

基于对偶学习的密集连接超分辨率网络

曾令贤

同济大学电子与信息工程学院, 上海

收稿日期: 2022年11月20日; 录用日期: 2022年12月19日; 发布日期: 2022年12月26日

摘要

深度学习的方法大大提升了超分辨率技术的性能,但是由于超分辨率重建本身是一个不适定的工作,导致生成的效果不稳定,且对于不同通道的平均化处理也会导致图像的特征难以凸显。本文使用了对偶学习的策略,通过引入SR图像下采样的子任务,减小解空间,进一步约束超分图像的生成过程;同时使用了密集连接的残差通道注意力模块,使图像的特征有所侧重的同时,又能较好地传递。最后通过与经典算法的对比和消融实验,验证了本方法具有较好的生成效果,且我们的改进也切实有效。

关键词

超分辨率, 深度学习, 对偶学习

Dense Connected Super-Resolution Networks Based on Dual Learning

Lingxian Zeng

College of Electronic and Information Engineering, Tongji University, Shanghai

Received: Nov. 20th, 2022; accepted: Dec. 19th, 2022; published: Dec. 26th, 2022

Abstract

The deep learning method greatly improves the performance of super-resolution. However, since the super-resolution reconstruction itself is an ill-posed work, the generated effect is unstable, and the averaging processing for different channels will also cause the image features to be difficult to highlight. In this paper, the dual learning strategy is used to reduce the solution space and further constrain the SR image generation process by introducing the sub task of SR image down-sampling; at the same time, the Residual Dense Channel Attention Blocks are used, so that the image features are focused and can be better transmitted. Finally, through the comparison with the classical algorithm and the ablation experiment, it is verified that this method has a better generation effect, and our improvement is also effective.

Keywords

Super Resolution, Deep Learning, Dual Learning

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

图像, 作为现代社会中非常常见的一种信息载体, 能够传递丰富的信息。它通过色彩、明暗、结构等多维度的信息, 传递信息到人类的视觉系统, 有着语言、文字无法比拟的优势。而图像分辨率, 是对图像信息量大小的一种衡量, 是评价图像在细节信息表现力方面是否优秀的一个重要指标。相对于低分辨率(Low Resolution, LR)图像来说, 高分辨率(High Resolution, HR)象征着更高的像素密度, 也因此能为人们提供更加丰富的纹理细节以便于做出更准确的判断, 故而高分辨率图像在众多数字图像应用领域中都有着极大的需求[1]。然而在现实生活当中, 由于环境干扰、成像设备缺陷、传输失真、图像降质模型自身影响等各种约束条件的存在, 直接获得细节清晰、边缘完整的理想高分辨率图像通常是很困难的[2]。在这种情况下, 采取一些有效的措施来提高图像的分辨率就显得尤为重要。

在提升图像分辨率的方法中, 对光学硬件设备进行改良是最为直截了当的一种举措, 但是一般来说高精度的成像设施其造价都较为高昂; 另一方面, 对一些特殊场合(如医疗影像、航空航天、卫星遥感、军事侦察等)下的成像系统而言, 受限于技术水平方面的瓶颈, 其制造工艺难以大幅改进, 面临着一些很难攻克的问题[3]。因此, 以软件和算法作为出发点来获得更加符合要求的高清图片是非常迫切且至关重要的, 超分辨率重建(Super-Resolution, SR)技术就在这种背景下诞生了。顾名思义, 超分辨率重建技术就是运用一些特殊的方式对给定的 LR 图像进行处理, 根据其所含有的碎片信息最终恢复出一幅 HR 图像的方法[4]。相比起直接优化图像的采集设备、成像设备, 通过 SR 技术来直接改善图像效果, 更加方便, 成本也更加低廉, 具有大范围推广的可行性, 应用价值较高。

