14 | $\times 2$ | 0.8687 | 0.9039 | 0.9056 | 0.9063 |
| | $\times 3$ | 0.7736 | 0.8145 | 0.8188 | 0.8168 |
| | $\times 4$ | 0.7019 | 0.7413 | 0.7471 | 0.8623 |
| BSDS | $\times 2$ | 0.8625 | 0.9024 | 0.9026 | 0.9030 |
| | $\times 3$ | 0.7638 | 0.8038 | 0.8078 | 0.8074 |
| | $\times 4$ | 0.6949 | 0.7291 | 0.7359 | 0.7394 |

4.2.2 主观图像仿真效果比较结果

在案件侦查过程中会因为得到的原始图像效果不理想而无法进行同一认定。为了更好地展现本文

算法在公安实战工作中的应用价值, 分别选用一段案件视频当中截取的模糊人脸图像和模糊车牌图像进行图像超分辨率重建的实验。将模糊图像输入到网络结构当中, 分别将经改进之后的CNN重建算法和其他几种算法进行对比, 可以发现经本文方法处

理后的效果图相比于其他传统图像重建算法在清晰度和图像精确度方面均得到显著提升。图3为随机截取的一段监控视频当中的模糊人脸图像, 先将该图像放大3倍, 然后分别使用几种图像重建方法对该模糊人脸进行重建并比对重建效果。

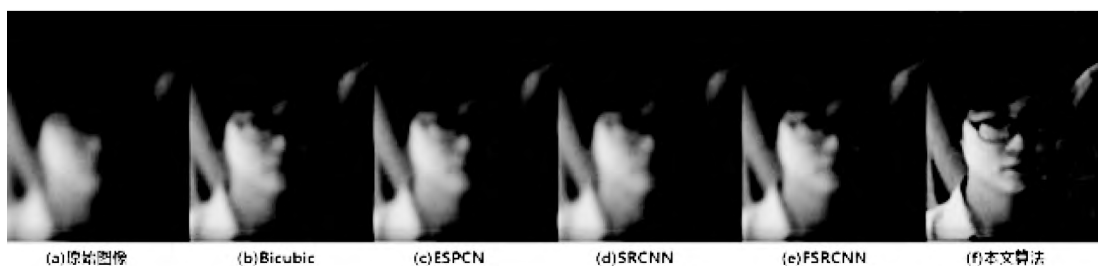


图3 3倍放大倍率下模糊人脸图像重建效果比对

图4为本文算法对人脸细节特征的重建效果展示, 由于本文算法采用了较大的感受野和更加复杂的神经网络结构, 相比于其他几种算法在眉毛、眼

睛等部位的边缘会更加锐利, 模糊部位较少, 可以更好地展现脸部眉眼的细节特征。



图4 3倍放大倍率人脸细节重建效果比对

图5分别使用Bicubic、ESPCN、SRCNN、FSRCNN和本文的基于神经网络的算法对经典训练

集CASPEAL-R1的正脸图像进行重建测试, 分别比较各个重建方法对人脸的重建效果。



图5 3倍放大倍率 CASPEAL-R1 库的正脸图像的重建效果比对

为了进一步验证本文方法在实际案件中的应用价值, 截取了一段真实案件视频监控中的模糊车牌图像作为输入端输入到网络结构, 将截取图片放大3倍来比较重建效果。图6是各个算法在模糊车牌图

