DOI:10.11918/j.issn.0367-6234.201812115

改进残差块和对抗损失的GAN图像超分辨率重建

张杨忆,林 泓,管钰华,刘 春

(武汉理工大学 计算机科学与技术学院,武汉 430063)

摘 要: 图像超分辨率(Super Resolution,SR)重建是计算机视觉领域中提高图像和视频分辨率的一种重要图像处理技术,针对基于深度学习的图像重建模型层次过多以及梯度传输困难导致训练时间长、重建图像视觉效果不理想的问题,本文提出了一种改进残差块和对抗损失的GAN(Generative Adversarial Networks)图像超分辨率重建模型.首先,在模型结构上,设计剔除多余批规范化操作的残差块并组合成生成模型,将深度卷积网络作为判别模型把控重建图像的训练方向,以减少模型的计算量;然后,在损失函数中,引入Earth-Mover距离设计对抗损失以缓解模型梯度消失的问题,采用L1距离作为重建图像与高分辨率图像相似程度的度量以指导模型权重更新来提高重建视觉效果.在DIV2K、Set5、Set14数据集上的实验结果表明:该模型剔除多余批规范化后的训练时间相比改进前模型减少约14%并有效提高图像的重建效果,结合Earth-Mover距离与L1距离的损失函数有效地缓解了梯度消失的问题.模型相较于双三次插值、SRCNN、VDSR、DSRN模型,提高了对低分辨率图像的超分辨率重建效率和视觉效果.

关键词: 超分辨率重建;生成对抗网络;深度学习;Earth-Mover距离;对抗损失

中图分类号: TP391 **文献标志码:** A **文章编号:** 0367-6234(2019)11-0128-10

GAN image super-resolution reconstruction model with improved residual block and adversarial loss

ZHANG Yangyi, LIN Hong, GUAN Yuhua, LIU Chun

(College of Computer Science and Technology, Wuhan University of Technology, Wuhan 430063, China)

Abstract: Image super-resolution (SR) reconstruction is an important image processing technology to improve the resolution of image and video in computer vision. Image reconstruction model based on deep learning has not been satisfactory due to the too many layers involved, the excessively long training time resulting from difficult gradient transmission, and the unsatisfactory reconstructed image. This paper proposes a generative adversarial networks (GAN) image SR reconstruction model with improved residual block and adversarial loss. Firstly, on the model structure, the residual blocks of the excess batch normalization were designed and combined into a generative model, and the deep convolution network was used as the discriminant model to control the training direction of the reconstructed image to reduce the model's calculation amount. Then, in the loss function, the Earth-Mover distance was designed to alleviate model gradient disappearance. The L1 distance was used as the measure of the degree of similarity between the reconstructed image and the high-resolution image to guide the model weight update to improve the reconstructed visual effect. Experimental results from the DIV2K, Set5, and Set14 datasets demonstrate that compared with the model before improvement, the training time of the proposed model was reduced by about 14% and the image reconstruction effect was effectively improved. For the loss function combined with Earth-Mover distance and L1 distance, gradient disappearance was effectively alleviated. Therefore, the proposed model significantly improved the SR reconstruction efficiency and visual effect of low-resolution images compared with Bicubic, SRCNN, VDSR, and DSRN model.

Keywords: super-resolution construction; Generative Adversarial Networks (GAN); deep learning; Earth-Mover distance; adversarial loss

分辨率是图像清晰度的一个重要评价指标,高分辨率图像清晰度好,拥有高像素密度,提供更多细节,使得人们可以迅速、准确地获取到更多信息,在医学、军事、遥感、视频监控等领域都有广泛应用.图像超分辨率重建技术能够将一幅低分辨率(Low Resolution,LR)图像或图像序列恢复出高分辨率(High Resolution,HR)图像,是近年来图像领域一个重要研究方向,一般可分为单帧图像重建和多帧图像重建,经历了插值重建、序列图像重建和基于学习的图像超分辨率重建三个重要发展阶段[1].

收稿日期: 2018-12-20
作者简介: 张杨忆(1995—),女,硕士研究生
通信作者: 刘 春,liuchun@whut.edu.cn

传统的基于学习的超分辨率重建模型如 Steerable filters[2]、Image Analogies[3] 等需要从若干成对出现的低、高分辨率图像中学习先验信息，对训练样本库依赖性强，训练模型形式单一且受人为干扰较大，近年来，基于深度学习的方法[4]取得了突破性进展，主要思路是充分利用图像本身的先验信息，将深度神经网络作为学习模型，直接学习从低分辨率图像到高分辨率图像的端到端的映射关系.

2014 年 Dong 等[5]提出的基于卷积神经网络的图像超分辨率重建模型(SRCNN)在 3 个卷积层上进行非线性映射得到高分辨率图像，然而由于输入是由双三次插值[6]得到的较高分辨率图像，卷积计算复杂，重建效率低下. 针对该问题，Shi 等[7]提出的 ESPCN 模型采用在低分辨率图像上直接计算卷积进行特征提取后使用子像素卷积层进行图像大小变换的方式，有效提高了图像重建效率，然而一旦模型设置较大的放大倍数时，会导致得到的结果过于平滑，细节缺乏真实感，不能很好提高重建图像的视觉效果.

另外，Kim 等[8]根据低分辨率图像和高分辨率图像拥有相似的低频信息这一理论，提出 VDSR 模型，利用残差学习来加快模型收敛速度，但仍然存在收敛和梯度传输困难的问题. 于是在残差网络和 VDSR 模型的启发下，Tai 等[9]提出 DRRN 模型，将深度增加至 52 层，并递归使用权重共享的网络模块，得到的训练结果相对于 VDSR 模型有提高但并不明显.

针对上述模型层次过多以及梯度传输困难导致训练时间长、重建图像视觉效果不理想的问题，在单帧低分辨率图像重建上本文提出一种改进残差块和对抗损失的 GAN[10] 图像超分辨率重建模型，旨在实现高效、高质量的图像重建，主要贡献如下：

1)设计剔除多余批规范化操作的残差块. 生成模型由剔除块中冗余批规范化操作的残差块组合而成，同时使用深度卷积网络作为判别模型来指导整个模型的训练，减少模型计算量，加快训练速度，提高图像重建效率.

