



计算机工程
Computer Engineering
ISSN 1000-3428, CN 31-1289/TP

《计算机工程》网络首发论文

题目: 基于生成对抗网络的多幅离焦图像超分辨率重建算法
作者: 斯捷, 肖雄, 李涇, 马明勋, 毛玉星
DOI: 10.19678/j.issn.1000-3428.0058848
网络首发日期: 2020-10-12
引用格式: 斯捷, 肖雄, 李涇, 马明勋, 毛玉星. 基于生成对抗网络的多幅离焦图像超分辨率重建算法. 计算机工程. <https://doi.org/10.19678/j.issn.1000-3428.0058848>



网络首发: 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式 (包括网络呈现版式) 排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

出版确认: 纸质期刊编辑部通过与《中国学术期刊 (光盘版)》电子杂志社有限公司签约, 在《中国学术期刊 (网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊 (网络版)》是国家新闻出版广电总局批准的网络连续型出版物 (ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。



基于生成对抗网络的多幅离焦图像 超分辨率重建算法

斯捷¹, 肖雄², 李泾³, 马明勋², 毛玉星²

(1.浙江图盛输变电工程有限公司, 浙江温州 325000; 2.重庆大学电气工程学院, 重庆 400044;

3.国网浙江电力有限公司温州供电公司, 浙江温州 325000;)

摘 要: 为了提高超分辨率算法重建出的图像质量, 提出了一种基于对抗网络, 融合多幅离焦图像的超分辨率重建算法。以离焦图像作为切入点, 首先利用自编码器提取了离焦图像中的重要特征; 然后利用空间特征变换层结构, 将离焦特征与原始特征相结合; 最后完成图像的超分辨率重建。在 Celeb A 人脸数据集上进行了测试, 并且与传统插值算法及 SRGAN 算法进行了对比。测试结果证明, 从峰值信噪比及结构相似性两个指标来看, 本文算法在大多数情况下能获得更高的数值。这意味着基于多幅离焦图像的超分辨率算法能生成更为优秀的重建图像, 在电力传输与运维场景视觉监控领域具有一定的实用价值。

关键词: 自编码器; 图像特征提取; 生成对抗网络; 超分辨率重建; 深度神经网络

开放科学(资源服务)标识码(OSID):



Super-Resolution Reconstruction with Multi-Frame Defocused Images Based on Generative Adversarial Network

Si Jie¹, Xiao Xiong², Li Jing³, Ma Mingxun², Mao Yuxing²

(1. Zhejiang Tusheng Transmission and Transformation Engineering Co. Ltd., Wenzhou 325000, China;

2. College of Electrical Engineering, Chongqing University, Chongqing 400044, China;

3. State Grid Zhejiang Electric Power Co., Ltd. Wenzhou Power Supply Company, Wenzhou 325000, China)

【Abstract】 In order to improve the quality of the image reconstructed by the super-resolution algorithm, a super-resolution reconstruction algorithm based on adversarial network and multi-frame defocused images is proposed. The algorithm uses the defocused image as the entry point. First, it uses an auto-encoder to extract the important features in the defocused image; then uses the spatial feature transformation layer structure to combine the defocused features with the original features; and finally completes the super-resolution of the image reconstruction. Tested on the Celeb A face data set, and compared with the traditional interpolation algorithm and SRGAN algorithm. The test results prove that from the two indicators of peak signal-to-noise ratio and structural similarity, the algorithm we use can obtain higher values in most cases. This means that the super-resolution algorithm based on multi-frame defocused images can generate better reconstructed images, so it has certain practical value in the field of power transmission and operation and maintenance scene visual monitoring.

【Key words】 auto encoder; image feature extraction; generative adversarial net; super resolution; deep neural network

DOI: 10.19678/j.issn.1000-3428.0058848.

0 概述

图像是人类视觉的基础, 生活中的大多数感知均与图像息息相关。而图像中存储的信息量大小以及图像细

节的精细程度则取决于图像分辨率的高低。通常情况下, 一张细节信息丰富且视觉效果清晰的图像, 大都具有更多的像素, 也就是更高的分辨率。随着信息化时代的不断

基金项目: 国家重点研发计划项目“物联网智能感知终端平台系统与应用验证”(项目号 2018YFB2100100)

作者简介: 斯捷(1976—), 男, 工程师, 主要研究方向: 计算机视觉、配电施工和运维; 肖雄(1997—), 男, 硕士研究生, 主要研究方向: 图像处理、计算机视觉; 李泾(1988—), 男, 工程师, 主要研究方向: 配电网视觉监测、运维检修; 马明勋(1994—), 男, 硕士研究生, 主要研究方向: 图像处理、计算机视觉; 毛玉星(1967—), 男, 副教授, 工学博士, 主要研究方向: 图像处理、计算机视觉、智能电网通信技术。

E-mail: myx_group@163.com

断发展,为了获得更加良好的观感体验,人们对分辨率的要求越来越高。然而,受限于信息传输、硬件设备、成像原理等客观因素的制约,图像的分辨率往往难以达到令人满意的水平。因此,图像超分辨率重建(Super Resolution, SR)作为一门能够将图像由低分辨率转化为高分辨率,丰富图像细节信息的图像处理技术,引起了国内外研究学者的广泛关注。对超分辨率算法的研究不仅能促使相关理论技术的完善,同时还能在人们的日常生活中产生重要的应用价值。

