

通道加权下的双判别 GAN 超分辨率网络

张港 陈东方 王晓峰

(武汉科技大学计算机科学与技术学院 湖北 武汉 430065)

(武汉科技大学智能信息处理与实时工业系统湖北省重点实验室 湖北 武汉 430065)

摘要 针对现有基于生成对抗网络的单图超分辨率重构方法特征利用率不足,生成图像包含少量无意义噪声的问题,提出一种基于通道注意力机制的双判别生成对抗网络。通过对生成网络中密集残差块进行通道加权,优化网络的特征利用率。同时在对抗网络中对生成图像进行像素域和特征域的双重判别,促使生成网络产生更丰富的结构特征和低频信息。实验结果表明,与现有的 SRGAN、ESRGAN 两种算法相比,该算法能够重构出感官质量更高的图像。

关键词 图像超分辨率重建 生成对抗网络 通道注意力 特征判别器 视觉质量

中图分类号 TP391.41

文献标志码 A

DOI:10.3969/j.issn.1000-386x.2024.09.035

IMAGE SUPER-RESOLUTION WITH DUAL DISCRIMINANT GAN UNDER CHANNEL WEIGHTING

Zhang Gang Chen Dongfang Wang Xiaofeng

(School of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan 430065, Hubei, China)

(Hubei Province Key Laboratory of Intelligent Information Processing and Real-time Industrial System, Wuhan University of Science and Technology, Wuhan 430065, Hubei, China)

Abstract Existing single image super-resolution methods based on the generative adversarial network cannot make full use of features, and the generated image contains a small amount of meaningless noise. Therefore, this paper proposes a dual discriminant generative adversarial network based on channel attention mechanism. In the generation network, channel attention mechanism was used in the dense residual blocks to improve feature utilization rate. Simultaneously by dual discrimination of pixels and features on the generated image, richer structural features and high frequency information was promoted to produce. Experimental results show that compared with the existing SRGAN and ESRGAN algorithms, the proposed algorithm achieves lower NIQE and PI values and can reconstruct images with better perceptual quality.

Keywords Image super-resolution reconstruction Generative adversarial network Channel attention Feature discriminator Visual quality

0 引言

单图超分辨率重构^[1] (Single Image Super Resolution, SISR) 技术一直是计算机图像处理的一个研究热点。SISR 旨在将一幅低分辨率 (Low Resolution, LR) 图像恢复成高分辨 (High Resolution, HR) 图像, 常用于

安防监控, 医学图像对象识别等领域。随着机器学习技术的发展, SISR 的研究主要分为基于卷积神经网络 (Convolutional Neural Networks, CNN) 的方法与基于生成对抗网络 (Generative Adversarial Networks, GAN) 方法。

基于 CNN 的图像超分辨率方法将 SISR 看作重建问题, 通常使用峰值信噪比 (Peak Signal-to-Noise Ratio,

PSNR)作为评价指标,通过缩小高低分辨率图像之间的误差进行重建。如 SRCNN^[2]、VDSR^[3]、DBPN^[4]、ResNet^[5]等,重建出的图像在极大地提升了图像保真度的同时,往往也因为缺少高频细节而显得平滑模糊。基于 GAN 的图像超分辨率方法将 SISR 看作生成问题,通常以感知分数^[6](Perceptual Index, PI)、自然图像质量评价^[7](Natural Image Quality Evaluator, NIQE)为评价标准,追求更好的视觉质量。这种方法一般由生成器和判别器组成,在二者相互博弈中不断提高图像的视觉质量。但之前的方法生成的图像往往存在伪影,而且在生成图像的过程中忽略了特征图的不同通道之间表达能力的差异。为此,本文提出通道加权下的双判别 GAN 超分辨率网络(Channel Weighting Super-Resolution Generative Adversarial Network, CWSRGAN),算法优点如下:

(1) 提出通道加权下的密集残差块(Channel Weighting Residual Dense Block, CWRDB),对深层网络中的特征通道进行加权,提升生成网络特征的利用率。

(2) 使用双判别网络对生成图像的像素域和特征域进行判别,形成对生成图像结构信息和高频信息的有效监督,促使生成网络产生更与真实图像一致性更高的细节信息。

1 相关工作

1.1 基于生成对抗网络的超分辨率重建

Ledig 等^[8]首次将 GAN 生成对抗的思想引入超分辨率领域,提出 SRGAN 网络。SRGAN 使用视觉损失函数代替均方误差进行优化,模拟生成真实图像的高频信息,使用 VGG19 (Visual Geometry Group Network) 网络,对伪造图像的分布进行判别并对更新生成器的权重,能够重构出具有更丰富纹理细节的图像。本文基于标准的 GAN 框架构建模型^[9],主要解决的问题可表示为:

$$\min_G \max_D (E_{y \sim p_{\text{train}}(y)} [\ln(D(y))] + E_{x \sim p_G(x)} [\ln(1 - D(G(x)))]) \quad (1)$$

式中: x, y 为低分辨率图像和真实图像数据, $p_{\text{train}}(y)$ 、 $P_G(x)$ 为真实图像和噪声图像的分布, G 为生成网络,通过模拟真实数据的分布来生成逼真的噪声图像 $G(x)$, D 为判别网络,对生成的噪声图像和真实图像进行判别。训练生成网络 G 目的是伪造图像减小和真实图像分布的误差去欺骗判别网络 D ,减小 $\ln(1 - D(G(x)))$ 的值。判别网络 D 负责对来自真实数据和 G 伪造的数据进行判别,最大化真实数据和噪声数据

的样本标签分布的正确率 $D(\cdot)$ 。生成网络和判别网络相互博弈,促使生成网络生成视觉效果更好、与真实图像更为接近的图像。

1.2 通道注意力机制

近几年,通道注意力机制在基于深度学习方法的计算机视觉任务中引起了越来越多的关注。Hu 等^[10]提出挤压和激励(Squeeze and Excitation, SE)模块,通过全局平均池化将特征挤压到一维,然后在特征通道的压缩和扩张阶段使用阈值激励机制实现对通道特征表达能力的筛选。通过对表达能力的不同对通道赋予不同的权重实现对特征通道的加权,提高网络特征的利用率。Zhang 等^[11]提出 RCAN 网络,将通道注意力机制嵌入 RIR(Residual In Residual)结构,利用通道注意力机制来对特征图中通道赋予不同的权值来提高网络的特征表达能力。