目前, 基于深度学习的超分辨率技术取得了显著优于传统方法的效果, 随着网络的加深和不同性能模块的引入, 图像的生成质量不断提高, 但是客观评价指标的高低与人眼视觉感知的效果并不等价。基于像素层面的优化往往会缺失对高频信息的关注, 导致重构的图像趋于像素平均化, 与人眼感知效果有一定的出入, 同时因为超分辨率的生成过程是一个不定问题, 同一个低分辨率图像可能会对对应着多张不同的高分辨率图片, 导致生成的效果并不稳定, 会偏离人眼的感知。

本文提出了一种基于对偶学习的超分辨率方法, 通过添加 SR 图像下采样的子任务, 进一步生成约束图像的解空间, 使传统的 LR \rightarrow HR 映射转变成了 LR \leftrightarrow HR 的闭环双向映射, 使得生成图像更加贴近自然世界的图片; 同时引入了密集连接的残差通道注意力模块, 改变不同通道平均化的问题, 使得图像的特征能够更好地呈现, 获得了更好的主观视觉效果。

2. 超分辨率重构理论基础

2.1. 图像超分辨率的基本概念

图像超分辨率重构技术顾名思义, 便是基于原始信息, 获取成像极限之外的信息, 提高图像的分辨率, 也就是通过像素质量较低的低分辨率图像来恢复与之对应的高分辨率图像。我们所拥有的 LR 图像

可以看作是 HR 图像通过一定的退化过程而来, 如式(1):

$$I_{LR} = D(I_{HR}, \delta) \quad (1)$$

其中 D 代表一定的退化方式, δ 为退化过程中的相关参数(比如噪声、尺寸等)。我们在获取到一张 LR 图片等时候, 对于他的退化过程是未知的(即 D 和 δ 未知), 而我们则要通过一定的方法, 来生成一张超分辨率图片, 我们希望这张 SR 图片能够无限地趋近于真实的 HR 图片。超分辨率需要解决的其实就是退化过程的逆过程, 找到合适的方法来降低或者消除图像降质过程的模糊和噪声, 并上采样获得像素量更大的图像素, 提高分辨率, 它的过程即可用式(2)来简单表示

$$I_{SR} = F(I_{LR}, \theta) \quad (2)$$

其中, F 表示一定的超分辨率方式, θ 表示相关的参数。而我们获得的 SR 图片, 他越接近于 HR 图片, 就说明重构的效果越好。

超分辨率技术的实现方式可以分为基于插值、基于重建和基于学习的方法, 目前最为常用、效果最好的则是通过卷积神经网络来实现的基于深度学习的超分辨率方法。

2.2. 基于卷积神经网络的超分辨率算法

基于卷积神经网络的超分辨率算法是一种端到端的方法, 通过中间层的处理, 将输入的低分辨率图像, 生成高分率的图像。SRCNN [5]首次使用全卷积神经网络进行了单图的超分辨率重构, 开辟了用深度卷积网络来解决超分辨率问题的先河, 当下的多数模型都是在此之上的扩展和延伸。

该模型首先通过双三次插值的方法, 将输入的 LR 图像放大到目标尺寸, 然后通过三层的卷积神经网络映射和重建, 得到目标的 HR 图片, 结构如图 1。

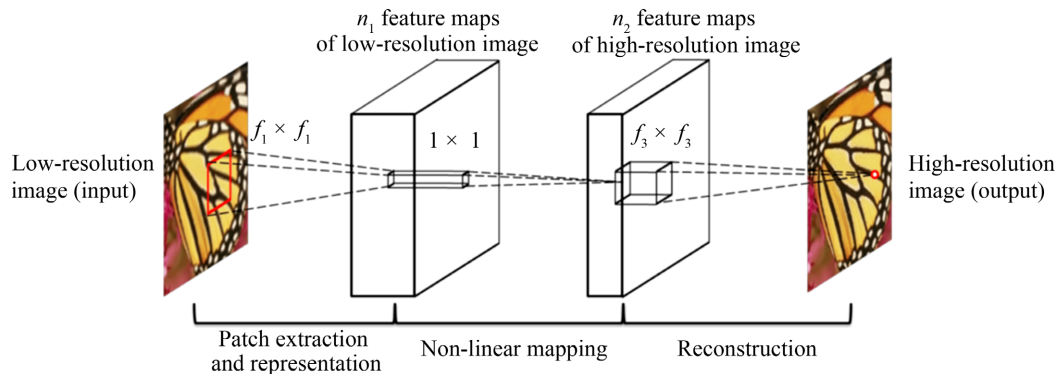


Figure 1. Diagram of SRCNN network structure

图 1. SRCNN 网络结构图

三个卷积层分别完成了特征块提取和表示、非线性映射和重构三个任务, 具体如下:

1) 特征块提取与表示: 将插值放大后的低分辨率图像提取 $f_1 * f_1$ 的图像块后输入卷积, 输出每个图像块的 n_1 维特征块。如式(3), 其中 $*$ 代表卷积操作, W_1 表示 n_1 个卷积核, B_1 为一个偏置向量。

$$F_1(Y) = \max(0, W_1 * Y + B_1) \quad (3)$$

2) 非线性映射: 将特征块输入第二层卷积, 通过非线性变化, 得到一个 n_2 维的矢量。如式(4), W_2 表示 n_2 个大小为 $n_1 * f_2 * f_2$ 个卷积核, B_2 为 n_2 维的向量。

$$F_2(Y) = \max(0, W_2 * F_1(Y) + B_2) \quad (4)$$

3) 重构:通过一个反卷积的操作,得到最终的高分辨率图像块。如式(5), W_3 表示 c 个大小为 $n_2 * f_3 * f_3$ 的滤波器, B_3 为 c 维的偏置向量。

$$F_3(Y) = W_3 * F_2(Y) + B_3 \quad (5)$$

该方法取得的效果显著优越于传统的超分辨率方法,在视觉效果和 PSNR (峰值信噪比, Peak Signal to Noise Ratio)值上都表现出了更好的性能,也成为了后续使用深度学习进行超分辨率研究的基础框架,本文的后续研究也是基于此框架进行的。

3. 基于对偶学习的残差注意力超分辨率算法

单图超分辨率的目的,是希望生成的 I_{SR} 能够尽量地逼近 I_{HR} , 我们将 I_{LR} 作为信息输入后,希望得到一个合适的解。但实际上这是一个不适定的问题,因为存在无数可能的 I_{HR} 在通过同一降采样手段后,会得到同一张的 I_{LR} 。所以在反向求解的过程中,就会存在解空间过大的问题,缺少约束条件,导致重构的图像尽管可能获得了较高的客观评价指标,但在视觉上却存在失真的问题。

本章通过引入对偶学习的框架,除了进行将 I_{LR} 生成 I_{SR} 的常规任务外,增加了将生成的 I_{SR} 再次下采样生成 I'_{LR} 的任务。两个任务互联关联,形成闭环的彼此监督,通过新增的对下采样的约束,进一步减小了 I_{SR} 的解空间范围,提高了生成图像的质量。

3.1. 任务框架

本文的对偶网络包含两个任务,一个是将 I_{LR} 生成 I_{SR} 的主任务,即常规 SR 任务,另一个是将生成的 I_{SR} 下采样为 I'_{LR} 的对偶任务。

训练的目标一方面希望生成的 I_{SR} 能尽可能的逼近 I_{HR} , 另一方面是希望 I_{SR} 的下采样结果 I'_{LR} 也能尽可能的逼近 I_{LR} , 用以对生成的 I_{SR} 做进一步的限制,减小解空间。因为如果生成的 I_{SR} 能接近于 I_{HR} , 那么他的下采样结果也理应和真实图像的下采样结果相似,由此将传统的 LR→HR 映射转变成了 LR←→HR 的闭环双向映射,以获得更好的 SR 效果。如图 2 所示为本网络的思路框架。

