doi:10.19665/j.issn1001-2400.2019.03.022

结合注意力机制的人脸超分辨率重建

陈晓范，申海杰，边 倩，王振铎，田新志
（西安思源学院 电子信息工程学院，陕西 西安 710038）

摘要： 因受成像设备限制，得到的人脸图像分辨率通常较低，针对此问题提出了一种将生成对抗网络和注意力机制相结合的方法，来对人脸图像进行多尺度超分辨率重建。将深度残差网络和深度神经网络分别作为生成器和判别器，并将注意力模块与深度残差网络中的残差块相结合，重建出与高分辨率图像高度相似且难以被判别器区分的超分辨率人脸图像。实验结果证明，所提出的方法能够有效地提升人脸图像的分辨率，同时也证明了注意力机制在图像细节信息重建中的重要作用。

关键词： 超分辨率重建；生成对抗网络；注意力机制；深度残差网络；深度神经网络
中图分类号：TP391 **文献标识码：**A **文章编号：**1001-2400(2019)03-0148-06

Face image super-resolution with an attention mechanism

CHEN Xiaofan, *SHEN Haijie*, *BIAN Qian* , *WANG Zhenduo* , *TIAN Xinzhi*
(School of Electronical and Information Engineering, Xi'an SiYuan Univ., Xi'an 710038, China)

Abstract: Because of the limitation of the imaging equipment, the face images captured by it usually have the problem of low resolution and low quality. This paper proposes a method based on the generative adversarial network and attention mechanism for the multi-scale super-resolution of face images. In this paper, the deep residual network and the deep convolutional neural network (VGG-net) are used as the generator and the discriminator, respectively. The attention modules are combined with the residual blocks in the deep residual network to reconstruct face images which are highly similar to the high-resolution images and difficult for the discriminator to distinguish. Experimental results demonstrate the effectiveness of the proposed method in multi-scale face image super-resolution and the important role of the attention mechanism in image detail reconstruction.

Key Words: super-resolution; generative adversarial network; attention mechanism; deep residual network; deep convolutional neural network

人脸图像为人类视觉感知和计算机分析提供了重要信息。但由于受到成像设备的限制，通常只能获得低分辨率的人脸图像，这就在一定程度上大大影响了我们对人脸信息的理解。而超分辨率重建是一种能够有效提高图像分辨率的方法。目前超分辨率重建的算法大致可分为两类，基于单帧图像的超分辨率重建以及基于多帧图像的超分辨率重建。其中，由单帧低分辨率图像重建出相应的高分辨率图像主要是通过深度学习的方法来实现的，该方法通过学习低分辨率(Low-Resolution, LR)图像到高分辨率(High-Resolution, HR)图像之间的映射关系从而达到提升图像分辨率的目的。目前，已通过大量研究结果证明，基于学习的超分辨率重建算法可获得更好的图像视觉效果并且可以实现多尺度的超分辨重建。自生成对抗网络(Generative Adversarial Network, GAN)被文献[1]所提出并证实了其在图像生成领域的惊人效果以来，

收稿日期：2018-11-06 **网络出版时间：**2019-03-04
基金项目：国家自然科学基金(81571772)；西安思源学院自然科学研究基金(XASY-B1803)；陕西省教育厅自然科学研究基金(17JK1073)
作者简介：陈晓范(1980－)，男，讲师，E-mail：xianfan_chen321@126.com.
网络出版地址：http://kns.cnki.net/kcms/detail/61.1076.TN.20190301.1706.002.html

GAN 便受到了广大图像超分辨率重建方向的研究者们的青睐。文献[2]提出了基于 GAN 的图像超分辨率重建算法(Super Resolution Generative Adversarial Network, SRGAN),利用生成对抗网络和感知损失函数实现了图像纹理信息的精确重建。文献[3]提出了一种基于 Wasserstein GAN (WGAN)的人脸超分辨算法,该算法克服了 GAN 本身训练困难、生成样本缺乏多样性等问题,使得人脸重建效果得到了明显的提升。虽然目前人脸超分辨率重建的研究成果有很多,但这些方法大多数仅通过加深网络层数[4]或堆叠残差块的个数来实现,这种做法并不能很有效地提升图像重建的精度。