2)引入 Earth-Mover(EM)距离设计对抗损失函数. 通过计算高、低分辨率图像之间的 EM 距离来反映真实数据集和生成数据集的分布距离远近以提供非零梯度来确保训练继续，有效缓解模型梯度消失问题.

3)引入 L1 损失函数计算图片间的 L1 距离. 使用 L1 损失函数替换普遍使用的 MSE 损失函数，将重建图像与对应高分辨率图像间的 L1 距离作为两张图片相似程度的度量，指导模型权重更新，从而提高重建图像的峰值信噪比，加强重建图像视觉效果.

1　改进残差块和对抗损失的 GAN 图像超分辨率重建模型

1.1　模型整体设计

为在单帧低分辨率图像重建中取得良好的重建效果，在模型结构设计中，借鉴 Goodfellow 提出的一种生成式模型—生成对抗网络 GAN. 该网络的核心思想来源于博弈论的纳什均衡[11-12]，结构如图 1 所示. 网络设定参与游戏的双方分别为一个生成模型和一个判别模型，生成模型的目标是尽量去学习真实的数据分布，而判别模型的目标是尽量正确判断输入数据是真实数据还是来自生成模型；为取得游戏胜利，这两个游戏参与者需要不断优化来各自提高自己的生成能力和判别能力，这个学习优化过程就是寻找二者之间的一个纳什均衡.

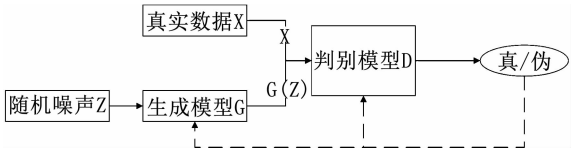


图 1　GAN 结构

Fig. 1　Structure of the GAN

本文通过建立一个基于生成对抗网络的模型进行图像超分辨率重建，模型由生成模型和判别模型两部分组成，分别学习生成与高分辨率图像相似的超分辨率图像和更准确地区分出超分辨率图像与真实高分辨率图像，模型整体设计如图 2 所示，其中，生成模型由剔除多余批规范化操作的残差块与子像素卷积层构成，判别模型由深度卷积网络与全连接层构成.

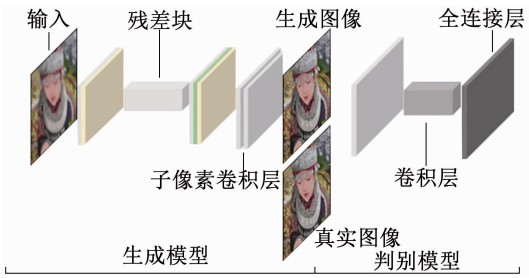


图 2　模型整体设计

Fig. 2　Overall structure of the model

模型训练过程中，低分辨率图像经过生成模型生成对应的超分辨率图像，再将生成的超分辨率图像和真实高分辨率图像一起输入判别模型进行判定，判别模型的输出代表对每张输入图像的真假判定，求取判定结果的误差后对模型参数进行调整，使模型在迭代学习中自动更新内部参数，模型目标函

数定义为

$$\min_{G}\max_{D}V(D,G)=E_{x\sim P_{data(x)}}[\log D(x)]+E_{z\sim P_{z(z)}}[\log(1-D(G(z)))]. \quad (1)$$

式中:x 为真实样本,$D(x)$ 为 x 通过判别模型被判断为真实样本的概率;z 为输入生成样本的噪声,$G(z)$ 为生成网络由噪声 z 生成的样本,而 $D(G(z))$ 为生成样本通过判别模型后,被判断为真实样本的概率. 生成模型的目标是要让生成样本越接近真实样本越好,即 $D(G(z))$ 越接近 1 越好,这时 $V(D,G)$ 会变小;判别模型的目标是要让 $D(x)$ 接近 1,而让 $D(G(z))$ 接近 0,这时 $V(D,G)$ 会增大. 当 $D(x)$ 收敛于 1/2 时,即判别模型无法区分生成样本和真实样本,训练终止. 整个训练过程中样本的分布如图 3 所示,其中,虚线为真实样本分布,实线 a 为判别模型的样本分布,实线 b 为生成模型的样本分布. 在训练过程中,生成模型的样本分布(实线 b)逐渐与真实样本分布(虚线)重合,判别模型的样本分布(实线 a)逐渐变为一条直线即无法正确区分真实样本与生成模型的样本从而完成模型的训练.

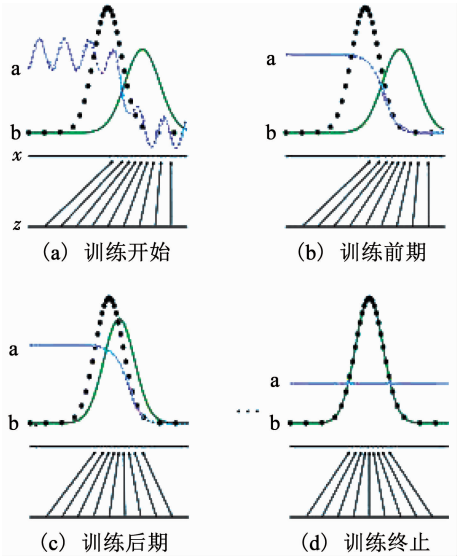


图 3 生成模型、判别模型样本示意

Fig. 3 Schematic diagram of generative model and discriminant model

1.2 改进残差块的生成模型

神经网络使用梯度下降法来计算网络损失,通过反向传播法对网络中所有权重计算损失函数的梯度,将其反馈给梯度下降法来更新神经网络中每个神经元的权值以最小化损失函数. 深度越深的神经网络能够表达越丰富的特征,拥有越出色的性能,然而梯度本身是一个很小的值,随着深度的增加经过多次连乘变得越来越小. 也就是说,网络深度越深,梯度越小,最后就会出现梯度消失的现象,导致深层神经网络无法训练. 何恺明等[13]提出的深度残差网络,通过残差学习有效简化模型训练、缓解梯度弥散现象. 同时,Kim 等[8]又提出低分辨率图像和高分辨率图像共享低频部分的信息,因此,只需要学习高分辨率图像和低分辨率图像之间的高频部分的残差就可以实现超分辨率重建.