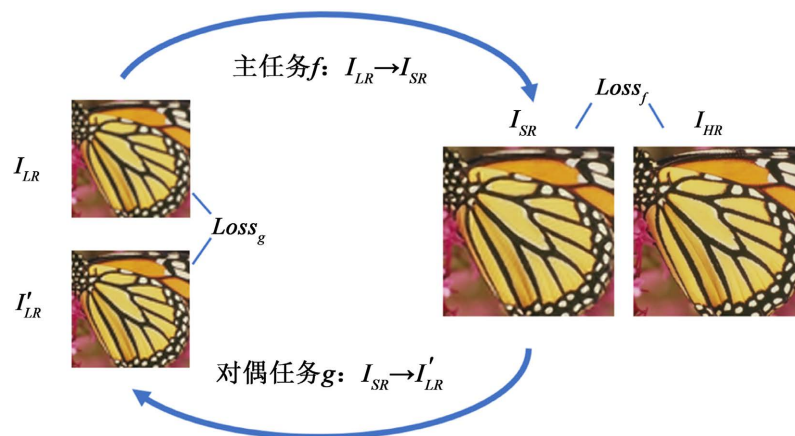


Figure 2. Diagram of dual task framework
图 2. 对偶任务框架图

3.2. 网络结构

3.2.1. 密集连接通道注意力模块

大多数基于深度学习的 SR 重构方法,在重构时忽略了分层特征的使用,而对于不同尺度的图像/特

征图来说, 特征的显著部分也存在着较大的差异。虽然使用残差块可以将原始 LR 图像的信息进行传递, 但是对于一个很深的网络来说, 在中间过程中的信息该如何利用依旧是一个问题。直接提取每个卷积层的输出很难, 为了更好地进行多层次信息的传递, 我将残差通道注意力模块(RCAB, Residual Channel Attention Block)进行了密集连接, 将每个层都与前面所有层在通道维度上连接在了一起, 作为下一层的输入, 以此构成了 Residual Dense Channel Attention Blocks (简称 RDCAB 模块)。比起简单的跳跃连接, 使用 concat 进行的密集连接, 将不同层的特征图进行了更好地复用, 使状态进行有效地传递, 让图像特征信息拥有更好的表征。如图 3 为 RDCAB 模块的结构图。

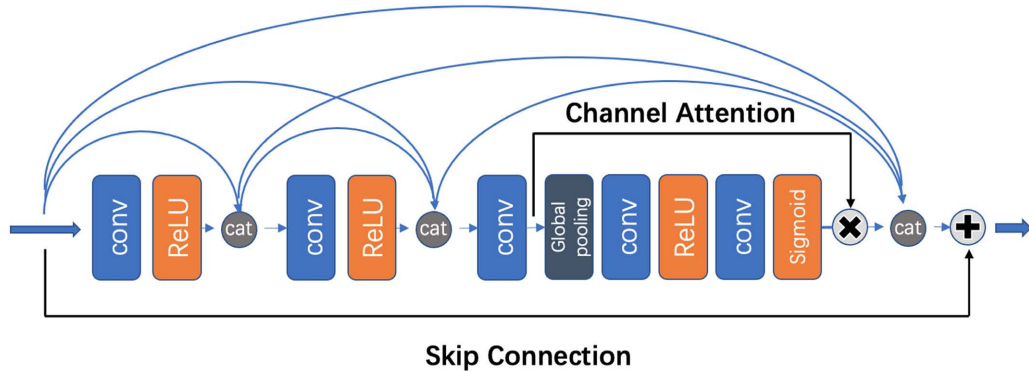


Figure 3. Structure diagram of RDCAB module
图 3. RDCAB 模块结构图

第 b 个 RDCAB 模块的输入为上一个模块输出的特征图 F_{b-1} , 经过 3×3 尺寸, 且步长、padding 皆为 1 的卷积核后, 特征图保持原有的尺寸不变, 通过 concat 进行信息的拼接。通过这一模块进一步利用不同层的特征, 可以提高模型学习的性能。

3.2.2. 模型细节

我们的网络借鉴了 U-Net [6] 结构, 模型包含编码和解码两个过程。 I_{LR} 首先通过编码层进行特征提取, 再通过解码层进行上采样, 重建获得 I_{SR} 。为了更好地传递不同层次的特征细节, 在阶段性的上下采样特征图中进行恒等映射, 使得同尺度的 I_{LR} 与 I_{SR} 中能够共享低频信息。如图 4, 为 $\times 4$ 尺寸的网络结构图。

