

基于Transformer和不可分加性小波的图像超分辨率重建

刘 斌, 杜丹丹

湖北大学计算机与信息工程学院, 湖北 武汉

收稿日期: 2022年12月12日; 录用日期: 2023年1月2日; 发布日期: 2023年1月17日

摘 要

针对目前超分辨率重建存在纹理模糊、扭曲等问题, 提出了一种基于Transformer和不可分加性小波的网络。该网络由小波分解模块、纹理提取模块、浅层特征提取模块、用于纹理匹配的相关嵌入模块、纹理传输模块、用于纹理融合的跨尺度集成模块共六个模块组成。我们对此网络的重建性能在相关的测试数据上与已有的典型方法进行了对比研究, 实验结果表明, 该网络不仅提升了视觉效果, 而且获得了较好的客观指标评价, 在所比较的六种方法中, 所构建网络在CUFED5、Sun80和Manga109数据集上的峰值信噪比(PSNR)和结构相似度(SSIM)都具有最佳性能。

关键词

Transformer, 不可分加性小波, 超分辨率重建, 计算机视觉, 深度学习, 注意力机制

Image Super-Resolution Reconstruction Based on Transformer and Non-Separable Additive Wavelet

Bin Liu, Dandan Du

School of Computer Science and Information Engineering, Hubei University, Wuhan Hubei

Received: Dec. 12th, 2022; accepted: Jan. 2nd, 2023; published: Jan. 17th, 2023

Abstract

Aiming at the problems of texture blur and distortion in super-resolution reconstruction, a net-

work based on Transformer and non-separable additive wavelet is proposed. The network consists of six modules: Wavelet Decomposition module, Texture Extraction module, Shallow Feature Extraction module, Relevance Embedding module for texture matching, Texture Transmission module, and Cross Scale Integration module for texture fusion. We compared the reconstruction performance of this network with the existing typical methods on the relevant test data. The experimental results show that this network not only improves the visual effect, but also obtains better objective index evaluation. Among the six methods compared, the peak signal to noise ratio (PSNR) and structure similarity (SSIM) of the constructed network on CUFED5, Sun80 and Manga109 datasets have the best performance.

Keywords

Transformer, Non-Separable Additive Wavelet, Super-Resolution Reconstruction, Computer Vision, Deep Learning, Attention Mechanism

Copyright © 2023 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

图像超分辨率(Super Resolution, SR)重建是在低分辨率(Low Resolution, LR)图像的基础上, 获得高分辨率(High Resolution, HR)图像, 属于计算机视觉领域的重要分支之一, 可以为目标检测、图像分类等任务提供服务, 以提高准确性和正确率。图像超分辨率重建不仅可以应用于医学、遥感等领域, 而且让图像分辨率不再受限于昂贵的硬件条件。

传统的超分辨率重建分为三类: ① 基于插值法的超分辨率重建。② 基于重建法的超分辨率重建。③ 基于学习法的超分辨率重建, 这里的学习是浅层学习。主要包括机器学习、稀疏编码等方法。近年来, 由于深度学习的快速发展, 浅层学习法中的机器学习逐渐演变成基于深度学习的超分辨率重建。

2014年, 香港中文大学的团队提出的 SRCNN [1]是首次使用深度学习来进行图像重建的算法, 之后该团队又提出了 FSRCNN [2]。Jiwon Kim 等人也改进了 SRCNN, 提出了 VDSR [3]和 DRCN [4]。2017年, Christian Ledig [5]等人提出了 SRResNet 和 SRGAN, 使用了带有跳跃连接的深度残差网络(ResNet)。而且 SRGAN 里首次将生成对抗网络(Generative Adversarial Networks, GAN)应用于 SR 领域, 并将感知损失函数[6] (Perceptual Loss)表示为内容损失(Content Loss)和对抗性损失(Adversarial Loss)的加权和。韩国首尔国立大学的团队以 SRResGAN 为基础, 提出了 EDSR [7], 改进了残差单元, 删除了 BN 层。Wang 等人则在 SRGAN 和 EDSR 的基础上, 提出了 ESRCGAN [8], 在网络结构上使用了 RRDB 作为基本构建模块。以上方法都是单图像超分辨率重建(Single Image Super-Resolution, SISR), 传统的 SISR 方法通常会会导致模糊效果, 因为高分辨率纹理在退化到低分辨率图像过程中被过度破坏, 无法恢复。