受 Zhang 等^[13]的启发,我们提出 CWRNet(Channel Weighting Residual Network)生成网络,将密集残差^[12]和通道注意力机制嵌入 RIR 结构,通过对特征图中的不同通道间重新分配权重,使网络提取出更加具有表现力的特征,从而提高图像的重构质量。

2 通道加权下的双判别 GAN 超分辨率网络

本文提出的通道加权下的双判别 GAN 超分辨率网络 CWSRGAN 整体结构如图 1 所示,包括生成网络(图左)和判别网络(图右)两个部分。

生成网络 CWRNet 主要分为三个模块:浅层特征提取模块、深层注意力残差模块以及图像重建模块。浅层特征提取模块是一个 3×3 的卷积层,在 LR 上提取出浅层特征,并将其作为深层特征提取的输入。深层注意力残差模块由多个 CWRDB 模块组成,对接收的浅层特征图的残差图像进行深层特征提取。CWRDB 由密集连接残差模块(Residual Dense Block, RDB)与特征通道加权模块(Channel Weighting Block, CWB)两部分组成。RDB 对流经的特征信息进行密集特征提取,其中每一层的特征图都与之前所有层的特征图连接起来作为下一层的输入。CWB 位于 RDB 块之后,对提取的特征进行通道加权然后将其作为下一个 RDB 块的输入。最后的重建模块使用亚像素卷积对图像进行上采样操作,完成生成网络的图像伪造。接下来对生成图像 SR 进行像素域与特征域的判别:

(1) 像素域判别:将包含生成图像 SR 与真实图像

HR 的数据集送入 VGG19 分类网络进行判别。这一过程主要是图像之间对应的像素值进行比对,通过交叉熵损失函数 $loss^{pixel}$ 对 VGG19 进行优化。

(2) 特征域判别:将 SR 与 HR 分别送入 VGG19 网络,取中间层特征图 $\phi(SR)$ 与 $\phi(HR)$ 送入 VGG19 网络进行判别。相对应地,使用交叉熵损失函数

$loss^{feature}$ 对 VGG19 网络进行优化。

最后,对 VGG19 网络中的 SR 与 HR 中间特征图 $\phi(SR)$ 、 $\phi(HR)$ 取 L1 范数作为内容损失。使用像素域与特征域的对抗性损失与内容损失一起对生成网络进行优化,完成图像的重建。下面对生成网络和判别网络进行详细介绍。

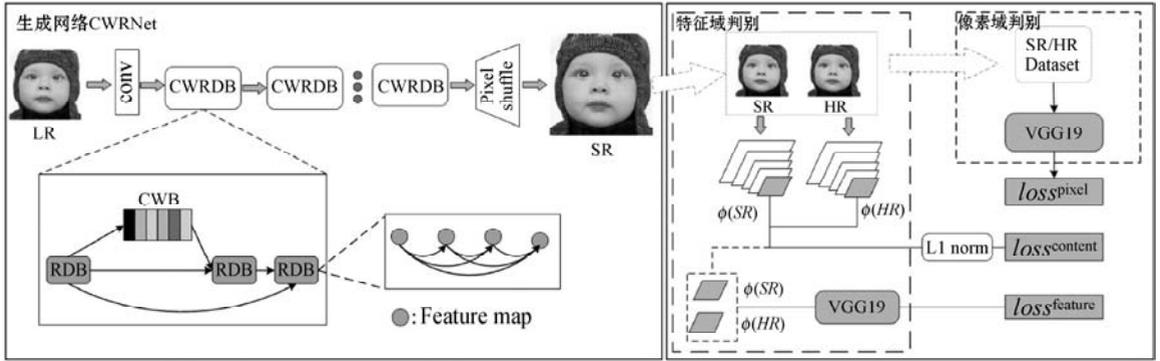


图 1 通道加权下的双判别 GAN 超分辨率网络 CWSRGAN 结构

2.1 通道加权下的密集残差生成网络

2.1.1 浅层特征提取模块

对于输入的低分辨率图像,使用一个 3×3 的卷积层来进行浅层特征提取:

$$F_{SF} = H_{SF}(I_{LR}) \quad (2)$$

式中: I_{LR} 表示输入的低分辨率图像; $H_{SF}(\cdot)$ 表示卷积操作; F_{SF} 为经过浅层提取后的特征。浅层特征提取模块将输入图像表示张量形式并对 LR 进行初步的特征提取,方便接下来的深层特征提取模块进行残差学习。

2.1.2 深层注意力残差模块

根据 RIR,深层注意力残差模块使用深层网络对浅层特征进一步提取,如图 1(左下)所示。

$$F_{DF} = F_{RIR}(F_{SF}) \quad (3)$$

式中: F_{DF} 表示深层注意力残差模块提取的深层特征; $F_{RIR}(\cdot)$ 表示本文搭建的 RIR 深层残差结构。深层注意力残差模块包含内外两层残差机制,如图 1 所示,对于输入的浅层特征图,外层残差机制使用跳跃连接将残差图像经过多个串联的 CWRDB 模块处理后的特征与浅层特征连接起来。

$$F_{DF} = F_{SF} + W_{SLC}(F_{last}) \quad (4)$$

式中: F_{last} 指最后一个 CWRDB 残差块的输出特征图。 W_{SLC} 表示 CWRDB 残差组尾部的卷积层操作,目的是将深度特征提取后的特征卷积到之前的尺寸,与浅层特征连接。