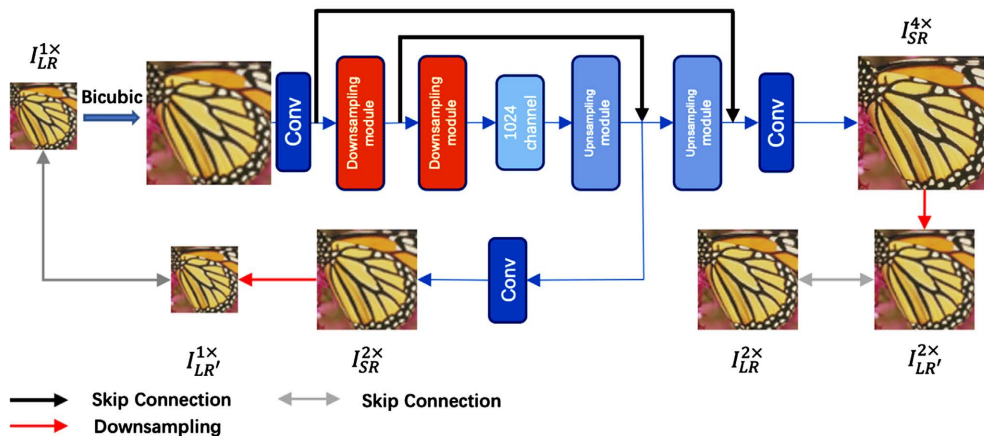


Figure 4. Diagram of network structure
图 4. 网络结构图

对于输入的 LR 图像，我们先通过双三次插值的方法，放大到目标尺寸。先经过一次卷积后，将通道数转化为 64，通过两个下采样模块，得到 1024 通道的特征图。下采样模块结构如图 5(a)，其中，先通过一个 size 为 3，步长为 2 的卷积，在进行特征提取的同时实现下采样，将 $c \times w \times h$ 的矩阵转化 $c_1 \times \frac{w}{2} \times \frac{h}{2}$ 的输出，其中 c_1 为卷积核的数目。然后通过激活函数实现非线性的变换，最后通过一个步长为 1 的常规卷积，得到下采样模块输出的特征图。下采样过程表现如式(6)所示：

$$F_n = h_{\text{down}}(F_{n-1}) \quad (6)$$

其中 F_n 表示第 n 个下采样模块输出的特征图， n 的数目根据放大尺寸而定， h_{down} 表示进行下采样模块操作，其中第一个模块的输入 F_0 为通过双三次插值放大后的 LR 图像。

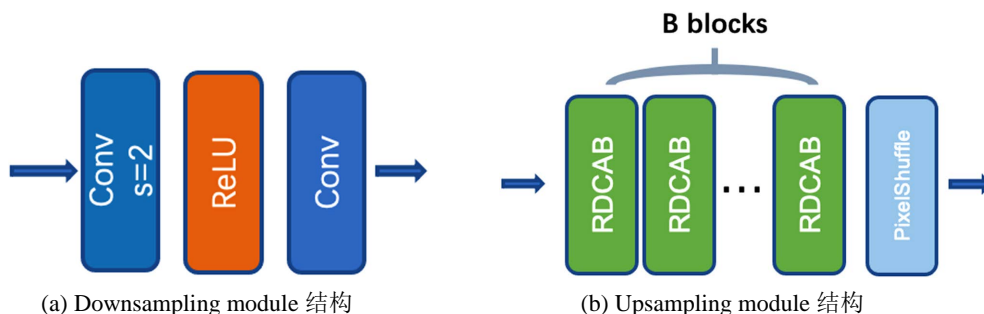


Figure 5. Structure diagram of downsampling/upsampling module
图 5. 下采样/上采样模块结构图

每个通过下采样模块的特征图，通过一个对称连接复制到上采样的结果中，进行特征信息的传递。上采样模块如图 5(b)，通过 B 个 RDCAB 模块进行连接后，使用 PixelShuffle 实现上采样，上采样模块的输出通过一个卷积层可以输出为对应尺寸 SR 图像。上下样过程表现如式(7)所示：

$$\begin{aligned} F_{2\times} &= h_{\text{up}}(F_2) + F_1 \\ F_{4\times} &= h_{\text{up}}(F_{2\times}) + F_0 \end{aligned} \quad (7)$$