为能够充分学习低分辨率图像到高分辨率图像端到端的映射关系,本文生成模型引入深度残差网络. 深度残差网络的残差块由卷积层、批规范化层[14](Batch-normalization, BN)、PReLU 激活函数[15]组成,如图 4(a)所示. 本文设计并使用仅由 2 个卷积层和 PReLU 激活函数构成的残差块,如图 4(b)所示. 在深度学习中,批规范化操作在一定程度上解决深度神经网络在训练中由于梯度随机下降而带来的梯度弥散现象,提高了模型的训练速度和准确度. 但在图像超分辨率重建问题上,图像经过批规范化操作会导致色彩分布被归一化、原始图像的空间表征被破坏,进而在重建中需要模型的部分层或者额外参数来恢复这种表征. 因此,本文将 Lim 等[16]剔除残差网络中批规范化的思想引入到本文设计的基于生成对抗网络的图像超分辨率重建中,在生成模型中剔除批规范化操作来减少模型计算量、防止提取到的图像特征被归一化,从而提高模型精度和重建效果.

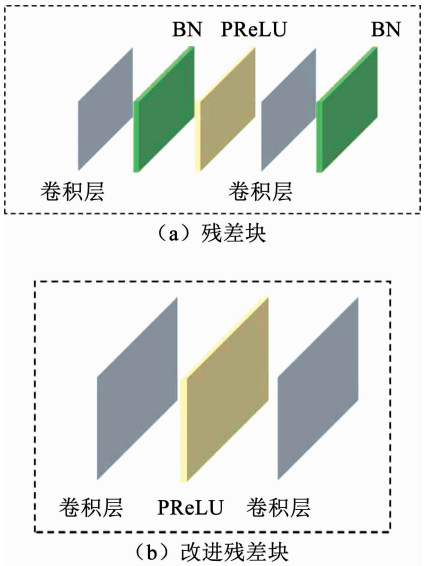


图 4 改进前后的残差块

Fig. 4 Residual blocks and modified residual blocks

本文生成模型的整体结构如图 5 所示,前一部分为深度残差网络,用于提取原始图像特征,后面连接 2 个子像素卷积层[7],用于放大特征图像. 其中,深度残差网络由 16 个剔除批规范化操作的残差块组合而成,图 5 中“16 个残差块”表示 16 个残差块依次连接,残差块使用图 4(b)所示结构,块内剔除

了批规范化操作即图 4(a)中的绿色块，使用步幅为 1、卷积核大小为 3 × 3 的卷积层来提取足够数量的特征图，块与块之间使用跳跃连接以促进梯度反向传播、加快模型训练速度；子像素卷积层输入通道为 1、大小为 $W \times H \times 1$ 的特征图，获得具有 c^2 个通道、大小为 $W \times H \times c^2$ 的特征图，再将特征图每个像素的 c^2 个通道重排成一个 $c \times c$ 的区域，对应于高分辨率图像中一个 $c \times c$ 的子区域，从而获得一个大小为 $cW \times cH \times 1$ 的高分辨率图像. 其中，c 是一个常数，表示一幅特征图经过一个子像素卷积层之后放大 c 倍. 在本文中，令 $c = 2$，即图像经过两个子像素卷积层后放大 4 倍. 生成模型剩余参数设置如表 1 所示，Conv_G1 ~ Conv_G5 为生成模型内卷积层，其中，Conv_G2 为残差块中的卷积层.

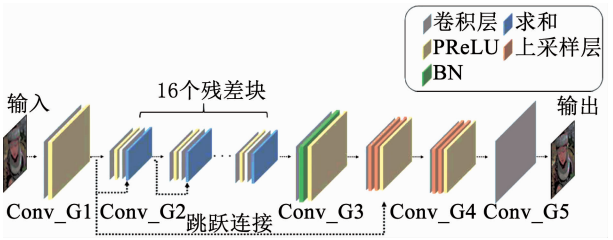


图 5　生成模型结构

Fig. 5　Structure of the generative model

表 1　生成模型卷积层参数设置

Tab. 1　Parameter setting of convolution neural network in the generative model

卷积层	卷积核大小	特征维度	步幅
Conv_G1	3 × 3	64	1
Conv_G2	3 × 3	64	1
Conv_G3	3 × 3	64	1
Conv_G4	3 × 3	256	1
Conv_G5	1 × 1	3	1

1.3　判别模型

为使生成模型生成的超分辨率图像更加接近于真实的高分辨率图像，本文训练了一个与之对应的判别模型.

本文的判别模型使用深度卷积网络，在搭建中参考 Radford 等[17]提出的判别模型结构. 模型内部共包括 11 个卷积层，所有卷积层均使用 Leaky ReLU[18]作为激活函数($\alpha = 0.2$)，从而避免在整个网络结构中使用 Max Pooling 操作，以简化计算过程. 除第一个卷积层，在每个卷积层之后使用批规范化操作以避免训练时梯度消失，加快模型的收敛速度，增加模型稳定性. 最后添加全连接层，将卷积层提取到的图像特征进行分类组合后输出. 由于原始生成对抗网络的判别模型用于判断输入是真实数据还是生成数据，输出结果映射在 0、1 之间，解决的是二分类问题，而本文模型引入 EM 距离，去掉模型最后的 sigmoid 层，连续的输出结果使判别模型转变为解决回归问题. 判别模型整体结构如图 6 所示，参数设置如表 2 所示，Conv_D1 ~ Conv_D11 为判别模型内卷积层.

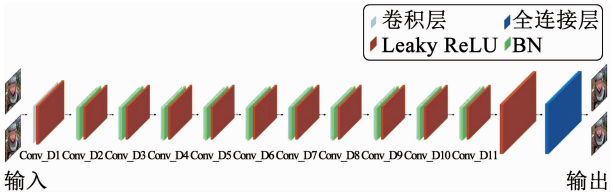


图 6　判别模型结构

Fig. 6　Structure of the discriminant model

表 2　判别模型卷积层参数设置

Tab. 2　Parameter setting of convolution neural network in the discriminant model

卷积层	卷积核大小	特征维度	步幅
Conv_D1	4 × 4	64	2
Conv_D2	4 × 4	128	2
Conv_D3	4 × 4	256	2
Conv_D4	4 × 4	512	2
Conv_D5	4 × 4	1 024	2
Conv_D6	4 × 4	2 048	2
Conv_D7	1 × 1	1 024	1
Conv_D8	1 × 1	512	1
Conv_D9	1 × 1	128	1
Conv_D10	3 × 3	128	1
Conv_D11	3 × 3	512	1

2　损失函数

定义高分辨率图像为 I_H，其对应的低分辨率图像为 I_L，通过模型重建得到的高分辨率图像被称为超分辨率图像 I_S. 定义低分辨率图像的宽度为 W，高度为 H，图像的颜色通道数目为 C，定义缩放倍数为 r，则 I_L 的大小为 $W \times H \times C$，I_H 和 I_S 的大小为 $rW \times rH \times C$.