超分辨率重建技术自从上个世纪 80 年代被提出,至今发展已有 40 余年,近年来国内外学者对超分辨率问题展开了许多方面的研究。从研究对象来分,超分辨率算法可以分为二维图像的超分辨率、二维流形数据的超分辨率和三维数据的超分辨率。而目前的超分辨率算法的研究大都集中于二维图像的超分辨率问题上,而关于二维流形数据和三维数据的研究仍较少,相关工作主要是将二维图像的超分辨率技术思想推广到不同领域。下面将重点从近年来基于学习的算法来介绍超分辨率重建。

2014 年, Dong 等人^[1-2]提出了基于卷积神经网络的超分辨率重建算法,通过学习低分辨率图像与高分辨率图像间的非线性映射关系,从而完成特征的提取以及图像的重建。与之前研究的方法相比,该方法具有更为优秀的重建效果。在这之后,各式各样的网络结构开始被提出。在 Kim 等人^[3-4]的研究中,他们大大增加了卷积网络的深度,同时引入残差学习结构,递归跳跃层机制来改善模型中的收敛率问题。在 2018 年的 CVPR 会议中, Yulun Zhang^[5]团队利用了残差密集结构,从密集连接的卷积层中提取出大量的局部特征,其作用是用于稳定更大范围网络的训练,并从以前的和当前的局部特征中学习有效的特征。Lim 等人^[6]通过移除不必要的批量归一化层来改善深度网络结构中的梯度消失问题,以稳定训练过程,同时提出了一种多尺度的超分辨率重建方法,这些都取得了良好的效果。

为了进一步改善图像的质量,牛津视觉几何小组提出了一种基于 VGG^[7]网络的感知损失函数^[8]。小组成员结合损失函数的优点,提出了在图像超分辨率任务中利用感知损失函数训练前馈网络的方法。该方法产生了令人满意的视觉效果,从而取代了逐像素损失函数。

随着生成对抗网络(Generative Adversarial Nets, GAN)的提出, Ledig^[9]将 GAN 的框架结构运用到了图像

超分辨率重建算法中并命名为 SRGAN。他们将对损失函数与感知损失函数相结合,利用生成网络和判别网络之间的博弈关系,使得重建出的图像在细节纹理和肉眼感知上更加真实,而非单单只是提升图像的峰值信噪比(PSNR)指标。作为一项开创性工作, SRGAN 为后一步的基于 GAN 的 SR 算法奠定了基础。基于这些工作, Bulat^[10]专注于利用 GAN 实现任意姿势的脸部 SR 算法并精确定位面部位置。另一篇文献^[11]通过附加一个在特征域中工作的额外判别网络来产生更真实的结果。通常来说,成对的训练集数据是通过高清图像下采样所获得,然而在 Zhang^[12]的研究中指出真正的图像降级不遵循这个假设,输入真实图像将产生糟糕的性能。因此,他们提出了一个框架,可将不同程度的模糊核和噪声水平输入到网络中,从而运用到真实图像中。面对同样的问题, Bulat^[13]提出了一种两步走的方法,在重建之前先训练一个将高分辨率图像转化为低分辨率图像的网络,用于正确的学习图像退化,再利用训练集进行重建训练。

最近,在文献^[14]中,作者将 GAN 网络的基础模块改为 residual-in-residual dense block (RRDB),并且改善了感知损失函数的输出方式。另外,还引入了相对判别网络来预测相对的真实性。在文献^[15]中,作者发现,通过将分割概率图结合在一起,可以恢复出忠实于语义类的纹理。在本文提出的算法中,其原理与参考文献^[16]有着密切的关系,使用卷积自编码器获得离焦特征图,并借用空间特征变换层将特征图作为联合输入到 GAN 网络中。

综上所述,本文提出了一种基于对抗网络,融合多幅离焦图像的超分辨率重建算法,主要工作如下:构建了基于运动模糊和高斯模糊的卷积去噪自编码器模型,实现图像的特征提取及去模糊过程;研究基于生成对抗网络的 SR 算法,完成与离焦图像的特征融合。基于 Celeb A 人脸图像数据集进行模型的训练学习,通过 PSNR (峰值信噪比)和 SSIM (结构相似性)这两种常用的图像质量评价指标以及对比图像的视觉效果,来定量且定性验证本文算法的有效性。

1 本文方法

1.1 基于自编码器的离焦图像特征提取

1.1.1 离焦图像的生成模拟

运动模糊的模型原理主要在水平匀速直线运动条件

下展开探讨。然而,在获取神经网络训练所需要的成对的清晰模糊图像时,并没有简单的方法可以达到实现很好的运动模糊效果。

在文献[17]中,作者首先利用运动向量来代表小块图像区域的运动核,然后对模糊向量进行离散化,根据不同的长度与方向生成 73 个候选线性运动核,最后与自然图像进行卷积,从而生成综合模糊图像。在文献[18]中同样使用线性运动内核创建综合模糊图像。文献[19]中,作者在一个有限尺寸的网格中随机采样 6 个点用线段相连,并将线段上每个像素的核值设置为从均值为 1、标准差为 0.5 的高斯分布中采样的值。然后将这些值裁剪为正,并将内核标准化为单位和,以此来创建模糊内核。