其中， h_{up} 表示进行上采样模块操作， F_n 为从下采样过程中传递过来的特征信息。 $F_{2\times}$ 、 $F_{4\times}$ 为对应放大尺寸的特征图，通过一个 1×1 卷积层后可得到对应尺寸的 SR 图像，如(8)所示：

$$I_{\text{SR}}^{n\times} = f_n(F_{n\times}) \quad (8)$$

f_n 为对应层的卷积操作， $I_{\text{SR}}^{n\times}$ 为 $n \times$ 放大倍数的 SR 图像，即 $I_{\text{SR}}^{4\times}$ 即是 4 倍超分任务中最终需要生成的 SR 图像。在用以限制生成效果的对偶任务中，将生成的 $I_{\text{SR}}^{2\times}$ 、 $I_{\text{SR}}^{4\times}$ 进行 2 倍的下采样，获得原尺寸的 $I_{\text{LR}}^{1\times}$ 和 2 倍放大的 $I_{\text{LR}}^{2\times}$ ，用他们和数据集中对应尺寸的 LR 图像进行比对，计算损失函数值。

3.3. 损失函数

损失函数的作用是对网络模型的输出和真实值进行差异度的计算，从而判断两者间的偏离程度，从而引导网络优化的目标。如果损失函数计算得到的值较小，那么就代表在我们选择的判断尺度内，网络模型的生成性能较好。因此，损失函数的类型对于训练和生成效果都起到了重要的指导作用。

因为本模型的任务包含主任务和回归任务，所以损失函数由两个部分组成：

$$L_{\text{total}} = L_p(P(x), y) + \lambda L_d(D(P(x)), x) \quad (9)$$

其中, L_p 为主任务的损失, x 为低分辨率的图像, P 表示图像的 SR 过程, y 表示对应的 HR 图像; L_d 为对偶任务的损失, λ 用来权衡对偶任务损失占比的大小, D 表示对 SR 图像进行下采样的过程。

对于 loss 的形式, 我们在相同的条件下分别用 L1 loss 和 L2 loss 进行了训练, 达到收敛后进行测试, 效果图如下图 6。

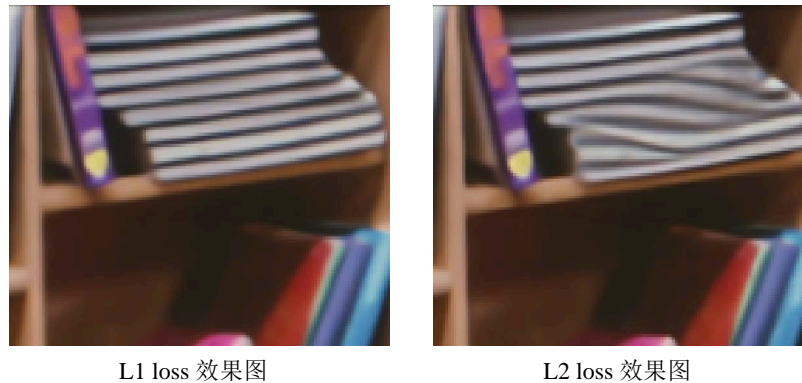


Figure 6. Effect picture of training with different loss functions
图 6. 不同损失函数训练的效果图

可以看到使用 L2 loss 生成的图像虽然整体上和 L1 loss 相似, 但对于纹理和边缘的处理, 明显存在模糊和失真的现象。

因此, 本文中使用 L1 loss 作为损失函数, 以产生更少的平滑现象。并且在 λ 的取值上, 我们使用了从 0.01 逐渐增大到 1 来进行实验, 在值为 0.1 的时候取得最好的效果。