对于内部的每一个 CWRDB 模块,同样使用跳跃连接将首先使用跳跃连接将特征提取前后的残差图像进行拼接。相比于普通的残差块堆叠,内外两层的残

差结构将局域与全域的特征进行了重复利用,能够提取更多的特征信息。对于每一个流经 CWRDB 的特征信息有:

$$F_g = H_g(F_{g-1}) = H_g(H_{g-1}(\dots H_1(F_{SF}) \dots)) \quad (5)$$

式中: $H_g(\cdot)$ 表示特征信息流经第 g 个 CWRDB 块的处理过程, F_{g-1} 和 F_g 表示第 g 个 CWRDB 块的输入和输出。

CWRDB 使用无 BN (Batch Normalization) 层的卷积网络以保证训练过程的稳定性,包含密集残差特征提取 RDB 与通道特征加权 CWB 两部分。RDB 结构如图 2 所示。

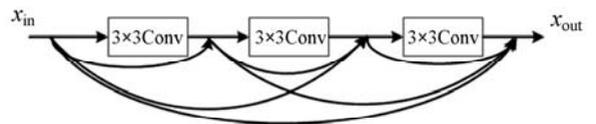


图 2 密集连接残差模块 (RDB)

输入的特征经过连续的 3×3 卷积层。跳跃连接将每一组卷积操作的特征图 (Feature map) x_{out} 与之前所有的特征图加起来作为下一组卷积操作的输入 x_{in} 。短跳跃连接密集地穿插在卷积层之间能够在保留上一层的信息同时保证网络进行残差学习。

特征通道加权部分 CWB 如图 3 所示,考虑到网络表达能力对不同特征通道的依赖性差异,引入通道注意力机制对密集残差提取过的特征进一步对特征通道加权。对于输入的特征信息 $X = [x_1, x_2, \dots, x_n]$,我们对其中的每一个特征图表示为 $H \times W \times C$,其中 $H \times W$ 为特征图的大小, C 为特征通道数。令 $Z \in R^C$ 表示对通道信息的统计值。

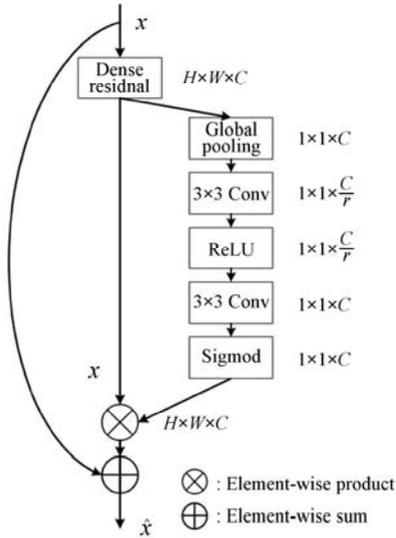


图3 特征通道加权模块(CWB)

首先使用平均池化层对特征图进行降维,此时第 n 个特征图通道的信息统计值可以表示为:

$$Z_n = H_{CP}(x_n) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W x_n(i, j) \quad (6)$$

式中: $x_n(i, j)$ 表示第 n 个通道的特征值, $H_{CP}(\cdot)$ 表示全局平均池化操作,此时的特征信息统计值 z_n 可以看作是特征图通道的一个描述符, m 描述特征通道表达能力的大小。为了捕捉不同通道之间表达能力的差异,设置如下阈值函数对通道描述符进行筛选并重新分配权重:

$$s = f(W_U \delta(W_D z)) \quad (7)$$

式中: f 和 δ 分别表示 Sigmoid 和 ReLU 激活函数, W_D 和 W_U 分别表示卷积操作对特征通道数量赋予的缩减和扩增系数 r 。然后引入残差的思想通过矩阵对应元素相乘将学习到的权重赋予特征信息:

$$\hat{x}_n = s_n \cdot x_n \quad (8)$$

式中: s_n 和 x_n 分别表示第 n 特征图通道的新权重和第 n 个特征图。最后使用矩阵相应元素求和将学习到的残差特征与输入特征进行连接。

2.1.3 图像重建模块

为了将图像恢复到真实图像大小,使用亚像素卷积层对特征图进行上采样并进行重建。

$$I_{SR} = H_{REC}(H_{UP}(F_{DF})) \quad (9)$$

式中: $H_{UP}(\cdot)$ 表示亚像素卷积上采样操作, $H_{REC}(\cdot)$ 表示重建操作。

为突出对残差网络进行通道加权的有效性,同时为了方便与之前的工作进行比较,本文采用 MSE 作为损失函数优化生成网络中重构图像与真实图像的误差:

$$Loss_{MSE} = \frac{1}{H \times W \times C} \sum_{i=1}^H \sum_{j=1}^W \sum_{k=1}^C (I_{i,j,k}^{SR} - I_{i,j,k}^{HR})^2 \quad (10)$$

式中: H 、 W 、 C 表示图像的高、宽和通道数。 $I_{i,j,k}^{SR}$ 和 $I_{i,j,k}^{HR}$

分别表示重建图像和 HR 图像上的相对应的像素值。

2.2 双判别网络

一般来说,生成网络在伪造图像对判别器进行欺骗时,最简单直接的方法便是生成任意能够弥补高低分辨率图像之间高频信息差距的噪声。这些噪声一部分作为图像的高频信息组成图像细节纹理,另一部分则构成图像的特征结构,使图像看上去更加自然。之前的工作在对生成图像进行判别时,只是同原始图像进行逐像素的比对。生成图像中的结构性特征和语义信息不能得到有效监督,这些没有经过约束的无意义的噪声就形成了图像中存在的少量伪影。本文使用 VGG19 分类网络作为判别网络对生成图像进行像素域和特征域上的双重判别。与像素值相比,生成图像的特征图包含更多的结构信息与语义信息。通过增加对图像特征域的判别能对生成网络拟合真实图像的分布加强监督,抑制重建图像伪影的产生。

判别网络结构如图 4 所示。整个网络卷积操作都使用同样大小的 3×3 卷积核。对于输入的图像,使用连续的 3×3 的卷积层对网络加深,随着每次卷积操作的进行,图像的特征图大小都会减半,通道数量扩大两倍。最后通过全连接层整合特征,输出判别结果。