本文模型的最终目标是训练一个生成函数 G，输入低分辨率图像 I_L，输出预测的超分辨率图像 I_S，即 $G(I_L) = I_S$. 为实现这个目标，需要训练一个卷积神经网络 G_{θ_G}，并优化损失函数 l_S 得到 $\theta_G = \{W_{1:L}; b_{1:L}\}$，其中 $W_{1:L}$ 为网络第 L 层的权重、$b_{1:L}$ 为偏差，进而可以通过给定高分辨率图像训练集 I_{H_n} ($n = 1, \cdots, N$) 和对应的低分辨率图像 I_{L_n} 优化式(2)

得到 θ_G：

$$\hat{\theta}_G = \arg\min_{\theta_G} \frac{1}{N}\sum_{n=1}^{N} l_S(G_{\theta_G}(I_{L_n}), I_{H_n}). \quad (2)$$

本文设计的损失函数根据感知相关特征来评估解决方案，由内容损失(l_X)和对抗损失(l_{Gen})两部分组成：

$$l_G = l_X + 10^{-3} l_{Gen}.$$

2.1 EM 距离下的对抗损失

在确定模型的对抗损失时，首先引入 Goodfellow 等[10]提出的生成模型损失函数：

$$E_{x\sim P_{data}}[\log(1 - D(x))]. \quad (3)$$

由最原始生成对抗网络的判别模型损失函数可以推出最优判别模型：

$$D^*(x) = \frac{P_{data}(x)}{P_{data}(x) + P_z(x)}. \quad (4)$$

式中：$P_{data}(x)$为实数据域的样本，$P_z(x)$为噪声 z 生成的样本. 式(3)加上一个不依赖于生成模型的项得到新的损失函数 $E_{x\sim P_{data}}[\log D(x)] + E_{x\sim P_z}[\log(1 - D(x))]$，最小化这个损失函数等价于最小化式(3)，将其带入式(4)，化简得到 JS 散度表示的损失函数为

$$2J_{JS}(P_{data} \| P_z) - 2\log 2.$$

然而，在训练判别模型的过程中，随着判别模型越来越接近最优效果时，由式(1)可知，最小化生成模型损失也会越来越接近于最小化真实数据域和生成数据域间的 JS 散度，但这建立在两个数据域有重叠的基础上，一旦没有重叠，JS 散度就会固定为常数 log2，梯度下降法中梯度变为 0，造成生成梯度消失. 与此同时，在最小化生成模型损失的过程中，模型在最小化生成分布与真实分布的 KL 散度的同时最大化两者的 JS 散度，这是一个矛盾的任务，并且 KL 散度是一个不对称的衡量，对生成样本多样性和准确性的衡量标准不一致会导致模式坍塌. 因此，本文引入 EM 距离来替代 JS 散度和 KL 散度. 相比于 JS 散度和 KL 散度，EM 距离在两个分布没有重叠或重叠部分可忽略时，仍能反映它们的远近，提供非零梯度，并且距离是平滑的、不会突变.

EM 距离的定义为

$$W(P_{data}, P_z) = \inf_{\gamma\sim\Pi(P_{data}, P_z)} E_{(x,y)\sim\gamma}[\|x - y\|]. \quad (5)$$

式中：$\Pi(P_{data}, P_z)$为 P_{data}和 P_z 组合起来所有可能的联合分布集合. 由于定义中 $\inf_{\gamma\sim\Pi(P_{data}, P_z)}$不能直接求解，根据 Kantorovich-Rubinstein 对偶原理将式(5)变换为

$$W(P_{data}, P_z) = \frac{1}{K}\sup_{\|f\|_L\leqslant K} E_{x\sim P_{data}}[f(x)] - E_{x\sim P_z}[f(x)]. \quad (6)$$

式中：K 为函数 f 的 Lipschitz 常数，使得定义域内任意两个元素 x_1 和 x_2 都满足 $|f(x_1) - f(x_2)| \leqslant K|x_1 - x_2|$. 特别地，可以用一组参数 ω 来定义一系列可能的函数 $f_\omega(x)$，此时求解式(6)可近似变为求解式(7)：

$$K \cdot W(P_{data}, P_z) \approx \max_{\omega:\ |f_\omega|_L\leqslant K} E_{x\sim P_{data}}[f_\omega(x)] - E_{x\sim P_z}[f_\omega(x)]. \quad (7)$$

因此，构造一个含参数 ω 且最后一层不是非线性激活层的判别模型 $D(x)$，使得 $L = E_{x\sim P_{data}}[D(x)] - E_{x\sim P_z}[D(x)]$ 尽可能最大，此时 L 就会近似于真实分布与生成分布之间的 EM 距离. 最小化 L 得到对抗损失函数如式(8)所示和判别模型的损失函数如式(9)所示：

$$l_{Gen} = -E_{x\sim P_z}[D(x)], \quad (8)$$

$$l_D = E_{x\sim P_z}[D(x)] - E_{x\sim P_{data}}[D(x)]. \quad (9)$$

2.2 L1 距离下的内容损失

在基于深度神经网络的超分辨率重建模型中，通常使用像素级 MSE 损失函数如式(10)所示，直接优化高、低分辨率图像像素间的平方差，使得生成的图像有较高的峰值信噪比.

$$l_{MSE} = \frac{1}{r^2 WH}\sum_{x=1}^{rW}\sum_{y=1}^{rH}((I_H)_{x,y} - (G_{\theta_G}(I_L))_{x,y})^2. \quad (10)$$

MSE 损失函数的目标形式包含平方项，梯度容易计算，同时具有很好的收敛特性. 然而对于图像超分辨率重建而言，MSE 损失函数具有对异常点进行大量加权的缺点，并由于对每一项都进行了平方，导致对大误差的惩罚多于小误差. 同时，MSE 损失函数基于原始误差度量的方式，没有考虑人眼的视觉特性，并不能提高人眼视觉感受的质量. 而 L1 损失函数属于原始误差度量方式，并不会过度惩罚大误差项，且每个像素的误差对 L1 损失的影响与误差的绝对值成正比.