而在基于参考的图像超分辨率重建(Reference-based Image Super-Resolution, RefSR)方面, 郑等人先利用块匹配提出了 SS-Net [9]。之后郑等人又提出了 CrossNet [10], 使用了编码器、跨尺度扭曲、解码器, 解决了块间不匹配、网格效应和优化不高效的问题。为了增强模型的鲁棒性, Zhang 等人提出了 SRNTT [11]。SRNTT 对预训练的 VGG 中提取的参考图像纹理和 LR 图像纹理进行多级匹配, 然后根据纹理相似

度自适应地完成纹理转换, 这样即使在参考图像和 LR 图像相似度极低时, 模型也与 SISR 性能相等。2020 年, Yang 等人首次将 Transformer 架构引入 RefSR 任务, 提出了 TTSR [12]模型。

但是有些模型在恢复纹理过程中, 会出现伪影、模糊、失真等问题, 针对这些问题, 本文提出了一种基于 Transformer 和不可分加性小波模型(Image Super-Resolution by Transformer and Non-Separable Additive Wavelet, 简称为 SRTW)。

2. 相关知识

2.1. Transformer 中的注意力机制

Transformer [13]是谷歌在 2017 年提出的一种网络结构, 它属于 Seq2Seq 模型, 但不同于 RNN、CNN, 而是完全采用自注意力机制, 如图 1 所示。应用注意力机制时, Transformer 首先通过矩阵运算将输入映射至三个关键向量 Q (Query, 查询)、 K (Key, 键)、 V (Value, 值), 通过将 Q 向量与要打分的 K 向量的点积相乘来计算关联度分数。为了避免点积之后获得太大的值, 需要对点积之后的矩阵进行缩放, 并通过 softmax 激活函数将数据归一化。然后将每个 V 向量乘 softmax 分数, 最后对加权向量进行求和, 得到自注意层的输出。

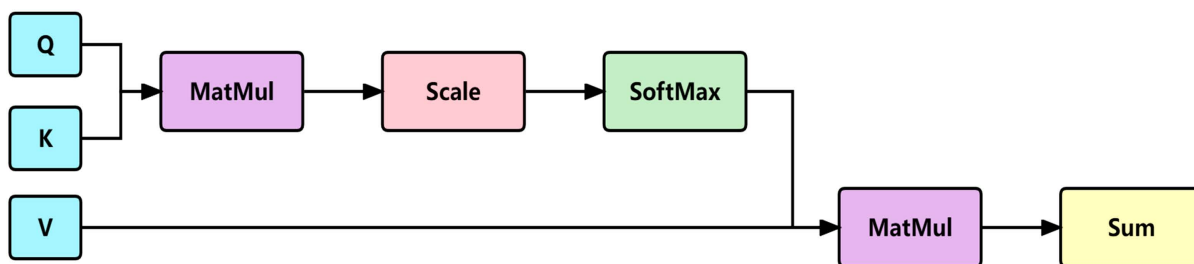


Figure 1. Attention mechanism in transformer

图 1. Transformer 中的注意力机制

2.2. 不可分小波

图像是二维的, 但传统的张量积小波只是用一维小波分别对图像的行和列进行处理, 并没有做到行列同时处理。而不可分小波在能对图像行、列一起处理的同时, 也解决了张量积小波不具有对称性(Haar 小波除外)所带来的边缘失真问题。

Chen [14]等人提出了高维具有紧支撑性和正交性的非张量积小波滤波器组的一般构造方法, 我们在此基础上构造二维二通道滤波器组[15], 设伸缩矩阵为 $\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, s = 2$, 构造 $X(x, y) = (1, xy)$,

$D(x, y) = \begin{bmatrix} 1 & 0 \\ 0 & xy \end{bmatrix}$, 那么滤波器组可表示为:

$$(m_0(x, y), m_1(x, y)) = \frac{1}{2}(1, xy) \prod_{j=1}^k (U_j D(x, y) U_j^T) V \tag{1}$$

其中 $x = \exp(-i\omega_1), y = \exp(-i\omega_2), U_j (j = 1, 2, \dots, K), V/\sqrt{2} = (V_0, V_1)/\sqrt{2}$ 均是正交阵, V_1 是 2×1 的向量, $V_0 = (1, 1)^T$ 。