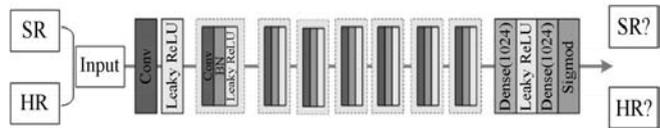


图4 判别网络结构

2.2.1 像素域判别

首先将生成图像和真实图像的混合数据集直接送入判别网络进行像素域判别。这一过程主要是通过检查图像的像素值来区分真实图像和生成图像,我们使用交叉熵损失函数的形式来优化判别网络:

$$loss_{dis}^{pixel} = -\ln(d^{pixel}(I^{HR})) - \ln(1 - d^{pixel}(I^{GEN})) \quad (11)$$

式中: I^{HR} 表示真实图像; I^{GEN} 表示生成图像; $d^{pixel}(\cdot)$ 表示像素域判别器的输出结果。

2.2.2 特征域判别

接着为了提高生成网络伪造图像的能力,在对图像的像素域判别基础上,增加对图像特征域的判别。我们依旧采用上图的判别网络结构,不同的是将经过 VGG19 提取后的特征图送入判别网络。相对应的,特征域判别网络的损失函数可表示如下:

$$loss_{dis}^{feature} = -\ln(d^{feature}(\phi^m(I^{HR}))) - \ln(1 - d^{feature}(\phi^m(I^{GEN}))) \quad (12)$$

式中: $d^{feature}(\phi^m)$ 表示在第 m 层图像的特征图来自真实图像的概率。

在对抗过程中,生成网络使用以视觉质量为导向的新损失函数进行优化:

$$Loss_{generate} = loss_{content} + loss_{adversarial}^{pixel} + loss_{adversarial}^{feature} \quad (13)$$

式中: $loss_{content}$ 表示 vgg19 网络特征提取的重建图像与真实图像特征图之间的内容损失,能促使网络产生更丰富的高频信息, $loss_{adversarial}^{pixel}$ 与 $loss_{adversarial}^{feature}$ 表示像素域与特征域的对抗性损失,通过判别网络对生成图像的判断结果反过来指导生成网络更好地拟合真实图像的分布。其中 $loss_{content}$ 、 $loss_{adversarial}^{pixel}$ 、 $loss_{adversarial}^{feature}$ 分别表示如下:

$$Loss_{content} = \frac{1}{H_m \times W_m \times C_m} \sum_{i=1}^{H_m} \sum_{j=1}^{W_m} \sum_{k=1}^{C_m} (\phi_{i,j,k}^m(I^{HR}) - \phi_{i,j,k}^m(I^{SR}))^2 \quad (14)$$

$$loss_{adversarial}^{pixel} = -\ln(d^{pixel}(I^{GEN})) \quad (15)$$

$$loss_{adversarial}^{feature} = -\ln(d^{feature}(I^{GEN})) \quad (16)$$

式中: H_m 、 W_m 、 C_m 分别指代 VGG19 结构第 m 层特征图的高度、宽度和通道的数目, $\phi_{i,j,k}^m(I^{HR})$ 表示重建图像在流经 VGG19 网络中第 m 层特征图的高度为 i 、宽度为 j 、通道数为 k 点的像素值。相对应的, $\phi_{i,j,k}^m(I^{SR})$ 表示真实图像在流经 VGG19 网络中第 m 层特征图的高度为 i 、宽度为 j 、通道数为 k 点的像素值。式(15)和式(16)分别表示像素域和特征域判别网络的对抗性损失, $d^{pixel}(I^{GEN})$ 表示生成图像来自真实图像的概率, $d^{pixel}(I^{GEN})$ 越大, $loss_{adversarial}^{pixel}$ 越接近 0,表示生成图像在像素域上越接近真实图像。相应地, $d^{feature}(\phi^m(I^{GEN}))$ 表示生成图像的特征图来自真实图像的概率, $d^{feature}(\phi^m(I^{GEN}))$ 越大, $loss_{adversarial}^{feature}$ 越接近 0,表示生成图像在特征域上越接近真实图像。

3 实验结果及分析

3.1 实验细节和参数设置

实验训练网络使用 Adam 优化器,动量参数为 0.9,初始学习率为 10^{-4} ,使用余弦退火方式来调节学习率的衰减,即随着迭代次数的增加,学习率随着余弦值的下降模式递减,损失函数可以更谨慎地接近全局最小值。根据文献[14],我们对于式(12)、式(14)中 m 的取值为 5,因为在 VGG19 网络中第 5 层的特征信息更为丰富。考虑到计算复杂度与时间成本,在深层注意力残差模块中我们将 CWRDB 个数设为 23。同时每个特征通道加权模块中扩增系数 r 为 4。实验使用 Ubuntu 18.04 系统,PyTorch 深度学习框架,CPU 为 Intel Core i7-4700k,GPU 为 RTX-2070。

为了与之前工作保持一致,我们使用 DIV2K^[15] 数据集中的 800 幅高分辨图像作为训练数据集,Set5 为验证集数,测试集使用 Set5、Set14、BSD100、Urban100 等多种公开数据集。为了扩充数据,对训练集进行 90、80、27 度的旋转,水平翻转,以及随机裁剪,使数量扩充到原来的 4 倍。考虑到内存占用及计算的复杂性,我们以步长为 240,将高清图像裁剪成 480×480 的子图像,然后使用双三次插值算法对训练集进行 4 倍的下采样,设置批大小为 16。因此训练网络的每一个最小批训练输入为 $16 \times 128 \times 128$ 。

模型训练分为预训练与对抗训练两个过程。首先使用 MSE 损失函数对生成网络 CWRNet 模型进行预训练。然后将预训练的 CWRNet 模型作为初始权重赋予生成网络,与判别网络进行对抗博弈。在对抗过程中,使用新的视觉损失函数对生成网络进行优化以产生视觉效果更好的图像。