本文将像素级 L1 损失函数计算得到的高分辨率图像$(I_H)_{x,y}$和重建图像$(G_{\theta_G}(I_L))_{x,y}$之间的 L1 距离作为 2 张图片相似程度的惩罚项，设计内容损失函数为

$$l_X = l_{L1} = \frac{1}{r^2 WH}\sum_{x=1}^{rW}\sum_{y=1}^{rH}|(I_H)_{x,y} - (G_{\theta_G}(I_L))_{x,y}|.$$

实验结果证明，相比于 MSE 损失函数，使用 L1 损失函数可以达到更好的实验效果. 同时，L1 损失函数是随着误差的增大线性增长，可以有效防止梯度爆炸的产生，并指导模型权重更新以使重建得到的超分辨率图像拥有更高的峰值信噪比.

3　实验结果及分析

3.1　实验环境及数据集

本文实验环境采用谷歌云计算平台，内存为 18G，显卡为 NVIDIA Tesla K80，显存为 12GB，平台搭载的操作系统为 Ubuntu16. 04 LTS，使用 Tensorflow-gpu 1.4，TensorLayer 1.8.0，Python 3.6.

实验使用 DIV2K 数据集，该数据集是一种新发布的用于图像复原任务的高质量图像数据集，包含 800 张训练图像，100 张验证图像和 100 张测试图像. 由于测试数据集资料尚未发布，因此在验证数据集上比较模型性能. 另外，增加 Set5，Set14 两个标准基准数据集（分别包含 5 张和 14 张图像）来进一步比较模型性能.

生成模型和判别模型的权重使用均值为 0、方差为 0.02 的高斯分布进行随机初始化. 误差反向传播采用随机梯度下降算法 Adam，beta1 为 0.9，初始学习率为 0.000 1，初始衰减率为 0.1，最小批尺寸为 16，训练过程中交替更新生成模型和判别模型.

3.2　评价指标

本文使用峰值信噪比（peak signal to noise ratio，PSNR）、结构相似性（structural similarity index，SSIM）、主观质量评分（mean opinion score，MOS）作为图像质量评估指标.

3.2.1　峰值信噪比

峰值信噪比（PSNR），单位 dB，是最普遍、最广泛使用的评估压缩图像质量指标，一般通过均方误差（Mean-square Error，MSE）进行定义. 两幅 $W \times H$ 单色图像 X 和 Y 的均方误差定义为

$$M_{\mathrm{MSE}} = \frac{1}{WH}\sum_{i=0}^{W-1}\sum_{j=0}^{H-1}[X(i,j) - Y(i,j)]^2.$$

对于 RGB 彩色图像，每个像素点有 RGB 3 个值，其均方误差的定义为所有值的方差总和除以图像大小再除以 3，如式（11）所示：

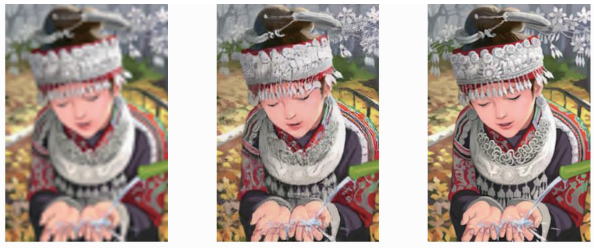
$$M_{\mathrm{MSE}} = \frac{1}{3WH}\sum_{i=0}^{W-1}\sum_{j=0}^{H-1}[X(i,j)_{\mathrm{R,G,B}} - Y(i,j)_{\mathrm{R,G,B}}]^2. \tag{11}$$

因此，峰值信噪比定义为

$$P_{\mathrm{PSNR}} = 10\log\left(\frac{X_{\mathrm{MAX}}^2}{M_{\mathrm{MSE}}}\right) = 20\log\left(\frac{X_{\mathrm{MAX}}}{\sqrt{M_{\mathrm{MSE}}}}\right).$$

式中：X_{MAX} 为图像中可能的最大的像素值. 如果每个采样点都有 B 位线性脉冲编码调制表示，那么 X_{MAX} 就是 $2^B - 1$，即如果每个采样点有 8 位表示时，$X_{\mathrm{MAX}} = 255$.

目前，峰值信噪比是超分辨率重建质量评估最常用的指标，峰值信噪比越高，代表图像重建质量越好. 然而峰值信噪比越高，并不代表图像的视觉效果越好. 如图 7 所示，图 7（a）图的 PSNR 值高于图 7（b）图，但图 7（a）图更加模糊，视觉效果差于图 7（b）图，这表明峰值信噪比在评价超分辨率重建视觉效果上存在一定的局限性.



（a）P_{PSNR}=20.47dB　（b）P_{PSNR}=20.03dB　（c）原图

图 7　PSNR 与图像视觉效果

Fig. 7　elationship between PSNR and image visual effect

3.2.2　结构相似性

结构相似性（SSIM）是测量两幅数字图像之间相似度的指标，能反映人眼的主观感受. 给定两幅图像 x 和 y，通过亮度 $\mathrm{L}(x,y)$、对比度 $\mathrm{C}(x,y)$ 和结构 $\mathrm{S}(x,y)$ 定义结构相似性：

$$\mathrm{L}(x,y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}, \tag{12}$$

$$\mathrm{C}(x,y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}, \tag{13}$$

$$\mathrm{S}(x,y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3}, \tag{14}$$

$$\mathrm{S}_{\mathrm{SSIM}}(x,y) = [\mathrm{L}(x,y)]^{\alpha}[\mathrm{C}(x,y)]^{\beta}[\mathrm{S}(x,y)]^{\gamma}. \tag{15}$$

式中：式（12）中 μ_x 与 μ_y 分别为 x 和 y 的平均值，式（13）、（14）中 σ_x 和 σ_y 分别为 x 和 y 的标准差，σ_{xy} 为 x 和 y 的共变异数字，C_1、C_2、C_3 皆为常数. 式（15）中 $\alpha > 0$，$\beta > 0$，$\gamma > 0$. 在实际应用中，通常 $\alpha = \beta = \gamma = 1$ 及 $C_3 = C_2/2$.

与峰值信噪比相比，结构相似性更符合对图像视觉效果的评估，但仍在反映图像的视觉效果上同样存在局限性.