4. 实验结果

本实验在一张 NVIDIA GeForce GTX-1080Ti GPU 上进行, 实验基于 Pytorch 深度学习框架实现。模型训练使用的数据集是 DIV2K [7]数据集, 这是目前较为流行的单图超分辨率数据集, 其中使用编号 1-800 的图片作为训练集, 801-900 的图像为测试集。选择的重构尺寸为 $\times 4$, 因此每张图片的输入都包含其 HR 图像和 2 倍、4 倍下采样的 LR 图像。

4.1. 对比实验

在视觉效果上, 以 Set14 [8]的 baboon 为例(如图 7), SRCNN 虽然整体上实现了超分重构, 但在结构复杂处还是呈现得相对模糊, 边缘平滑不清晰。而 EDSR 在此基础上有了较大的改进, 纹理边缘得到凸显, 但是部分区域仍呈现结构性不好的失真现象(如胡须呈现偏折、不连贯的现象), 同时它的模型体量也显著较大。本章提出的方法, 在细节处理上更胜于 EDSR, 胡须纹理边缘都较为清晰, 且视觉上连贯和谐、整体性较好。

如表 1 所示, 我们的方法在一系列经典超分辨率算法中取得了较好的效果, 其中加粗部分为最优指标。同时, 本文的模型参数(8.13 M)也少于其他表现优异的模型如 RCAN (15.6 M)、EDSR (43.1 M)。因此, 本模型可以在参数量相对较小的情况下, 实现较好的超分性能和视觉效果。

4.2. 消融实验

为了验证本文所使用的方法在模型优化中是切实有效的, 我们又进行了消融实验, 通过 RCAB/RD CAB 的替换以及去除对偶任务来进行。

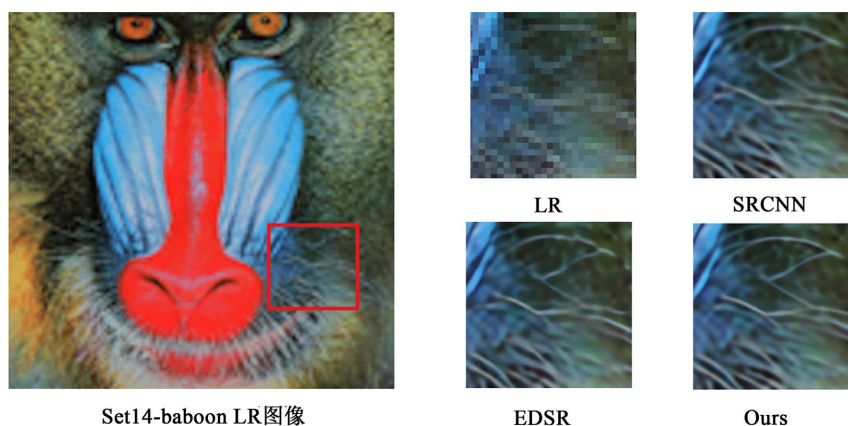


Figure 7. Visual comparison between our methods with other super-resolution methods

图 7. 本文方法与其他超分方法的效果图对比

Table 1. Performance comparison of super-resolution models at 4×

表 1. 4 倍超分下模型性能对比

Algorithms	Set5 [9] PSNR/SSIM	Set14 PSNR/SSIM	BSD100 [10] PSNR/SSIM	Urban100 [11] PSNR/SSIM
Bicubic	28.42/0.8104	26.00/0.7027	25.96/0.6675	23.14/0.6577
SRCNN	30.28/0.8628	27.49/0.7503	26.90/0.7101	24.52/0.7221
EDSR	32.48/0.8982	28.81/0.7873	27.72/0.7420	26.64/0.8029
SRResnet	32.05/0.8910	28.53/0.7804	27.57/0.7354	26.07/0.7839
RCAN	32.63/0.9001	28.85/0.7882	27.74/0.7429	26.74/0.8061
Ours	32.58/0.9001	28.86/0.7887	27.75/0.7431	26.67/0.8062