像3倍放大倍率重建效果示意图。可以发现, 本文算法不仅可以实现人脸重超分辨率, 在模糊车牌图像处理中也可以进一步提升清晰度。

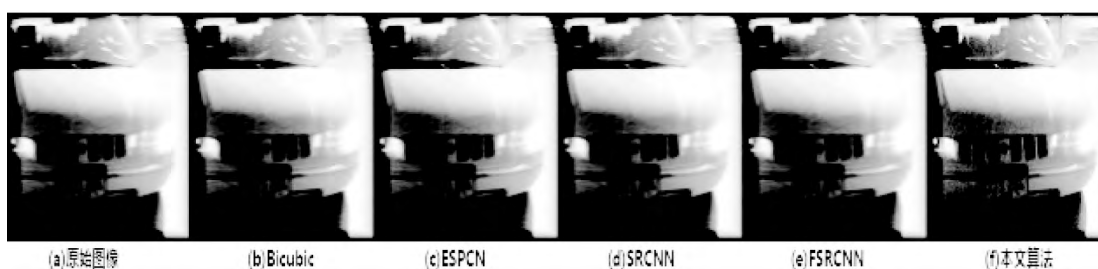


图6 3倍放大倍率模糊车牌的重建效果比对

5 结论

本文算法首先引入同态滤波器压缩人脸图像的动态范围满足了图像重建对于实时性的要求, 采用小卷积核降低图像维度从而提升重建效率, 将反卷积核与池化层结合在一起实现对图像敏锐特征的提取, 通过扩维理念进一步改善重建效果, 最后使用反卷积核对面脸图像进行上采样完成重建。结果表明, 通过卷积神经网络重建模糊人脸图像可以凸显人脸细节, 并且清晰度和人眼舒适度均得到提高, 重建速度加快。下一步可以更深层次的挖掘该算法的实用价值, 将其应用到公安机关的车牌识别监控、破损纸币快速识别等其他各种低分辨率图像重建领域。

参考文献:

- [1]Wang X G,Tang X O. Hallucinating face by eigentransformation[J]. IEEE Transactions on Systems Manand Cybernetics,2005(3): 425-434.
- [2]陆志芳,钟宝江.基于预测梯度的图像插值算法[J].自动化学报,2018(6):1072-1085.
- [3]南方哲,钱育蓉,行艳妮,等.基于深度学习的单图像超分辨率重建研究综述[J].计算机应用研究,2020(2):321-326.
- [4]孙京,袁强强,李冀玮,等.亮度-梯度联合约束的车牌图像超分辨率重建[J].中国图像图形学报,2018(6):802-813.
- [5]Dong C,Chen C L,He K,etal.Image super-resolution using deep convolutional networks[J].IEEE Transactions on Pattern Analysis & Machine Intelligence,2016(2):295-307.
- [6]杨洪臣,党京,蔡能斌,等.BP神经网络车牌识别技术优化研究[J].中国刑警学院学报,2019(3):120-124.
- [7]Marani R,Renò V,Nitti M,etal.Amodifiedinterative closest point algorithm for 3D point cloud registration [J].Computer-Aided Civil and Infrastructure Engineering,2016(7):515-534.
- [8]Zhou L Y,Yang M H,Fast direct super-resolution via sparse representation[J].Computer Engineering & Design,2016(12):3290-3294.
- [9]Jaderberg M,Simonyan K,Vedaldi A,etal.Reading text in the wild with convolutional neural networks [J].International Journal of Computer Vision,2014(1):1-20.
- [10]Yang R H, Hua X H,You Y S.Local n-TH order surface fitting and curvature estimation based on scattered-point cloud data[J].Journal of Geodesy and Geodynamics,2013(3):141-143.
- [11]Bevilacqua M ,Roumy A ,Guillemot C,et al.Single-image super-resolution via linear mapping of interpolated self-examples [J].IEEE Transactions on Image Processing,2014(12):5334-5347.
- [12]Martin D R ,Fowlkes C C ,Malik J.Learning to detect natural image boundaries using local brightness color and texture cues [J].IEEE Transactions on Pattern Analysis & Machine Intelligence,2004(5):530-549.
- [13]苏衡,周杰,张志浩.超分辨率图像重建方法综述[J].自动化学报,2013(8):1202-1213.

(责任编辑: 于 萍)

Face Super-resolution Reconstruction Algorithm for Convolutional Neural Networks

YANG Hong-Chen HAO Yun

(School of Criminal Science and Technology, Criminal Investigation Police University of China, Liaoning Shenyang 110035)

Abstract: Aiming at the problems of inefficiency and reconstruction distortion in the current face super-resolution algorithm, an improved face super-resolution network based on CNN is proposed. In order to solve the problem of the effect of invalid information on the reconstruction and the distortion of reconstruction, the paper introduces homomorphic filter to enrich the features of face detail and optimize the logical weight and operation process of convolutional neural network. The experimental results show that the algorithm is superior to other existing algorithms in terms of subjective visual quality, peak signal-to-noise ratio and structural similarity and has certain practical value.

Key words: Super-resolution Reconstruction, Convolutional Neural Network, Homomorphic Filter