3.2.3　主观质量评分

主观质量评分（MOS）是图像质量最具代表性的主观评价方法，可以直接反映人眼对图像质量的评价. 其中，主观质量评分使用绝对分类评分标准，将差评和好评分映射到 1 至 5 之间的数字，如表 3 所示.

在主观质量评估测试中，最终评分被计算为 N 位受试者针对给定图像进行的单一评分的算术平均值为

$$M_{\mathrm{MOS}} = \frac{\sum_{i=1}^{N} R_i}{N}.$$

式中:R_i 为单人评分. 本文的主观质量评分是基于随机抽取的 50 名学生对给定图像的评分计算得到.

表 3 绝对分类评分标准

Tab. 3 Absolute classification criteria

评分	质量评价
5	非常好
4	好
3	一般
2	差
1	非常差

3.3 残差块质量评价

本文在深度残差网络的基础上,剔除了网络残差块中多余的批规范化操作,为证明改进残差块的有效性,相同实验环境下,在 Set5,Set14 两个标准基准数据集上,对原始残差网络作为生成模型的模型、改进残差网络作为生成模型的模型分别进行迭代训练,且两个模型均使用本文设计的损失函数. 原始模型耗时约 28 小时,改进后的模型耗时约 24 小时.

固定模型迭代次数 500、600、800,比较 PSNR、SSIM、MOS 和单次迭代耗时,如表 4 所示. 在不同迭代次数下,本文模型相比于改进前模型,PSNR 在 Set5 数据集上平均约提高 1.15 dB,在 Set14 数据集上平均约提高 0.16 dB;SSIM 在 Set5 数据集上平均约提高 0.030,在 Set14 数据集上平均约提高 0.017;MOS 在 Set5 数据集上平均约提高 0.19,在 Set14 数据集上平均约提高 0.18. 改进后的残差块,剔除块中冗余批规范化操作,减少网络计算量,使得迭代耗时由原来的 200 秒变为 175 秒,减少 25 秒,证明改进残差块对加快模型训练速度、提高重建效率的有效性.

表 4 残差块实验对比

Tab. 4 Results of residual blocks experiment

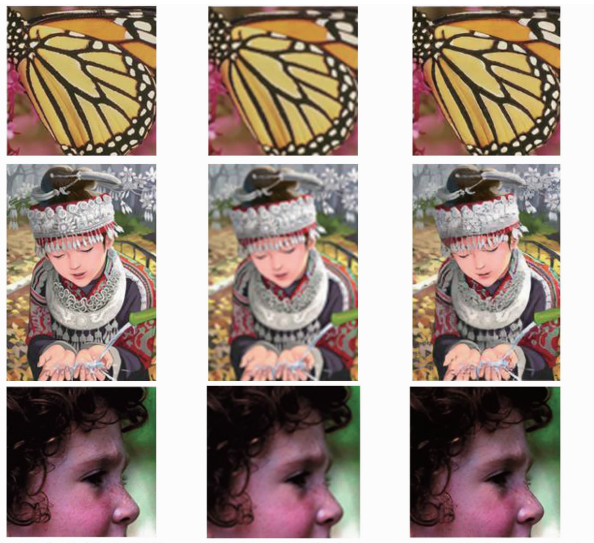
迭代次数	数据集	残差块改进前的模型				本文模型			
		P_{PSNR}/dB	SSIM	MOS	单次迭代耗时/s	P_{PSNR}/dB	SSIM	MOS	单次迭代耗时/s
500	Set5	28.92	0.832	3.48	200	30.14	0.866	3.68	175
	Set14	25.77	0.722	3.28		25.93	0.741	3.49	
600	Set5	29.03	0.837	3.51		30.19	0.869	3.72	
	Set14	25.81	0.731	3.32		25.98	0.749	3.48	
800	Set5	29.09	0.844	3.49		30.21	0.872	3.67	
	Set14	25.83	0.735	3.29		25.99	0.751	3.47	
1 000	Set5	29.11	0.845	3.49		30.21	0.872	3.67	
	Set14	25.84	0.736	3.29		25.98	0.751	3.46	

表 4 数据显示,当迭代次数达到 500 次以上时,本文模型的评价指标值趋于收敛,增长并不迅速且有降低趋势,因此,本文以下实验均采用模型 500 次迭代时的结果来验证模型有效性.

3.4 损失函数性能对比

本文引入 L1 损失函数替换普遍使用的 MSE 损失函数,并在此基础上引入 EM 距离,为证明改进损失函数的有效性和保证实验的公平性,模型均使用本文改进的残差块来进行实验.

首先,在 Set5,Set14 两个标准基准数据集上,将使用 MSE 损失的和使用 L1 损失及 EM 距离的本文模型在相同实验环境下分别进行 500 次迭代训练,两者均耗时约 24 小时,实验部分结果如图 8 所示. 其中图 8(a)列是高分辨率图像;图 8(b)列是改进前模型的重建图像;图 8(c)列是本文模型的重建图像. 观察图 8 可以发现本文模型重建得到的图片清晰度较高,视觉效果较好,细节较丰富,改进前模型重建得到的图片清晰度较低,图像过于平滑,细节较为粗糙.



(a) 高分辨率图像 (b) MSE损失 (c) L1损失+EM距离

图 8 重建效果

Fig. 8 The reconstructed images

其次，为进一步证明 L1 损失和 EM 距离的有效性，在 Set5 和 Set14 数据集上比较本文模型和改进前模型的 PSNR、SSIM 和 MOS，如表 5 所示. 本文模型相比于改进前的模型，PSNR 在 Set5 数据集上提高约 0.19 dB，在 Set14 数据集上提高约 0.24 dB；SSIM 在 Set5 数据集上提高约 0.012，在 Set14 数据集上提高约 0.017；MOS 在 Set5 数据集上提高约 0.26，在 Set14 数据集上提高约 0.35，实验数据证明引入 L1 损失和 EM 距离对提高重建效果的有效性.