受文献[5-7]中注意力机制思想的启发,笔者将其与生成对抗网络相结合,并提出了一种新的人脸超分辨率重建的网络。该网络由两部分构成,即生成网络和对抗网络。其中,生成网络主要是基于深度残差网络,可以大大提升训练速度和训练效果,将注意力机制模块与深度残差网络中的残差块相结合,使得网络能够有目的地进行学习,这样更有助于图像细节信息的准确重建。

笔者所提出的基于生成对抗网络与注意力机制的人脸超分辨重建算法主要有以下优点:

(1) 引入注意力机制,并将其与残差块相结合,使得网络能够更有效地学习,从而大大提升了图像的重建效果;

(2) 将用于衡量图像像素空间相似度的 L_1 损失函数与用于衡量图像特征空间相似度的感知损失函数相结合,使得网络在注重图像像素信息重建的同时,兼顾了图像的特征信息,L_1 损失函数的引入也使得网络的收敛速度有所提升。

1　生成对抗网络结构

由于受到文献[5]和文献[7]中所提出工作的启发,设计了一个基于注意力机制的对抗生成网络来完成人脸超分辨率重建的任务,其网络结构如图 1 和图 2 所示。该网络是由基于深度残差网络的生成网络(图 1(a))和基于深度卷积网络(Visual Geometry Group ,VGG)的判别网络(图 2)构成。

1.1　生成网络

如图 1(a)所示,生成网络主要是由浅层特征提取模块、基于残差块的深度特征提取模块、图像上采样模块和图像重建模块 4 部分组成,以此来学习低分辨率图像 I_{HR} 与高分辨率图像 之间的映射关系,从而生成出相应的超分辨率(Super Resolution, SR)图像。参考文献[2]、[8]中的研究,这里使用一层卷积层来从低分辨率图像 I_{LR} 中提取浅层特征信息 F_0,其表达式如下:

$$F_0 = G_{\mathrm{SF}}(\mathrm{I}_{\mathrm{LR}}), \tag{1}$$

其中,G_{SF} 表示卷积操作。接着,提取的浅层信息进入深度残差网络中来提取更深层的特征信息 F_0,然后,使用图像上采样模块进行相应尺度的放大操作。

$$F_{\mathrm{DF}} = G_{\mathrm{DR}}(\mathrm{F}_0), \tag{2}$$

$$F_{\mathrm{UP}} = G_{\mathrm{UP}}(F_{\mathrm{DF}}), \tag{3}$$

其中,G_{DR} 表示深度残差网络,共包含 16 个残差块,G_{UP} 表示图像上采样模块。经上采样操作后的特征图由一层卷积层进行重建,生成最终的超分辨重建后的图像 I_{SR}。

$$I_{\mathrm{SR}} = G_{\mathrm{REC}}(\mathrm{F}_{\mathrm{UP}}), \tag{4}$$

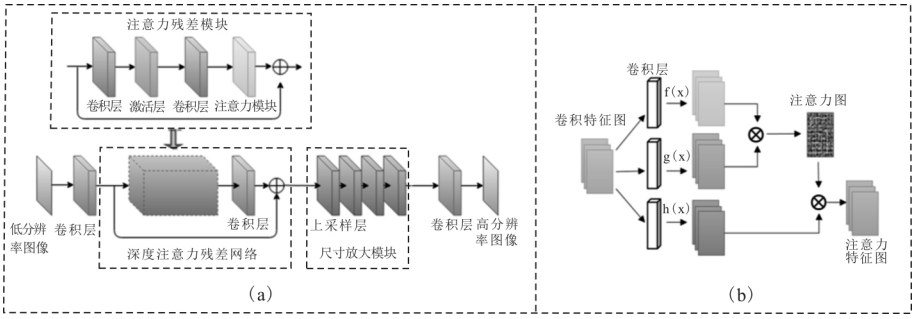
其中,G_{REC} 表示图像重建层。

1.2　判别网络

为了对真实的 HR 图像与生成器生成的 SR 图像进行区分,这里设计了一个判别网络(图 2)。该判别网络是由 8 个卷积层所构成,并且每两层卷积层其卷积核由 64×64 递增至 512×512。由式(5)可以知道,判别网络的主要思想就是使生成器生成的超分辨图像 $G_{\theta_{\mathrm{G}}}(\mathrm{I}_{\mathrm{LR}})$ 能够骗过判别器 D,使得判别器无法判断出该图像是生成器生成的还是原始的高分辨图像。通过生成器与判别器之间的相互博弈,使得的网络最终能重