在运动模糊部分,根据前人的实验,本文使用了能更加真实地模拟了复杂的模糊核的算法。这种算法遵循了 Boracchi 和 Foi 于 2012 年在文献[20]中所描述的随机轨迹生成的概念。轨迹是通过在一个规则的像素网格上连续采样得到,且每条轨迹由一个粒子在连续域中的二维随机运动的位置组成。粒子有一个初始速度矢量,在每次迭代中,它受到指向当前粒子位置的高斯扰动和确定性惯性分量的影响。此外,在很小的概率下,会出现一个使粒子速度反转的脉冲扰动,以此模拟用户按下相机按钮或补偿相机抖动时发生的突然运动。在每一步中,速度都会被归一化以保证相同曝光的轨迹有相同的长度。每个扰动都由自己的参数控制,当所有扰动参数设置为 0 时生成直线轨迹。

每个轨迹矢量都是一个复杂矢量,由马尔可夫过程完成整个轨迹的生成。如图 1 所示,轨迹的下一个位置由前一个位置的速度、脉冲参数及高斯扰动方程所共同决定。最后对轨迹矢量应用亚像素插值法生成模糊核。

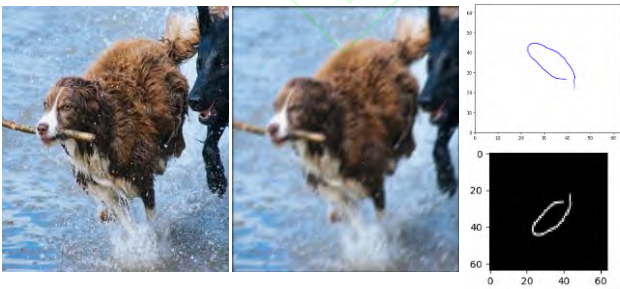


图 1 运动模糊模拟示意图

Fig.1 Diagram of motion blur simulation

1.1.2 基于自编码器的图像特征提取

去噪自编码器是一类接受损失数据作为输入,并训练来预测原始未被损坏的数据作为输出的自编码器,其

与普通自编码器相比增加了一个数据的有损处理,例如噪声,模糊等。在有限的损失处理下,数据间的结构相关性仍然得到了保存,通过特征提取后,依旧可以重建出原始图像。

本文将模糊后的图像视作有损数据,构建去模糊自编码器来对模糊图像进行编码处理,去除图像冗余信息;再通过解码操作,恢复原始清晰图像,从而训练卷积神经网络更好的获取图像中的关键特征部分,并加以压缩,为后续图像的超分辨率重建部分打下基础。

去模糊自编码器的训练过程如下:

1) 首先引入一个损坏过程 $C(\tilde{x}|x)$, 这个条件代表针对给定数据样本 x , 对其进行损坏处理,例如噪声、模糊等,生成样本 \tilde{x} 。

2) 将损坏样本 \tilde{x} 输入到编码器中,可表示为如下公式,其中 s 为激活函数:

$$h = f_{\theta}(\tilde{x}) = s(W\tilde{x} + b) \quad (1)$$

$$\theta = \{W, b\} \quad (2)$$

3) 经过编码后的数据输入到解码器中,解码部分表示为:

$$z = g_{\theta'}(h) = s(W'h + b') \quad (3)$$

$$\theta' = \{W', b'\} \quad (4)$$

4) 针对单个训练样本数据对,构建损失函数 $L(x, z)$, 用于衡量解码输出 z 与原始数据样本 x 之间接近程度:

$$L(x, z) = \|x - z\|_2^2 \quad (5)$$

5) 对于具有 n 个样本的训练集,每一个样本数据 $\tilde{x}^{(i)}$ 可以得到相应的 $h^{(i)}$ 与 $z^{(i)}$ 。当平均重构误差最小时可以获取最优的编码器解码器网络参数。

$$\theta, \theta' = \arg \min_{\theta, \theta'} \frac{1}{n} \sum_{i=1}^n L(x^{(i)}, z^{(i)}) \quad (6)$$

同时,为防止权值过大引起的过拟合,对上式(6)设置一个权重衰减项:

$$D = \frac{1}{2} \|\omega\|_2^2 \quad (7)$$

整合上面两式得到目标函数,其中 λ 为权重衰减参数,此时 $h^{(i)}$ 即为 $\tilde{x}^{(i)}$ 的低维度特征表示:

$$J = \frac{1}{2n} \sum_{i=1}^n L(x^{(i)}, z^{(i)}) + \lambda D \quad (8)$$

1.2 基于离焦图像的对抗超分辨率重建算法

1.2.1 网络结构

1) 生成网络

生成对抗网络是通过对抗的方式,去学习数据分布,从而产生新的数据。整个框架由生成网络和判别网络两部分组成。在超分辨率重建中,生成网络定义为 G_θ , 负责将一张低分辨率 LR 图像重建为超分辨率 SR 图像;判别网络定义为 D_η , 作用是将重建出的 SR 图像与原始 HR 图像进行对比,判断重建图像的真实性。生成网络与判别网络分别由参数 θ 与 η 所确定,整个过程交替进行,通过对抗的方式解决一个最大化最小化问题:

$$\min_{\theta} \max_{\eta} \mathbb{E}_{I^{HR} \sim p_{HR}} \log D_\eta(I^{HR}) + \mathbb{E}_{I^{LR} \sim p_{LR}} \log(1 - D_\eta(G_\theta(I^{LR}))) \quad (9)$$