表 5　损失函数实验对比

Tab. 5　Results of loss experiment

数据集	MSE 损失函数			L1 损失函数 + EM 距离		
	P_{PSNR}/dB	SSIM	MOS	P_{PSNR}/dB	SSIM	MOS
Set5	29.95	0.854	3.42	30.14	0.866	3.68
Set14	25.69	0.724	3.14	25.93	0.741	3.49

3.5　图像重建效果评估

选择典型的双三次插值（Bicubic）模型、SRCNN 模型、VDSR 模型以及 DSRN 模型[19]与本文模型在 PSNR、SSIM 和 MOS 3 个图像质量评价指标上进行对比实验. 双三次插值直接根据图像中距离每个像素点最近的 16 个像素点计算得到目标超分辨率图像，而 SRCNN 模型、VDSR 模型和 DSRN 模型是 3 个典型的基于深度学习的重建模型，通过学习低分辨率图像到高分辨率图像的端到端的映射关系来得到目标超分辨图像.

本次实验通过 Python 中的 OpenCV 库对图像进行双三次插值；SRCNN 模型和 DSRN 模型使用作者提供的源代码进行重建；VDSR 模型使用 Tensorflow 框架下的代码进行重建. 实验结果如图 9 所示，分别展示每个模型在 Set5 数据集上重建的超分辨率图像的效果. 其中图 9（a）列是原始高分辨率图像；图 9（b）列是 Bicubic 模型的重建图像；图 9（c）列是 SRCNN 模型的重建图像；图 9（d）列是 VDSR 模型的重建图像；图 9（e）列是 DSRN 模型的重建图像；图 9（f）列是本文模型的重建图像.

图 9　Set5 数据集的重建图像

Fig. 9　Reconstructed images based on Set5 dataset

其中，第 1、4、5 行分别是在 baby、head、woman 人物图像上的重建结果，采用双 3 次插值法重建的图像人物面部特征都比较模糊，SRCNN、VDSR、DSRN 和本文模型重建的图像都比较清晰，其中本

文模型重建的图像细节更丰富、与原图更接近;第 2 行是在 bird 图像上的重建结果,5 种模型重建的图像背景整体效果类似,而采用双 3 次插值法的鸟的头部比较模糊,SRCNN、VDSR 和 DSRN 模型的结果则比较清晰,但都比本文模型重建的结果要模糊;第 3行是在 butterfly 图像上的重建结果,采用双 3 次插值法重建的图像锯齿较多,SRCNN 和 VDSR 模型重建的图像虽然锯齿较少但比较模糊,DSRN 和本文模型重建的图像则比较平滑,都与原图比较接近.

为对比本文模型与其他模型的重建效果,在 Set5 数据集上分别计算每个模型重建得到每张图像的 PSNR、SSIM 和 MOS,如表 6 所示.

为进一步验证本文模型的有效性,在 Set14 数据集上进行实验. Set14 数据集中 14 幅图像的实验结果如表 7 所示. 从表 6 和表 7 中各项数据可知,本文模型重建得到的图像虽然在个别图像上 PSNR 和 SSIM 得分并不是最高,但在 MOS 评分整体上是最优的,再次证明其在提高视觉效果方面的有效性.

表 6 在 Set5 上的实验结果对比

Tab. 6 Test results based on Set5 dataset

重建图像	Bicubic[6]			SRCNN[5]			VDSR[8]			DSRN[19]			本文模型		
	P_{PSNR}/dB	SSIM	MOS	P_{PSNR}/dB	SSIM	MOS	P_{PSNR}/dB	SSIM	MOS	P_{PSNR}/dB	SSIM	MOS	P_{PSNR}/dB	SSIM	MOS
Woman	24.35	0.806	1.52	27.60	0.875	3.14	27.54	0.902	3.28	28.19	**0.904**	3.56	**28.53**	0.900	**3.57**
Butterfly	19.86	0.706	1.42	24.20	0.841	3.42	25.58	0.866	3.51	25.59	0.875	3.67	**25.66**	**0.878**	**3.71**
Bird	27.73	0.84	1.28	31.27	0.903	3.42	32.50	0.917	3.44	**32.52**	**0.918**	**3.46**	31.84	0.917	3.28
Head	29.76	0.721	1.57	31.11	0.763	3.28	**31.89**	0.778	3.40	31.87	0.778	3.57	29.02	**0.789**	**3.85**
Baby	29.47	0.827	2.14	31.78	0.872	3.85	**32.56**	**0.880**	3.92	32.45	0.878	3.96	30.65	0.816	**4.00**
平均值	26.23	0.780	1.60	29.19	0.851	3.42	30.01	0.869	3.51	**30.12**	**0.871**	3.64	29.14	0.856	**3.68**

注:表中加粗数据分别对应 Set5 数据集图像由不同模型重建的结果在 PSNR、SSIM、MOS 指标上的最优值

表 7 在 Set14 上的实验结果对比

Tab. 7 Test results based on Set14 dataset

重建图像	Bicubic[6]			SRCNN[5]			VDSR[8]			DSRN[19]			本文模型		
	P_{PSNR}/dB	SSIM	MOS	P_{PSNR}/dB	SSIM	MOS	P_{PSNR}/dB	SSIM	MOS	P_{PSNR}/dB	SSIM	MOS	P_{PSNR}/dB	SSIM	MOS
Foreman	27.68	0.844	2.57	29.04	0.902	3.00	30.43	**0.908**	3.35	**30.47**	**0.908**	3.37	28.23	0.905	**3.42**
Man	23.85	0.636	1.28	**25.45**	**0.729**	3.00	25.06	0.714	3.16	25.13	0.716	**3.17**	24.51	0.695	3.14
Pepper	28.42	0.806	1.42	**31.28**	0.852	**3.57**	31.01	**0.854**	**3.57**	31.09	0.853	3.56	30.02	0.853	**3.57**
Comic	19.86	0.556	2.28	19.23	0.597	2.57	21.91	0.635	2.98	**21.93**	0.635	3.01	21.87	**0.699**	**3.14**
Coastguard	24.02	0.472	2.71	24.66	0.487	2.71	24.79	0.500	3.21	**24.81**	**0.501**	3.33	24.53	0.475	**3.42**
Baboon	20.9	0.426	2.14	21.4	0.513	3.28	21.48	0.513	3.47	21.52	0.514	3.51	**21.62**	**0.516**	**3.57**
Monarch	25.16	0.859	1.85	28.85	0.915	2.57	28.99	0.927	3.29	29.13	0.928	3.67	**29.38**	**0.933**	**3.85**
Flowers	23.49	0.687	1.57	23.64	0.737	2.85	24.85	0.766	3.00	24.98	0.769	3.09	**26.48**	**0.789**	**3.14**
Face	29.71	0.719	1.14	**30.84**	**0.771**	3.42	29.41	0.766	3.50	29.38	0.758	3.48	29.3	0.748	**3.71**
Ppt3	19.97	0.787	1.14	21.81	0.869	3.45	24.07	0.898	**3.52**	24.12	0.903	3.46	**24.75**	**0.921**	3.42
Zebra	21.87	0.636	1.42	19.07	0.552	2.85	23.75	0.693	3.37	23.98	0.699	3.47	**25.37**	**0.723**	**3.48**
Bridge	22.75	0.537	1.57	23.70	0.637	3.57	23.81	**0.626**	3.60	**23.82**	**0.626**	**3.61**	23.3	0.567	3.57
Barbara	23.45	0.657	2.00	**24.30**	**0.727**	3.57	23.77	0.715	3.69	23.79	0.714	3.75	23.83	0.712	**3.85**
Lenna	27.71	0.779	1.42	29.94	0.834	3.00	**29.82**	0.831	3.13	29.81	0.830	3.50	29.8	**0.835**	**3.57**
平均值	24.20	0.671	1.75	25.23	0.723	3.10	25.94	0.739	3.35	**26.00**	0.740	3.43	25.93	**0.741**	**3.49**