建出与 HR 图像高度相似且难以被判别器区分的超分辨率图像。

$$\min_{\theta_G}\max_{\theta_D\in L} E_{I_{HR}\sim P_{train}(I_{HR})}[D(I_{HR})]-E_{I_{LR}\sim G(I_{LR})}[D(G_{\theta_G}(I_{LR}))] \quad 。 \tag{5}$$



(a) 生成网络结构图　　(b)注意力机制模块结构图

图 1　人脸超分辨重建网络结构图

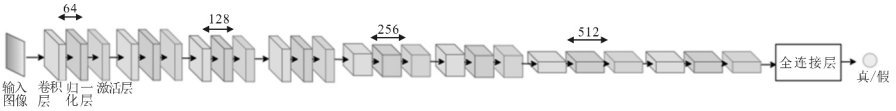


图 2　人脸超分辨重建网络结构图

1.3　损失函数

文中设计的模型所使用的损失函数 l_{SR} 是由感知损失函数 l_{SR}^{VGG}（VGG 损失函数），对抗损失函数 l_{SR}^{adv} 以及正则化损失函数 l_{SR}^{r} 组成，其表达式为

$$l_{SR}=\alpha l_{SR}^{VGG}+\beta l_{SR}^{adv}+(1-\alpha-\beta)I_{SR}^{r} \quad , \tag{6}$$

其中，$\alpha>0$，$\beta>0$，$\alpha+\beta<1$，均为衡量每个损失函数在损失函数 l_{SR} 中所占比例的参数。

1.3.1　感知损失函数（VGG 损失函数）

为衡量网络重建出的超分辨率图像与目标高分辨率图像之间的相似程度，通常是将相应模型产生的 SR 图像与目标 HR 图像以像素为单位计算其均方误差损失函数，该评价方式在一定程度上削弱了模型的泛化能力，使其仅局限于像素级信息的重建。而感知损失，则是利用已训练好的网络来计算网络生成的 SR 图像与目标 HR 图像的相应特征值，通过特征值来计算相应的损失函数，这样便可以使网络学习到鲁棒性能更加的效果。

文中是将生成器生成的 SR 图像与目标 HR 图像放入训练好的 VGG-19 网络中，通过计算 SR 图像特征图与 HR 图像特征图之间的欧氏距离来得到其 VGG 损失的。

$$l_{SR}^{VGG/i,j}=\frac{1}{W_{i,j}H_{i,j}}\sum_{x=1}^{W_{i,j}}\sum_{y=1}^{H_{i,j}}(\phi_{i,j}(I_{HR})_{x,y}-\phi_{i,j}(G_{\theta_G}(I_{LR})_{x,y})^2 \quad , \tag{7}$$

其中，$\phi_{i,j}$ 是第 i 个池化层之前的第 j 个卷积所获得的特征映射，$W_{i,j}$ 和 $H_{i,j}$ 是 ϕ 的维度。

1.3.2　对抗损失函数

对抗损失函数用于评价生成的 SR 图像与原始 HR 图像的相似性，对抗损失 l_{SR}^{adv} 越小，说明生成的 SR 图像越接近真实 HR 图像，生成网络的性能越好，文中模型训练过程中的对抗损失函数[9]如下所示：

$$l_{SR}^{adv}=\sum_{n=1}^{N}-\log D_{\theta_D}(G_{\theta_D}(I_{LR})), \tag{8}$$

其中，$D_{\theta_D}(G_{\theta_G}(I_{LR}))$ 是 $G_{\theta_G}(I_{LR})$ 真实 HR 图像的概率。

1.3.3　正则化损失函数

正则化损失用于提供像素空间上的正则化，用来保证生成的 SR 图像与真实的 HR 图像在内容上不会有很大的偏差，文中模型使用的正则化损失基于 L_1 损失函数，被定义为

$$l_{SR}^{r}=\frac{1}{WH}\sum_{x=1}^{W}\sum_{y=1}^{H}|(I_{HR})_{x,y}-(G_{\theta_G}(I_{LR})_{x,y})。 \tag{9}$$