3.2 评价标准

本文采用 PSNR^[16] (峰值信噪比)、SSIM^[17] (结构相似性)、NIQE(自然图像质量评价)和 PI(感知指数)来对实验的结果进行度量。

给定 $M \times N$ 大小的当前图像 X 和参考图像 Y ,PSNR 计算式如下:

$$NSE = \frac{1}{M \times N} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} (X(i,j) - Y(i,j))^2 \quad (17)$$

$$PSNR = 10 \log_{10} \left(\frac{(2^{bits} - 1)^2}{MSE} \right) \quad (18)$$

式中: MSE 表示当前图像和参考图像的均方误差, $bits$ 表示当前像素的二进制数,通常为 8。 $PSNR$ 单位是 dB, $PSNR$ 的值越大,表示当前图像与参考图像越接近,重建效果越好。

SSIM 基于当前图像的亮度,对比度和结构三个方面对图像质量进行评估。给定 $M \times N$ 大小的当前图像 X 和参考图像 Y ,SSIM 计算式如下:

$$L(X,Y) = \frac{2\mu_X\mu_Y + C_1}{\mu_X^2 + \mu_Y^2 + C_1} \quad (19)$$

$$C(X,Y) = \frac{2\sigma_X\sigma_Y + C_2}{\sigma_X^2 + \sigma_Y^2 + C_2} \quad (20)$$

$$S(X,Y) = \frac{\sigma_{XY} + C_3}{\sigma_X\sigma_Y + C_3} \quad (21)$$

$$SSIM(X,Y) = L(X,Y) \cdot C(X,Y) \cdot S(X,Y) \quad (22)$$

式中: μ_X 、 μ_Y 表示 X 和 Y 的均值; σ_X^2 、 σ_Y^2 为 X 和 Y 的方差; σ_{XY} 表示 X 、 Y 的协方差。为避除零, $C_1 = (k_1(2^b - 1))^2$, $C_2 = (k_2(2^b - 1))^2$ 为两个常数,通常默认

k_1 取 0.01, k_2 取 0.03, B 为 8, $C_3 = \frac{C_2}{2}$ 。 L 、 C 、 S 分别表示图像在亮度、对比度和结构三个方面的评价指标。SSIM 的取值范围是 0 到 1, SSIM 的值越大, 表示图像的失真越小, 重建效果越好。

NIQ 是一种无参考图像质量评价指标^[18]。它通过原始图库提取的图像特征建立多元高斯 (Multivariate Gaussian, MVG) 模型, 对提取的失真图像特征进行拟合计算与自然图像之间的距离来评估图像的质量。MVGE 模型如下:

$$f_X(x_1, x_2, \dots, x_k) = \frac{1}{(2\pi)^{\frac{k}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x - \nu)^T \Sigma^{-1}(x - \nu)\right) \quad (23)$$

式中: x_1, x_2, \dots, x_k 指原始图库提取的图像特征, 通过最大似然估计可以得到表示 MVG 模型的均值向量和协方差矩阵 ν, Σ 。NIQE 的最终得分通过衡量待测图像与自然图像参数之间的距离得到:

$$D(\nu_1, \nu_2, \Sigma_1, \Sigma_2) = \sqrt{\left(\nu_1 - \nu_2\right)^T \left(\frac{\Sigma_1 + \Sigma_2}{2}\right)^{-1} (\nu_1 - \nu_2)} \quad (24)$$

式中: ν_1, Σ_1 为自然图像的均值向量和协方差矩阵, ν_2, Σ_2 表示当前待测图像的均值向量和协方差矩阵。NIQE 的评分越低, 表示图像的视觉质量越高, 重建效果越好。

PI[] 的计算式如下:

$$Ma = \frac{1}{T} \sum_{i=0}^T p_i(x_n | \Theta) \quad (25)$$

$$PI = \frac{1}{2}((10 - Ma) + NIQE) \quad (26)$$

式中: Ma 使用三个独立的回归森林拟合图像的局部频域特征, 全局频域特征以及空间不连续特征。它们的输出在 PI 上线性回归, 用来对待测图像进行评估打分。其中 x_n 表示图像的一种低级特征, $p_i(x_n | \Theta)$ 表示森林中第 i 个决策树的得分。

3.3 CWRDB 模块对重构结果的影响

本节验证 CWRDB 模块对模型性能的有效性。实验使用 Set5 数据集对加入普通残差模块和 CWRDB 的 CWRNet 在 RGB 空间上的平均 PSNR 值进行对比。如图 5 所示, R-CWRNet 表示加入了残差块的 CWRNet 网络, CR-CWRNet 表示加入了 CWRDB 的 CWRNet 网络。从实验结果可以看出, 本文提出的 CWRDB 模块相较于普通残差块在图像的重建效果上优势明显, 其原因为 CWRDB 模块通过对流经的特征信息重新加权, 同时扩大了感受野, 重建出的图像能够得到更多的高频信息, 因此重建出的图像效果更好。

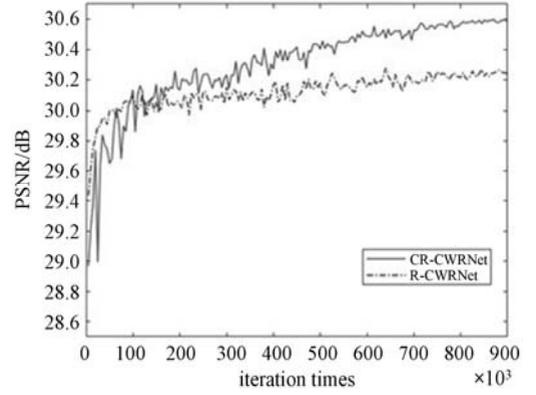


图5 CWRDB 模块加入前后 CWRNet 的 PSNR 变化与迭代次数的关系

接着验证实验的统计结果如表 1 所示, 可以看到使用了 CWRDB 的 CWRNet 较使用普通残差的 CWRNet 最终在 PSNR 上平均提升 0.33 dB 左右。对于 Set5 数据集中的五幅图片, 加权后的残差块性能都要高于普通残差块。