Table 2. PSNR value of super-resolution ablation experiment at 4×

表 2. 4 倍超分下消融实验 PSNR 值

	模型一	模型二	模型三	模型四
RDCAB	χ	χ	\sqrt	\sqrt
对偶任务	χ	\sqrt	χ	\sqrt
Set5	32.36	32.45	32.54	32.58
Set14	28.74	28.77	28.82	28.86
B100	27.66	27.68	27.71	27.75
Urban100	26.43	27.49	26.62	26.67

在无对偶任务的 RCAB 集联网络上,我们分别增加对偶任务/将 RCAB 替换成 RDCAB 来进行了实验,每个模型都是在训练至 300 个 Epoch 已收敛的情况下进行测试。结果如表 2 所示,证明两项改进对于网络的性能都有着较为显著的提升,我们的方法是有效的。

5. 总结

本文对超分辨率的概念及其在深度学习基础上的实现进行了阐述,并基于对偶学习,在深度学习主任务的基础上,通过下采样的子任务约束超分辨率图片的训练过程,提高生成效果的整体性;同时使用

了密集连接的通道注意力模块,使得网络的性能也得到提升。最后,通过模型对比实验和消融实验,验证了本方法的高效性。

参考文献

- [1] Zhao, Y., Yi, C., Yang, J., *et al.* (2014) Coupled Hyperspectral Super-Resolution and Unmixing. 2014 *IEEE Geoscience and Remote Sensing Symposium*, Quebec City, 13-18 July 2014, 2641-2644.
- [2] Nishibori, K., Takahashi, T., Deguchi, D., *et al.* (2014) Exemplar-Based Human Body Super-Resolution for Surveillance Camera Systems. *Computer Vision Theory and Applications*, **45**, 115-121.
- [3] Bahrami, K., Shi, F., Rekić, I., *et al.* (2017) 7T-Guided Super-Resolution of 3T MRI. *Medical Physics*, **44**, 1661-1677. <https://doi.org/10.1002/mp.12132>
- [4] Straub, J. (2012) Robotic Applications of a Defensible Error-Aware Super-Resolution Technique. *The Seventeenth International Symposium on Artificial Life and Robotics*, **54**, 1421-1422.
- [5] Dong, C., Chen, C.L., He, K., *et al.* (2016) Image Super-Resolution Using Deep Convolutional Networks. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, **38**, 295-307. <https://doi.org/10.1109/TPAMI.2015.2439281>
- [6] Ronneberger, O., Fischer, P. and Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W. and Frangi, A., Eds., *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*, Springer, Cham, 234-241. https://doi.org/10.1007/978-3-319-24574-4_28
- [7] Agustsson, E. and Timofte, R. (2017) NTIRE 2017 Challenge on Single Image Super-Resolution: Dataset and Study. 2017 *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Honolulu, 21-26 July 2017, 126-135. <https://doi.org/10.1109/CVPRW.2017.150>
- [8] Zeyde, R., Elad, M. and Protter, M. (2010) On Single Image Scale-Up Using Sparse-Representations. *Proceedings of the 7th International Conference on Curves and Surfaces*, Avignon, 24-30 June 2010, 711-730.
- [9] Bevilacqua, M., Roumy, A., Guillemot, C. and Morel, M.-L.A. (2012) Low-Complexity Single-Image Super-Resolution Based on Nonnegative Neighbor Embedding.
- [10] Huang, J., Singh, A. and Ahuja, N. (2015) Single Image Super-Resolution from Transformed Self-Exemplars. 2015 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, 7-12 June 2015, 5197-5206. <https://doi.org/10.1109/CVPR.2015.7299156>
- [11] Matin, D., Fowlkes, C., Tal, D., *et al.* (2001) A Database of Human Segmented Natural Images and Its Application to Evaluating Segmentation Algorithms and Measuring Ecological Statistics. *Proceedings Eighth IEEE International Conference on Computer Vision, ICCV 2001*, Vancouver, 7-14 July 2001, 416-423.