2　注意力机制

视觉注意力机制源于对人类视觉的研究,它可以通过快速地对全局图像进行扫描来挑选出感兴趣的区域,然后向该区域投入更多的关注度。从本质上来讲,深度学习中的注意力机制和人类的选择性视觉注意力机制十分相似,其主要目标也是从冗杂的信息中选择出对当前目标更关键的信息[10]。近几年,注意力机制被广泛应用于图像分割[11,12],图像定位与理解[13]以及基于循环卷积网络的语义分割、语义理解[14]等工作中。

笔者将注意力机制[5]与生成网络中的残差块结合,使得生成器能够更好地重建出 HR 图像的细节信息。如图 1(b) 所示,从前面卷积层(图 1(a)中注意力残差模块中的第二层卷积层)提取的特征图 x 分别经过两个核为 1 的卷积层被变换到了两个特征空间 $f(x)$和 $g(x)$,其中,$f(x)=W_f x$ 用于提取像素特征,$g(x)=W_g x$ 用于提取全局特征。接着,通过对 $f(x)$和 $g(x)$进行变换来计算注意力图,如下式所示:

$$\beta_{ij}=\frac{\exp(S_{ij})}{\sum_{i=1}^{N}\exp(S_{ij})}\quad , \tag{10}$$

$$O_j=\sum_{i=1}^{N}\beta_{ij}h(x_i)\quad , \tag{11}$$

$$h(x_i)=W_h x_i\quad , \tag{12}$$

$$y_i=\gamma O_i+x_i\quad , \tag{13}$$

其中,β_{ij} 表示在第 i 个区域对第 j 个位置的关注度,且 $S_{ij}=f(x_i)^{\mathrm{T}}g(x_j)$;$O_j$ 表示注意力层的输出;y_i 表示最终的输出,输出的注意力特征图会进入下一个注意力机制网络中继续特征提取与学习的过程;γ 的初始值为 0,因为该开始训练时注意力模块的效果很差,随着训练的进行会增加权重。

注意力机制模块通常放于网络的中上层[5],会使其表现效果更好。因为高层网络接收到的信息多,特征图会更大,选择的自由度也会更大,从而使得生成器和判别器能够建立更加稳定的长期关系。

3　实验结果

3.1　数据准备及参数设置

将 Helen[15]和 celebA[16]数据集中的人脸图像作为训练集和测试集来对提出的网络进行训练。由于文中的网络中有大量参数需要训练,所以需要使用数据增强的方法来增加数据量。这里主要对数据集中的图像进行旋转操作,经数据增强后,数据集中总共有 4 200 张人脸图像,其中 4 000 张图像作为训练集,200 张作为测试集。数据集中的图像用 MATLAB 中的 imresize 函数分别以不同的放大因子(×2,×3,×4)进行下采样。

图像感受野的大小对图像的重建效果有很大的影响,感受野越大,图像所获得的信息就越多。故文中每个图像块的大小都设置为 41×41,训练每一批图片的数量均为 64,使用的优化器是 Adam,其动量和权重衰减分别为 0.9 和 0.000 1。初始学习率为 0.000 1 ,训练的总迭代次数为 200,且每迭代 100 次学习率乘以 0.1。所提出的模型在配置为一块 Intel Xeon E5-2620 CPU 和 1 块 NVIDIA GeForce GTX 1080Ti GPU 的台式电脑上总共训练 16 个小时。

3.2　实验结果与分析

对测试集中降采样后的低分辨率人脸图像进行了不同尺度下的重建,对重建结果采用图像峰值信噪比(Peak Signal-to-Noise Ratio, PSNR)和图像结构相似度(Structural SIMilarity, SSIM)进行了量化评估,并与双三次线性插值(bicubic)方法,SRGAN 方法以及 WGAN 的重建效果进行了定量的对比,其结果如表 1 所示。