其中 p_{HR} 与 p_{LR} 分别为 HR 图像与 LR 图像各自训练集的分布。

生成网络结构的中间部分由 16 个残差模块构成,层与层之间以及整个模块前后都有跳跃连接,从而减少每层信息承载,加速模型收敛。在网络最后设置 2 个亚像素卷积层,使初始 LR 图像分辨率倍增 4 倍,以实现超分辨率重建的效果。整个重建过程的思想是先使用较深神经网络对低分辨率图像信息进行整合,再摒弃复杂无效的反卷积,利用亚像素卷积操作完成重建。

2) 判别网络

在卷积神经网络中,感受野(Receptive Field)的定义是卷积神经网络每一层输出的特征图上的像素点在输入图片上映射的区域大小,通常感受野的设定是越大越好。然而,更大的感受野往往意味着更大的卷积核以及更多的网络参数。卷积操作中,输出矩阵高和宽可由公式(10)和(11)计算:

$$height_{out} = (height_{in} - height_{kernel}) / stride + 1 \quad (10)$$

$$width_{out} = (width_{in} - width_{kernel}) / stride + 1 \quad (11)$$

其中 $stride$ 代表步长,假设一张图像的大小为 28×28 , 利用 5×5 大小的卷积核,步长为 1,则输出图像大小为 $(28 - 5) / 1 + 1 = 24$, 参数为 $5 \times 5 \times c$, 其中 c 为输入输出矩阵深度的乘积。卷积核的大小与网络参数成正比。

在判别网络结构里的八个卷积层中,均采用 3×3 尺寸的卷积核。除此之外,使用了 Leaky ReLU 激活函数并取消了常规卷积过程中的最大池化操作。整个过程中,特征图由 64 张增加到 512 张,每次特征图倍增的同时均使用步长为 2 的卷积核,以减少输出尺寸。最后利用两个全连接网络和一个 sigmoid 激活函数,来获得样本分类的概率,以此判别真伪。结构如下图所示,其中每个卷积层对应的内核大小(k)、特征映射的数量(n)和步

长(s)

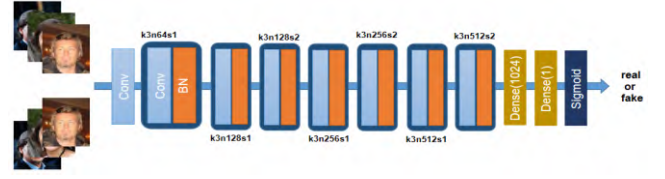


图 2 判别网络的架构

Fig.2 Structure of discriminator

1.2.2 损失函数

损失函数即目标函数,算法的求解过程即是对这个目标函数优化的过程。为了使超分辨率重建后的 SR 图像能与 HR 图像尽可能的相似,损失函数将分为像素域与感知域两个部分。像素域使用 MSE 损失函数,通过将两类不同图像逐像素对比,以保证获得更高的 PSNR 分数。基于 MSE 损失函数的算法会常常导致输出过度平滑的结果,使人眼感知效果不佳,因此需要感知域的损失函数进行限制。

作用于感知域的感知损失函数称为 perceptual loss,它度量了两幅图像特征之间的欧氏距离,因此使用它来约束 SR 图像与 HR 图像之间的特征相似性,以保证人眼感知效果。为了获取特征间的距离,感知损失函数通常使用一个已经过预训练的用于图像分类的网络模型,再将两类图像输入模型以对比深层次的特征。此处使用了经典图像识别网络 vgg-16。最后,基于交叉熵的对抗损失也被用来帮助重建图像。生成网络的损失函数定义为:

$$L_G = \eta L_{mse} + L_{vgg} + \lambda L_{Gan}^g \quad (12)$$

其中 η 与 λ 为平衡不同损失项的系数。

1.2.3 相对判别网络

在标准生成对抗网络中,就未经过函数激活的 $C(x)$ 而言,判别网络可被简单定义为 $D(x) = \text{sigmoid}(C(x))$, 根据 $D(x)$ 的大小来判断图像的真伪。

与论文[14]中相似,为了保持判别网络的输出与真伪图像两者均能产生联系,本文引入相对判别概念以改进判别网络。针对真伪图像对 (x_r, x_f) , 定义 $D(\hat{x}) = \text{sigmoid}(C(x_r) - C(x_f))$ 。这样的定义为整个网络增加了相对性,判别网络将会估计给定真实图像 x_r 比虚假图像 x_f 更真实的概率。同理, $D(\hat{x}) = \text{sigmoid}(C(x_r) - C(x_f))$ 代表着相对更虚假的概率。由此, (9)式的第一项可以改写为:

$$\log D_\eta(I^{HR}) \rightarrow \log D_\eta(\text{sigmoid}(C(x_r) - C(x_f))) \quad (13)$$

其中 $I^{HR} \sim p_{HR}$, $x_f \sim G_\theta(I^{LR})$, $I^{LR} \sim p_{LR}$ 。

由于原式再第一项没有生成网络 G_θ 的参与, 梯度为 0, 所以在优化中通常被忽略。而改进之后, 第一项受到 G_θ 生成的 x_f 的影响, 梯度不再为 0 后, 自然能够帮助生成网络完成损失函数优化, 使生成网络的训练变为整个公式的最小化过程而非仅有后半部分。