注:表中加粗数据分别对应 Set14 数据集图像由不同模型重建的结果在 PSNR、SSIM、MOS 指标上的最优值

4 结 论

本文针对基于深度学习的图像重建模型层次过多导致训练时间长、重建效率低的问题,设计剔除多余批规范化操作的残差块组合成生成模型,同时将深度卷积网络作为判别模型以指导模型训练. 针对重建模型中梯度传输困难导致训练难以继续的问题,引入 EM 距离设计对抗损失函数,同时使用 L1

损失函数来指导模型权重更新，加强重建图像视觉效果. 在不同数据集上的实验结果表明，本文模型针对单帧图像的重建效率与重建视觉效果相对于优秀的同类方法有一定的提升，但适用于特定的使用双三次插值的降质模型，对于通过不同降质过程获得的低分辨率图像，重建效果并不理想. 最后，本文对未来的工作归纳如下：

1）超分辨率重建模型使用的 L1 损失函数忽略了图像的部分高频特征，导致重建图像的细节与实际不太相符. 如何改进损失函数，从而使重建得到的图像细节更为丰富且与实际更为接近，将有助于提高重建图像的准确率.

2）基于生成对抗网络的超分辨率重建模型的计算量仍然太大，导致模型运行效率低下，无法达到实际应用水平. 如何简化生成模型结构或者使用更高效的计算方式来降低计算量，将有助于提高模型的重建效率.

参考文献

[1] 杨帅锋. 基于学习的超分辨率重建算法研究[D]. 北京：北京交通大学，2015
YANG Shuaifeng. Super-resolution reconstruction algorithms based on learning method[D]. Beijing：Beijing Jiaotong University，2015

[2] FREEMAN W T，ADELSON E H. The design and use of steerable filters[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence，1991，13(9)：891. DOI：10.1109/34.93808

[3] HERTZMANN A，JACOBS C E，OLIVER N，et al. Image analogies [C]// Conference on Computer Graphics & Interactive Techniques. ACM，2001. DOI：10.1145/383259.383295

[4] 孙旭，李晓光，李嘉锋，等. 基于深度学习的图像超分辨率复原研究进展[J]. 自动化学报，2017(5)：697. DOI：10.16383/j.aas.2017.c160629
SUN Xu，LI Xiaoguang，LI Jiafeng，et al. Review on deep learning based image super-resolution restoration algorithms [J]. Acta Automatica Sinica，2017(5)：697. DOI：10.16383/j.aas.2017.c160629

[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，2014，38(2)：295. DOI：10.1109/TPAMI.2015.2439281

[6] KEYS R G. Cubic convolution interpolation for digital image processing[J]. IEEE Transactions on Acoustics Speech & Signal Processing，2003，29(6)：1153. DOI：10.1109/TASSP.1981.1163711

[7] SHI W，CABALLERO J，HUSZAR F，et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network[C]//2016 IEEE Conference on Computer Vision & Pattern Recognition. IEEE Computer Society，2016：1874. DOI：10.1109/CVPR.2016.207

[8] KIM J，LEE J K，LEE K M. Accurate image super-resolution using very deep convolutional networks[C]// 2016 IEEE Conference on Computer Vision & Pattern Recognition. IEEE Computer Society，2016：1646. DOI：10.1109/CVPR.2016.182

[9] TAI Ying，YANG Jian，LIU Xiaoming. Image super-resolution via deep recursive residual network [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society，2017：2790. DOI：10.1109/CVPR.2017.298

[10] GOODFELLOW I J，POUGET-ABADIE J，MIRZA M，et al. Generative adversarial networks[J]. Advances in Neural Information Processing Systems，2014，3：2672

[11] NASH J F. Non-Cooperative games[J]. Annals of Mathematics，1951，54(2)：286. DOI：10.2307/1969529

[12] NASH J F. Equilibrium points in N-person games[J]. National Academy of Sciences，1950，36(1)：48

[13] HE Kaiming，ZHANG Xiongyu，REN Shaoqing，et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society，2016：770. DOI：10.1109/CVPR.2016.90

[14] LOFFE S，SZEGEDY C. Batch normalization：Accelerating deep network training by reducing internal covariate shift [J]. arXiv preprint arXiv：1502.03167，2015

[15] HE Kaiming，ZHANG Xiongyu，REN Shaoqing，et al. Delving deep into rectifiers：Surpassing human-level performance on ImageNet classification[C]//2015 IEEE International Conference on Computer Vision. IEEE Computer Society，2015：1026. DOI：10.1109/ICCV.2015.123

[16] 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. IEEE Computer Society，2017：1132. DOI：10.1109/CVPRW.2017.151

[17] RADFORD A，METZ L，CHINTALA S. Unsupervised representation learning with deep convolutional generative adversarial networks[J]. Computer Science，2015

[18] MAAS A L，HANNUN A Y，Ng A. Y. Rectifier nonlinearities improve neural network acoustic models[C]// Proc. icml. 2013，30(1)：3

[19] HAN Wei，CHANG Shiyu，LIU Ding，et al. Image super-resolution via dual-state recurrent networks [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition，2018：1654. DOI：10.1109/CVPR.2018.00178

（编辑　苗秀芝）