表 1 不同方法重建效果的比较

放大因子	Bicubic PSNR(dB)/SSIM	SRGAN[2] PSNR(dB)/SSIM	WGAN[3] PSNR(dB)/SSIM	文中方法 PSNR(dB)/SSIM
2	28.98/0.856 4	30.53/0.893 5	31.23/0.901 5	32.96/0.932 2
3	26.57/0.798 9	28.57/0.864 5	29.36/0.876 6	31.65/0.913 2
4	24.57/0.749 5	25.89/0.827 6	27.13/0.835 4	29.63/0.897 7

如表所示，随着放大倍数的增大，重建的难度也会随之加大，所以放大倍数越大，所得到的结果也会有一定程度的消减。但是相比于传统的双三次线性插值(bicubic)，SRGAN 及 WGAN 方法，文中的方法在 PSNR 值与 SSIM 值上均有一定程度的提升。在放大因子为 4 的情况下，该方法相较于 SRGAN 方法，PSNR 值提升了 3.74dB，SSIM 值提升了 0.070 1；相较于 WGAN 方法，PSNR 值和 SSIM 值分别提升了 2.5dB 和 0.062 3，这说明了笔者所提出的基于生成对抗网络和注意力机制的方法在人脸超分辨重建中的有效性。

接下来，分别使用文中所提出的方法、双三次线性插值、SRGAN 以及 WGAN 方法对测试集的图像进行超分辨重建，如图 3 和图 4 所示，分别为使用这 4 种方法进行 3 倍和 4 倍超分辨率重建后的图像视觉效果对比。由于人眼是人脸图像中信号能量较大的区域，也是人脸中最重要的特征，故图 3 和图 4 中也对每种方法重建后的人脸图像中的人眼部位进行了局部放大，以更好地进行比较。从图中可以看出，文中的方法重建后的图像与 bicubic 放大后的图像相比，人脸轮廓更加清晰。由于文中引入了注意力机制，故相较于 SRGAN 和 WGAN 方法，在眼睛、眉毛等部位的纹理细节信息更加丰富，更符合真实图像。

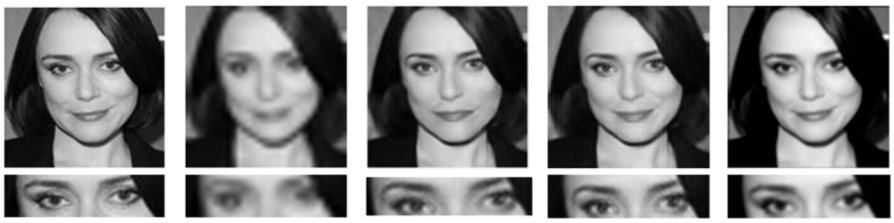


图 3 在放大 3 倍后的人脸图像超分辨重建效果对比

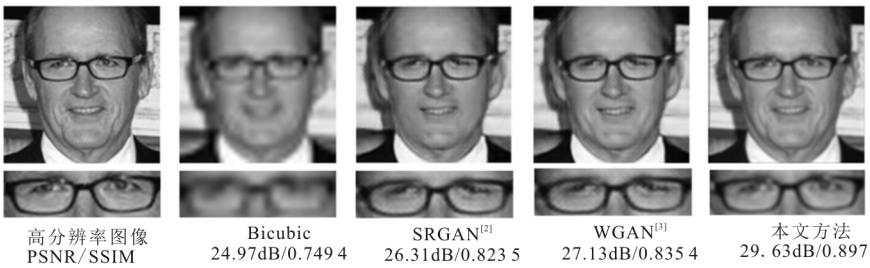


图 4 在放大 4 倍后的人脸图像超分辨率重建效果对比

4 结束语

笔者提出了一种基于生成对抗网络和注意力机制的方法来对人脸图像进行超分辨重建。结合了注意力模块的深度残差网络作为生成器与基于 VGG 网络的判别器相互作用，相互制约，从而重建出与目标高分辨图像高度相似的超分辨人脸图像。将 bicubic 方法、SRGAN 方法以及 WGAN 方法的重建效果与文中方法的重建效果进行了定性定量的对比，实验结果表明，经文中算法重建后图像的 PSNR 值与 SSIM 值都有显著的提升，证实了该方法在人脸图像分辨率提升中的有效性。