由于批量输入的图像对存在不同的组合方式,为了简化算法时间复杂度,可加入均值概念,进一步对判别网络进行重新定义:

$$\tilde{D}(x) = \begin{cases} \text{sigmoid}(C(x) - \mathbb{E}_{f \sim \mathbb{Q}} C(x_f)) & \text{if } x \text{ is real} \\ \text{sigmoid}(C(x) - \mathbb{E}_{r \sim \mathbb{N}} C(x_r)) & \text{if } x \text{ is fake} \end{cases} \quad (14)$$

当损失函数修改为相对损失之后，损失函数将衡量真假图片间的相对真实性。在对抗性训练中，两种损失可以从生成数据和真实数据的梯度中均获得收益，从而获得更好的效果。

1.2.4 基于空间特征变换层的信息融合

本文利用改进的去噪自编码器，针对离焦图像的特征进行压缩重组训练，网络中经过训练的编码部分能实现特征的提取功能。而要利用这些特征图对超分辨率重建过程产生帮助，则需要用到信息特征融合技术。

为了将自编码器与 GAN 的框架相结合,利用已经训练好的自编码器的编码部分,输入离焦图像;再将编码部分输出的离焦特征图添加到图像重建的过程中;最后,使用空间特征变换(Spatial Feature Transform, SFT)网络层结构来实现特征的融合。在空间特征变换过程中,针对不同的模糊类型特征图像,先进行一个合并运算操作:

$$I = \text{Concat}(i_\sigma, i_m) \quad (15)$$

其中 i_σ 与 i_m 分别代表高斯模糊及运动模糊的特征图。

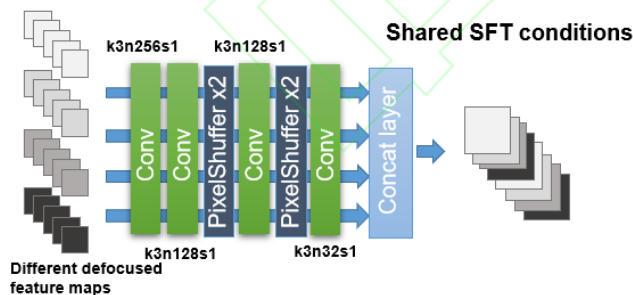


图 3 特征图合并操作

Fig.3 Feature map merge operation

基于合并而成的特征图这一先验信息 I ， SFT 网络层通过一个映射关系输出调制参数对 (γ, β) ：

$$\gamma = M(I) \quad \beta = N(I) \quad (16)$$

函数 M 与 N 可以是任意映射算法，在本文中使用了具有两层深度的神经网络作为映射算法以便优化。在获得

了 (γ, β) 之后, 将 γ 定义为缩放参数, β 定义为移动参数, 运用类仿射变换的方式, 将合并运算与条件伸缩运算相结合:

$$SFT(F | \gamma, \beta) = F \odot \gamma + \beta \quad (17)$$

其中 F 代表 SFT 层输出作为后续网络输入的特征图,其维度与 γ 和 β 相同。 \odot 代表按元素相乘。由于保留了空间维度, SFT 层不仅执行特征操作, 还执行空间转换。这样的结构能够有效的转换和操作空间信息。

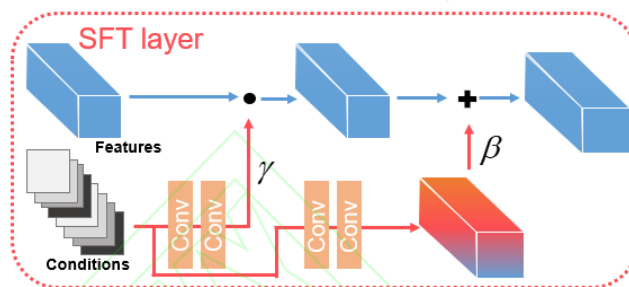


图 4 空间特征变换层

Fig.4 Spatial Feature Transform network

由于 *SFT* 层计算量不大，并且无法得知模型何处需要使用到条件信息，故在每一个残差模块中均加入 *SFT* 层，整个生成器网络结构如下：

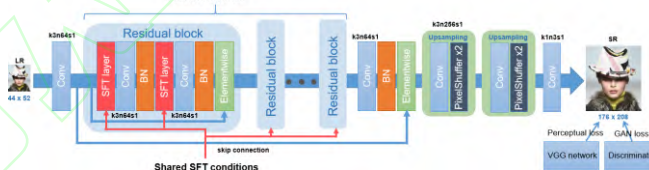


图 5 生成网络的架构

Fig.5 Structure of generator

2 实验结果与分析

2.1 离焦图像特征提取实验

整个训练过程中使用 Celeb A 数据集，在离焦图像训练集获取时，分别对数据集进行了高斯模糊与运动模糊处理，高斯模糊直接使用了 Opencv 中的高斯模糊函数，半径参数设置为 2 以下的随机数。首先，将离焦图像训练集输入到自编码器网络中，在前面的网络层中完成特征提取降维工作。然后利用后续的网络层将隐藏层中的特征进行解码放大，输出一张尺寸与输入相同的图像。利用均方误差损失函数通过与未离焦原图的对比完成优化，整个编码解码过程类似于图像去模糊过程。在参数设置上，将学习率设置为 1×10^{-4} ，Adam 优化器参数设置为 0.9，训练次数 30000 次，低分辨率图像的尺寸设置为 44×52 。具体结构如下图所示，包含核尺寸 (k)、特征图数量 (n) 与步长 (s)。