表 1 使用不同残差模块在 Set5 数据集上 PSNR 值的对比

Set5	普通残差块	加权后的残差块
baby	32.20	32.33
bird	32.79	33.43
butterfly	27.28	27.85
head	29.93	29.96
woman	29.04	29.43

3.4 双判别网络对重构效果的影响

3.4.1 单双判别网络对重构效果影响的对比

本节验证双判别网络相对于单判别网络在重构图像视觉质量提升的有效性。实验使用 Set5 数据集对两个模型的平均 NIQE 值进行对比。如表 2 所示, 只含单像素域判别网络的为 P-CWSRGAN, 包含像素和特征双重判别的为 FP-CWSRGAN。实验结果中双判别网络下的 CWSRGAN 要比单像素域判别重构网络平均低 0.18 个百分点左右, 这表明本文提出的双判别网络比传统的单判别网络更好。

表 2 使用不同判别网络在 Set5 数据集下的 NIQE 值

Set5	P-CWSRGAN	FP-CWSRGAN
baby	2.55	2.47
bird	3.42	3.25
butterfly	6.01	5.84
head	3.18	2.81
woman	3.42	3.25

3.4.2 双判别网络对重构效果的影响

本节对本文模型 GAN 网络对抗训练前后作对比

试验。实验选取 Set14 数据集中的一幅 comic 图像进行重构并对其进入判别网络对抗训练前后的视觉效果进行对比。实验结果图像使用 sobel 算子边缘检测^[19]来对图像结构信息与高频细节丰富程度进一步刻画。如图 6 所示,(a)为原始图像及其 sobel 边缘检测结果,(b)、(c)为 4 倍效果下生成网络和判别网络对抗训练后的重构图像及 sobel 边缘检测结果。以图像中少女的头冠纹饰为例,图 6(b)中相同区域中的纹饰细节是一堆模糊的色块,而在图 6(c)中生成的细节信息与原始区域虽然有一定差距,但是可以较为清楚地看出纹饰的纹路,在视觉效果上比 CWRNet 的结构要更加清楚,结构信息更加丰富。从实验结果可以看出,CWRNet 的边缘检测结果与真实图像相比存在大量的空白区域,而 CWSRGAN 的边缘信息的分布与细密程度上与真实图像更为接近。因此在结构信息的丰富程度上,经过双判别 GAN 网络对抗训练后的模型重建出的图像更接近于真实图像。

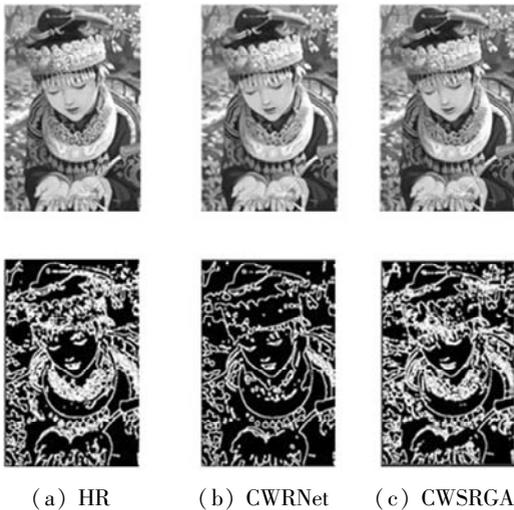


图 6 Set14 数据集中 comic 图像在 4 倍重构过程中的效果对比

3.5 与当前先进方法比较

3.5.1 主观效果比较

本节将提出的 CWRNet 生成网络与 Bicubic、CWRNet 生成网络以及当前流行的几种算法^[2,3,12,20]进行主观效果对比实验。实验选用 Set5、Set14、Urban100 中纹理细节较为丰富的三幅图像进行重建。如图 7、图 8 所示,实验在缩放系数为 4 的情况下进行,将图像的特定区域放大处理并在原始图像中用红框标注。以图 7 中 Set5 数据集的 Baby 图像为例,对于 baby 帽子的额头矩形区域进行放大,可以看到本文提出的 CWSRGAN 效果要明显比其他几种算法好,其中 Bicubic 算法重构的效果失真最严重,而基于 MSE 损失函数的 SRCNN、RDN^[12],以及我们的生成网络 CWRNet 都因为细节信息不足,相比之下显得模糊。除此之外,SRGAN 算法

重构出的图像虽然包含有丰富的信息,但是通过观察可以发现,在毛线交织的沟壑间存在白色噪声,而且局部结构坍塌,图像纹路不清晰。ESRGAN^[21]的结果与 SRGAN 相比,结构稍显分明,纹理稍显清楚,但是噪声点的出现仍然存在。本文提出的 CWSRGAN 算法重建出的图像细节丰富,脉络鲜明,看上去更加自然,视觉效果更好。其原因为在使用 CWRNet 生成丰富细节的同时还包含更多的结构信息,使用的双判别网络抑制了无意义噪声点的产生,对图像中线条的纹理处理更加自然。

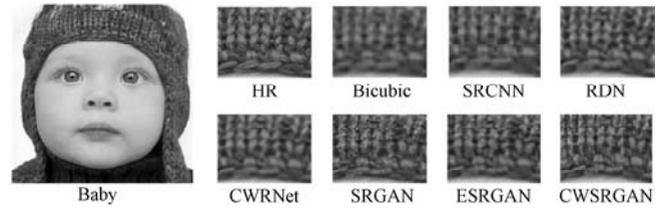


图 7 4 倍放大系数下 baby 在不同算法下的重构效果对比图

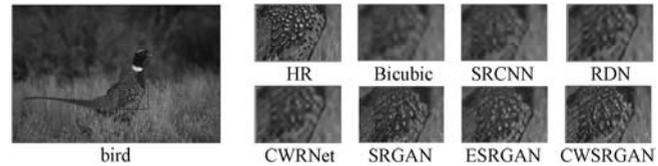


图 8 4 倍放大系数下 bird 在不同算法下的重构效果对比图

3.5.2 客观定量结果对比

最后将本文提出的算法与现有的流行算法进行客观定量对比实验。实验数据如表 3 所示,实验首先选用 Bicubic、SRCNN、VDSR、IDN、SRResNet、SRFeat 与 CWRNet 生成网络进行对比实验。图像在 2、3、4 倍缩放因子下都进行了实验。从实验结果可以看出,本文提出的 CWRNet 与传统的 Bicubic、SRCNN 与 VDSR 算法相比有较为明显的优势,对于 IDN 和 SRFeat, CWRNet 在 PSNR 上平均提升了 0.3 dB 左右,在 SSIM 上提升了 0.04 左右。实验结果表明,本文提出的通道加权的密集残差生成网络在生成图像的效果上达到了预期。