参考文献:

[1] GOODFELLOW I J, POUGET-ABADIE J, MIRZA M, et al. Generative Adversarial Nets[C]//Advances in Neural Information Processing Systems. Vancouver, Canada: Neural Information Processing Systems Foundation, 2014: 2672-2680.

[2] LEDIG C, THEIS L, HUSZAR F, et al. Photo-realistic Single Image Super-resolution Using a Generative Adversarial Network[C]// Proceedings of the 2017 30th IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 105-114.

[3] CHEN Z, TONG Y. Face Super-resolution through Wasserstein GANs[CA/OL]//[2018-10-20]. https://arxiv.org/pdf/1705.02438v1.pdf.

[4] **孙毅堂,宋慧慧,张开华,等. 基于极深卷积神经网络的人脸超分辨率重建算法**[J]. **计算机应用**, 2018, 38(4): 1141-1145.
SUN Yitang, SONG Huihui, ZHANG Kaihua, et al. Face Super-resolution via Deep Convolutional Neural Networks[J]. Journal of Computer Application, 2018, 38(4): 1141-1145.

[5] ZHANG H, GOODFELLOW I, METAXAS D, et al. Self-attention Generative Adversarial Networks[CA/OL]//[2018-10-20]. https://arxiv.org/pdf/1805.08318.pdf.

[6] WANG F, JIANG M Q, QIAN C, et al. Residual Attention Network for Image Classification[C]//Proceedings of the 2017 30th IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 6450-6458.

[7] ZHANG Y, LI K P, LI K, et al. Image Super-resolution Using Very Deep Residual Channel Attention Networks[C]// Lecture Notes in Computer Science: 11211. Heidelberg: Springer Verlag, 2018: 294-310.

[8] LIM B, SON S, KIM H, et al. Enhanced Deep Residual Networks for Single Image Super-resolution[C] // Proceedings of the 2017 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. Washington: IEEE Computer Society, 2017: 1132-1140.

[9] ARJOVSKY M, CHINTALA S, BOTTOU L. Wasserstein GAN[CA/OL]//[2018-10-20]. https://arxiv.org/pdf/1701.07875.pdf.

[10] JETLEY S, LORD N A, LEE N, et al. Learn to Pay Attention[CA/OL]//[2018-10-20]. https://arxiv.org/pdf/1804.02391.pdf.

[11] SIMONYAN K, VRDALDI A, ZISSERMAN A. Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps[CA/OL]//[2018-10-20]. https://arxiv.org/pdf/1312.6034.pdf.

[12] CAO C, LIU X, YANG Y, et al. Look and Think Twice: Capturing Top-down Visual Attention with Feedback Convolutional Neural Networks[C]//Proceedings of the 2015 IEEE International Conference on Computer Vision. Piscataway: IEEE, 2015: 2956-2964.

[13] XU K, BA J L, KIROS R, et al. Show, Attend and Tell: Neural Image Caption Generation with Visual Attention[C]// Proceedings of the 32nd International Conference on Machine Learning. Lille: International Machine Learning Society, 2015: 2048-2057.

[14] BAHDANAU D, CHO K, BENGIO Y. Neural Machine Translation by Jointly Learning to Align and Translate[CA/OL]//[2018-10-20]. https://arxiv.org/pdf/1409.0473.pdf.

[15] LE V, BRANDT J, LIN Z, et al. Interactive Facial Feature Localization[C]//Lecture Notes in Computer Science: 7574. Heidelberg: Springer Verlag, 2012: 679-692.

[16] LIU Z, LUO P, WANG X, et al. Deep Learning Face Attributes in the Wild[C]//Proceedings of the 2015 IEEE International Conference on Computer Vision. Piscataway: IEEE, 2015: 3730-3738.

(编辑: 王　瑞)