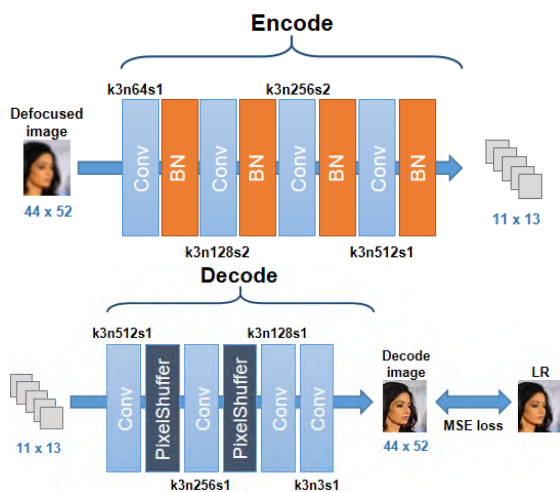


图 6 自编码器结构及实现过程

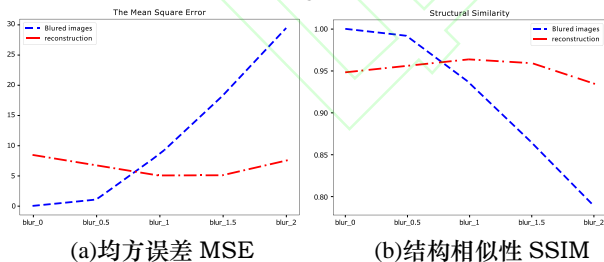
Fig.6 Structure and process of AE

实验过程中,离焦图像作为有损数据输入到模型中,图像的损失程度与解码后的效果息息相关。为探究模型能够处理的模糊程度的边界,将不同离焦度的图像作为输入,再对比重建后的效果。当离焦程度从 0 到 2 变化时,重建图像效果如图 7 所示,其中第一行为离焦图,第二行为重建图。



图 7 不同离焦程度图像的重建效果示例

Fig.7 Examples of image reconstruction with different defocus degrees



(a)均方误差 MSE

(b)结构相似性 SSIM

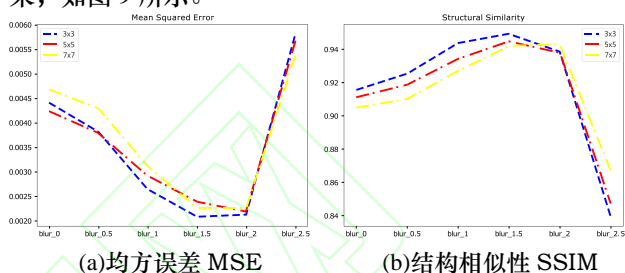
图 8 不同离焦程度图像效果对比

Fig.8 Index comparison of different defocus degrees

图 8 中横轴表示图像的不同离焦程度,纵轴表示评价指标的数值。对比图像的均方误差 (MSE) 与结构相似性 (SSIM) 指标,从蓝色线条代表的离焦图像可以得出随着模糊程度的加大,结构相似性在下降,均方误差值在增加;由红色线条代表的重建图像可以得出,模糊

程度为 1 时重建图像取得最佳效果,而当模糊程度较小或较大时,模型的重建效果较为一般。考虑到收敛速度及网络参数的原因,以上实验中使用到的卷积核均为 3×3 大小。

根据传统反卷积维纳算法,其逆副核虽然有限但能够提供足够大的空间支持,这表明反卷积可以通过够大的空间卷积近似。由于离焦图像的获取基于卷积过程,所以在不考虑收敛速度的情况下,同样对比了三种不同规模卷积核 (3×3 , 5×5 , 7×7) 所训练模型的重建效果,如图 9 所示。



(a)均方误差 MSE

(b)结构相似性 SSIM

图 9 不同尺寸卷积核模型效果对比

Fig.9 Comparison of effects of different size convolution kernel models

图 9 中横轴表示图像的不同离焦程度,纵轴表示均方误差 (MSE) 与结构相似性 (SSIM) 评价指标的数值。根据重建出的效果可以看出,作为有损降维算法,卷积自编码器能够完成离焦图像的降维重建工作。虽然较大卷积核需要更多的训练时间,在后续实验证明,在离焦程度较小时,卷积核选择 3×3 尺寸效果最佳;随着离焦程度的不断增大,较大的卷积核能够更好的完成空间卷积的近似;而面对训练集中未出现的更大范围的离焦现象,模型的重建效果急剧下滑。

2.2 超分辨率重建实验

考虑到实验设备与环境,本文的实验基于 Celeb A 人脸图像数据集,人脸图像特征更集中,利于模型的训练学习。此外,在感知损失函数部分,使用了基于 ImageNet 预训练的 VGG-19 网络。实验的分辨率倍增数设置为 4 倍,以便与文献 SRGAN 进行对比。所有的低分辨率图像均从高分辨率图像插值下采样获得,以便重建后的图像与原高分辨率图像进行质量指标的相关计算。