接着我们对本文提出的 CWSRGAN 与 SRGAN、ESRGAN 和 CWRNet 进行对比实验。实验选用 NIQE 与 PI 两种与人类感知具有更好一致性的评价指标。实验结果如表 4 所示,在四种不同的公开数据集上 CWSRGAN 都取得了比较好的实验结果。其中,在 NIQE 上与 SRGAN 和 ESRGAN 相比分别平均降低了 0.75 个百分点和 0.55 个百分点。在 PI 上与 SRGAN 和 ESRGAN 相比分别平均降低了 0.37 个百分点左右和 0.13 个百分点左右。实验证明,对于 Set5、Set14、BSD100、Urban100 数据集,本文提出的 CWSRGAN 在 NIQE 和 PI 上都要优于 SRGAN 与 ESRGAN 两种算法。

表 3 本文算法 CWRNet 与其他 6 种算法在不同测试集下的 PSNR、SSIM 值对比

datasets	scale	Bicubic	SRCNN	VDSR	IDN	SRResNet	SRFeat	CWRNet
		PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
Set5	2	33.66/0.929 8	36.66/0.954 1	37.54/0.958 6	37.83/0.960 2	37.90/0.959 4	37.95/0.959 8	38.27/0.961 0
	3	30.39/0.867 9	32.74/0.909 0	33.65/0.921 8	34.12/0.925 1	34.32/0.926 2	34.25/0.921 0	34.70/0.929 5
	4	28.41/0.821 1	30.48/0.862 8	31.18/0.882 6	31.92/0.887 6	32.17/0.896 1	32.14/0.891 8	32.55/0.899 8
Set14	2	30.24/0.868 8	32.44/0.907 2	33.07/0.913 0	33.30/0.914 7	33.65/0.917 2	34.05/0.911 0	34.10/0.921 5
	3	27.52/0.772 8	29.29/0.821 3	29.76/0.832 0	29.98/0.834 5	30.39/0.841 5	30.58/0.846 3	30.75/0.849 0
	4	26.08/0.747 2	27.48/0.758 5	27.79/0.826 8	28.51/0.783 3	28.60/0.783 0	28.61/0.781 6	28.81/0.788 3
BSD100	2	29.56/0.843 1	31.36/0.887 9	31.90/0.896 0	32.10/0.899 0	32.21/0.900 5	32.48/0.903 8	32.46/0.903 5
	3	27.21/0.738 5	27.41/0.786 3	28.82/0.797 6	28.90/0.801 3	29.13/0.807 5	29.53/0.812 0	29.40/0.812 4
	4	25.96/0.667 6	26.96/0.710 8	27.29/0.762 6	27.48/0.734 5	27.61/0.739 0	27.59/0.735 7	27.73/0.743 7
Urban100	2	26.88/0.839 8	29.52/0.894 6	30.77/0.914 1	31.27/0.919 5	31.85/0.926 0	32.05/0.930 0	33.54/0.9401
	3	24.45/0.735 2	26.24/0.798 8	27.15/0.828 9	27.41/0.834 8	28.25/0.846 3	28.37/0.855 5	29.32/0.873 8
	4	23.15/0.657 7	24.52/0.722 1	25.18/0.754 6	25.86/0.785 6	26.18/0.789 8	26.53/0.803 7	26.70/0.805 7

表 4 4 倍放大下算法 CWSRGAN 与其他 3 种算法在不同测试集下的 NIQE、PI 对比(%)

数据集	Set14		Set14		BSD100		Urban100	
	NIQE	PI	NIQE	PI	NIQE	PI	NIQE	PI
SRGAN	4.81	3.05	4.40	2.98	4.53	2.37	4.73	2.09
ESRGAN	4.67	2.65	4.21	2.61	4.15	2.28	4.64	1.98
CWSRGAN	4.36	2.47	3.56	2.54	3.65	2.16	3.92	1.85

4 结 语

本文提出了一种通道加权下的双判别 GAN 超分辨率重构方法。首先基于通道注意力机制通过对低分辨率图像特征密集地逐层提取并进行通道加权,在增加网络感受野的基础上进一步提高了特征利用率。然后引入生成对抗网络的架构,增加对生成图像特征域的判别促使网络生成图像蕴含丰富的结构信息,同时抑制了图像无意义噪声点的产生。实验结果表明,在 Set5、BSD100 等四种公开的数据集上,本文算法能够重建出细节丰富,结构明晰的高分辨率图像,且对于现有的几种流行算法在定量指标以及主观效果上都有显著的提升。但是需要指出的是,由于去除网络中的 BN 层,本文提出的模型在生成网络训练的收敛速度较慢。今后的工作将集中在网络结构的优化方面,开展基于 GAN 的轻量级网络研究。

参 考 文 献

[1] 李彬,喻夏琼,王平,等. 基于深度学习的单幅图像超分辨率重建综述[J]. 计算机工程与科学,2021,43(1):112 -

124.