设置高分辨率图像 HR 的尺寸为 176×208 ,相应的低分辨率图像 LR 的尺寸为 44×52 。为了更好的从离焦图像中获取可能的补充信息,在离焦图像的处理上实验并未设置特别大的离焦半径,高斯模糊最大半径设置为

1, 而运动模糊最大设置为 4。在这样的设置下, 自编码器的解码部分仍能重建出较好的结果。

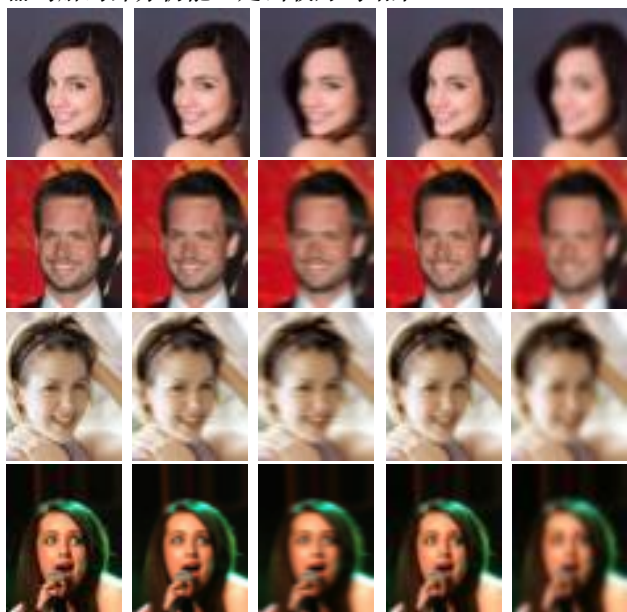


图 10 输入图像示例

Fig.10 Sample of input images

在上一部分的实验中, 自编码器完成了离焦图像的特征提取以及重建工作。接着利用已训练好的自编码器的编码部分, 通过 SFT 网络层结构, 与生成对抗网络相结合。上图为输入图像示例, 分别为低分辨率图像以及不同模糊处理后的图像。

为了更好的比较实验结果, 本文引入其他的超分辨率重建算法作为比较, 例如双三次插值以及同样在 CelebA 上训练出的 SRGAN 算法。基于多幅离焦图像重建的效果如图 11 所示, 从左到右依次为双三次插值算法, SRGAN 算法, 本文提出的算法以及原 HR 图像。

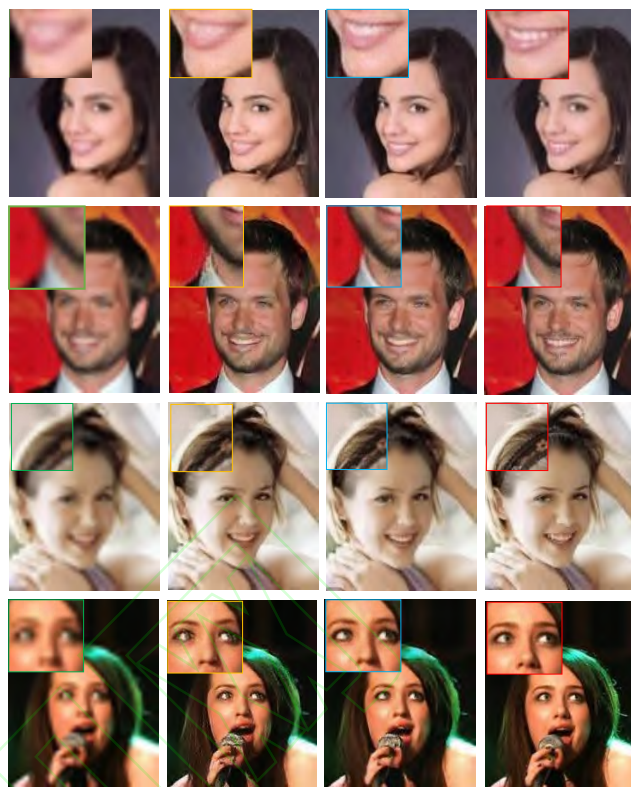


图 11 重建效果图对比

Fig.11 Comparision of reconstruction

从视觉对比中可以看出, 双三次插值方法产生的图像质量较差, 大量细节丢失, 导致边缘高度模糊, 出现一些不希望出现的伪影。与双三次插值图像相比, SRGAN 图像有明显的改善, 图像更加清晰; 然而, 由于一些虚假细节的存在, 该算法依然导致了整张脸部看起来不那么真实。相比之下, 通过多幅离焦图像重建出的图像视觉效果更为真实, 并且边缘信息也得到了保留。为了能更加精确的比较几种算法, 本文通过峰值信噪比 (PSNR) 与结果相似性 (SSIM) 这两种常用的评价图像质量的指标进行计算, 几种算法的结果如下表所示。

表 1 三种超分算法测试结果对比

Table 1. Comparison of test results of three Super-Resolution algorithms

Number in Celeb A	Indexes	Bicubic	SRGAN	Proposed	HR
202497	PSNR	28.5180	28.5072	29.0340	∞
	SSIM	0.9728	0.9748	0.9766	1
202498	PSNR	27.9415	30.8092	31.0274	∞
	SSIM	0.9794	0.9811	0.9807	1
202499	PSNR	25.7485	27.0001	29.4404	∞
	SSIM	0.9740	0.9763	0.9754	1