- [2] Dong C, Loy C, He K M, et al. Image super-resolution using deep convolutional networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2016,38(2):295 - 307.
- [3] Kim J, Lee J K, Lee K M. Accurate image super-resolution using very deep convolutional networks[C]//IEEE Conference on Computer Vision and Pattern Recognition, 2016: 1646 - 1654.
- [4] Haris M, Shakhnarovich G, Ukita N. Deep back-projection networks for super-resolution [C]//IEEE Conference on Computer Vision and Pattern Recognition, 2018: 1664 - 1673.
- [5] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//IEEE Conference on Computer Vision and Pattern Recognition,2016:770 - 778.
- [6] Blau Y, Mechrez R, Timofte R, et al. The 2018 Pirm challenge on perceptual image super-resolution [C]//European Conference on Computer VisionWorkshops,2018:334 - 355.
- [7] Mittal A, Soundarajan R, Bovil A C. Making a 'completely blind' image quality analyzer [J]. IEEE Signal Processing Letters,2013,20(3):209 - 212.
- [8] Ledig C, Theis L, Huszár F, et al. Photo-realistic single

- image super-resolution using a generative adversarial network [C]//IEEE Conference on Computer Vision and Pattern Recognition,2017:4681-4690.
- [9] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[C]//27th International Conference on Neural Information Processing Systems,2014:2672-2680.
- [10] Hu J, Shen L, Sun G. Squeeze-and-excitation networks [C]//IEEE Conference on Computer Vision and Pattern Recognition,2018:7132-7141.
- [11] Zhang Y L, Li K P, Li K C, et al. Image super-resolution using very deep residual channel attention networks [C]//European Conference on Computer Vision,2018:286-301.
- [12] Huang G, Liu Z, Maaten L, et al. Densely connected convolutional networks [C]//IEEE Conference on Computer Vision and Pattern Recognition,2017:4700-4708.
- [13] Dai T, Cai J R, Zhang Y B, et al. Second-order attention network for single image super-resolution [C]//IEEE Conference on Computer Vision and Pattern Recognition, 2019: 11065-11074.
- [14] Park S J, Son H, Cho S, et al. Srfat: Single image super-resolution with feature discrimination [C]//European Conference on Computer Vision,2018:439-455.
- [15] Timofte R, Agustsson E, Van Gool L, et al. Ntire 2017 challenge on single image super-resolution: Methods and results [C]//IEEE Conference on Computer Vision and Pattern Recognition Workshops,2017:114-125.
- [16] Horé A, Ziou D. Image quality metrics: PSNR vs. SSIM [C]//20th International Conference on Pattern Recognition, 2010:2366-2369.
- [17] 佟雨兵,张其善,祁云平.基于 PSNR 与 SSIM 联合的图像质量评价模型[J].中国图象图形学报,2006,11(12):19-24.
- [18] 王慧,冯金顺,程正兴.基于局部路径特征信息神经网络的图像去噪[J].液晶与显示,2020,35(1):70-79.
- [19] 张宇伟,王耀明,蒋慧钧.一种结合 sobel 算子和小波变换的图像边缘检测方法[J].计算机应用与软件,2007,24(4):133-134,161.
- [20] Park S J, Son H, Cho S, et al. SRFeat: Single image super-resolution with feature discrimination [C]//European Conference on Computer Vision,2018:455-471.
- [21] Wang X T, Yu K, Wu S X, et al. ESRGAN: Enhanced super-resolution generative adversarial networks [C]//European Conference on Computer Vision,2018:63-79.
- 图构建方法综述[J].计算机辅助设计与图形学学报,2016(28):855-868.
- [2] 陈超,李强,闫青.基于异质传感器信息融合的移动机器人同步定位与构图[J].科学技术与工程,2018,18(13):86-91.
- [3] Mouragnon E, Dekeyser F, Sayd P, et al. Real time localization and 3D reconstruction [C]//IEEE Conference on Computer Vision & Pattern Recognition. IEEE, 2006.
- [4] Klein G, Murray D. Parallel tracking and mapping for small AR workspaces [C]//IEEE & ACM International Symposium on Mixed & Augmented Reality, 2008.
- [5] Mur-Artal R, Montiel J M M, Tardos J D. ORB-SLAM: A versatile and accurate monocular SLAM system [J]. IEEE Transactions on Robotics, 2017, 31(5):1147-1163.
- [6] Qin T, Li P, Shen S. VINS-Mono: A robust and versatile monocular visual-inertial state estimator [J]. IEEE Transactions on Robotics, 2018, 34(4):1004-1020.
- [7] Yijia H, Ji Z, Yue G, et al. PL-VIO: Tightly-coupled monocular visual-inertial odometry using point and line features [J]. Sensors, 2018, 18(4):1159.
- [8] Nister D. An efficient solution to the five-point relative pose problem [C]//2003 IEEE Conference on Computer Vision and Pattern Recognition, 2003.
- [9] Ng P C, Henikoff S. SIFT: predicting amino acid changes that affect protein function [J]. Nucl Acids Research, 2003, 31(13):3812-3814.
- [10] Rosten E, Porter R, Drummond T. Faster and better: A machine learning approach to corner detection [J]. IEEE Transactions on Software Engineering, 2010, 32(1):105-119.
- [11] Harris C, Stephens M. A combined corner and edge detector [C]//Alvey Vision Conference, 1988.
- [12] Gioi R G V, Jérémie J, Morel J M, et al. LSD: A line segment detector [J]. Image Processing On Line, 2012, 2(4):35-55.
- [13] Zhang L, Koch R. An efficient and robust line segment matching approach based on LBD descriptor and pairwise geometric consistency [J]. Journal of Visual Communication and Image Representation, 2013, 24(7):794-805.
- [14] Shen S, Michael N, Kumar V. Tightly-coupled monocular visual-inertial fusion for autonomous flight of rotorcraft MAVs [J]. Proceedings IEEE International Conference on Robotics & Automation, 2015, 2015:5303-5310.
- [15] Triggs B, Zisserman A, Szeliski R. Bundle adjustment—A modern synthesis [C]//International Workshop on Vision Algorithms,1999:298-372.

(上接第 235 页)

参 考 文 献

[1] 刘浩敏,章国锋,鲍虎军.基于单目视觉的同时定位与地