202500	PSNR	27.7923	27.6750	28.6462	∞
	SSIM	0.9686	0.9705	0.9752	1
202506	PSNR	28.9671	27.6459	28.7575	∞
	SSIM	0.9648	0.9586	0.9664	1
202508	PSNR	27.9254	28.0268	27.6590	∞
	SSIM	0.9683	0.9749	0.9736	1
202513	PSNR	27.0109	26.6491	27.1592	∞
	SSIM	0.9556	0.9544	0.9570	1
202529	PSNR	27.6403	25.9164	27.0461	∞
	SSIM	0.9732	0.9696	0.9766	1

从表中显示出的结果来看,某些情况下,双三次插值算法获得了高于其他的 PSNR 值,结合其视觉质量效果可以发现,这与前文关于 PSNR 的介绍以及论文 SRGAN 中的结论一致:衡量像素间相似度的 PSNR 无法真实反映图像的感知质量,获得最高的 PSNR 值有时并不意味着最好的超分辨率重建效果。从 SSIM 值来看,本文所提出的算法在大多数情况下能获得更高的数值,即使在未取得最高的时候也与数值最高的 SRGAN 相近。这意味着基于多幅离焦图像的超分辨率算法能生成更为优秀的重建图像。

3 结束语

本文以离焦图像作为切入点,研究图像超分辨率重建算法。针对离焦图像的特性,结合去噪自编码器模型,完成了特征提取与重建。在融合离焦特征信息时,采用空间特征变换层网络,以类仿射变换的方式将自编码器模型的编码部分与每一个残差块结合。将模型在 Celeb A 人脸数据集上进行训练,对比双三次插值,SRGAN 超分辨率的重建效果可以看出,本文提出的算法能更好的恢复图像细节,并且再大多数情况下也获得了更高的 PSNR 与 SSIM 数值。实验结果表明,本文提出的结合离焦图像的超分辨率重建算法能获得更好的图像重建效果,在电力传输与运维场景视觉监控领域具有一定的实用价值。在后续的研究中,应朝着深度学习模型轻量化,加快训练速度,降低权重参数数量的方向研究,以保证图像重建的效果看起来更加真实。

参考文献

[1] Dong C, Loy C C, He K, et al. Learning a deep convolutional network for image super-resolution[C]// 2014 European Conference on Computer Vision. Springer, Cham,

2014.

[2] Dong C, Loy C C, He K, et al. Image super-resolution using deep convolutional networks[J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2016, 38(2):295-307.

[3] Kim J, Lee J K, Lee K M. Accurate image super-resolution using very deep convolutional networks[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2016.

[4] Kim J, Lee J K, Lee K M. Deeply-recursive convolutional network for image super resolution [C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2016.

[5] Zhang Y, Tian Y, Kong Y, et al. Residual dense network for image super-restoration[C]// 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2018.

[6] Lim B, Son S, Kim H, et al. Enhanced deep residual networks for single image super-resolution[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE, 2017

[7] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. OALib Journal, Computer Science, 2014, 2(1): 136-144.

[8] Johnson J, Alahi A, Fei-Fei L. Perceptual losses for real-time style transfer and super-resolution[C]// 2016 European Conference on Computer Vision. Springer, Cham, 2016.

[9] Ledig C, Theis L, Huszar F, et al. Photo-realistic single image super-resolution using a generative adversarial network[J]. arXiv e-prints, 2016, 11(1):1-14.

[10] Bulat A, Tzimiropoulos G. Super-FAN: Integrated facial landmark localization and super-resolution of real-world low resolution faces in arbitrary poses with GANs[C]// 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2018.

[11] Park S, Son H, Cho S, et al. SRFeat: single image super-resolution with feature discrimination[C]// 2018 European Conference on Computer Vision. Springer, Cham, 2018.

[12] Zhang K, Zuo W, Zhang L. Learning a single convolutional super-resolution network for multiple degradations[J]. Proceedings / CVPR, IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2018, 6(1):124-133.

[13] Bulat A, Yang J, Tzimiropoulos G. To learn image super-resolution, use a GAN to learn how to do image degradation first[C]// 2018 European Conference on Computer Vision. Springer, Cham, 2018.

[14] Wang X, Yu K, Wu S, et al. Esrgan: enhanced

super-resolution generative adversarial networks[C]// 2018 European Conference on Computer Vision workshops. Springer, Cham, 2018.

[15] Wang X, Yu K, Dong C, et al. Recovering realistic texture in image super-resolution by deep spatial feature transform[C]// 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2018.

[16] Larsen A B L, Sonderby Søren Kaae, Winther O, et al. Autoencoding beyond pixels using a learned similarity metric[C]// 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2015.

[17] Sun J, Cao W, Xu Z, et al. Learning a convolutional neural network for non-uniform motion blur removal[C]// 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2015.

[18] Li Xu, Jimmy SJ Ren, Ce Liu, et al. Deep convolutional neural network for image deconvolution[C]// 2014 Neural Information Processing Systems. (NIPS). MIT Press, 2014.

[19] Leibe B, Matas J, Sebe N, et al. A Neural Approach to Blind Motion Deblurring[J]. [Lecture Notes in Computer Science] Computer Vision – ECCV, 2016, 9907(1):221-235.

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