为构造二通道 6×6 的小波滤波器组, 取 $K = 4, U_j$ 和 V 如式(2):

$$\begin{aligned}
 U_1 &= \begin{pmatrix} \cos(\alpha_1) & -\sin(\alpha_1) \\ \sin(\alpha_1) & \cos(\alpha_1) \end{pmatrix}, \\
 U_2 &= \begin{pmatrix} \cos(\alpha_2) & \sin(\alpha_2) \\ -\sin(\alpha_2) & \cos(\alpha_2) \end{pmatrix}, \\
 U_3 &= \begin{pmatrix} -\sin(\alpha_3) & -\cos(\alpha_3) \\ \cos(\alpha_3) & -\sin(\alpha_3) \end{pmatrix}, \\
 U_4 &= \begin{pmatrix} \sin(\alpha_4) & \cos(\alpha_4) \\ -\cos(\alpha_4) & \sin(\alpha_4) \end{pmatrix}, \\
 V &= \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}
 \end{aligned} \tag{2}$$

可验证 $U_j (j=1,2,3,4)$ 均为正交矩阵, $V/\sqrt{2}$ 为正交阵。

为构造具有对称的滤波器组, 取 $\alpha_1 = \frac{3\pi}{4}, \alpha_2 = \frac{\pi}{2}, \alpha_3 = -\frac{\pi}{4}, \alpha_4 = -\frac{\pi}{2}$, 代入式(1)后, 可得 $m_0(x, y)$ 的时域形式:

$$H_0 = \frac{1}{8} \begin{bmatrix} -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix} \tag{3}$$

可以验证, 此滤波器不能分解为两个一维滤波器的张量积, 因此是二通道对称不可分小波滤波器。

3. 方法

3.1. 整体结构

SRTW 的网络结构如图 2 所示。输入为 LR、Ref、LR↑、Ref↑↓四种图像, LR 是低分辨率图像, Ref 是参考图像, LR↑是 LR 图像经过 4 倍双三次插值上采样的图像, Ref↑↓是 Ref 图像经过 4 倍双三次插值上下采样的图像, 目的是和 LR↑图像保持域一致[12] (即都是经过双三次变换得到的)。

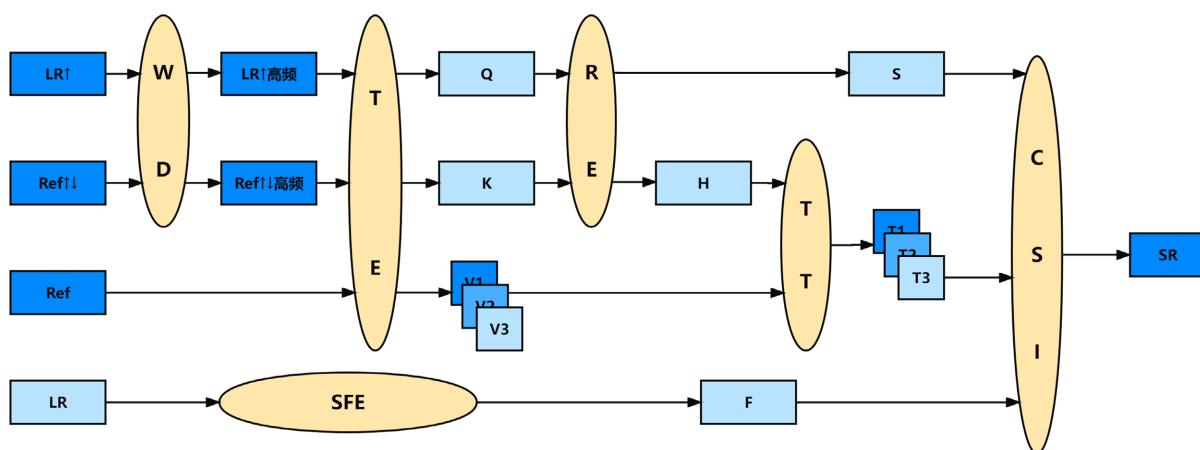


Figure 2. SRTW network structure

图 2. SRTW 网络结构

LR↑、Ref↑↓先通过小波分解模块(WD)得到 LR↑的高频、Ref↑↓的高频, 再和 Ref 一起送入纹理提取模块(TE), 得到向量 Q 、 K 、 $V(1\times, 2\times, 4\times)$ 。 Q 、 K 通过相关嵌入模块(RE)得到 Q 、 K 最相关位置 H 和值 S 。通过纹理传输模块(TT)对 V 取 H 位置得到 $T(1\times, 2\times, 4\times)$ 。 T 、 S 和 LR 图像经过浅层特征提取(SFE)得到的特征图 F 一起送入跨尺度集成模块(CSI), 将不同尺度的图像进行融合, 得到最终的超分辨率图像。

3.2. 小波分解模块(WD)

Nunez [16]等人在 1999 年提出了加性小波, 不需要高通滤波器, 只需要低通滤波器 H_0 对图像进行卷积得到低频信息 p_i , 而高频信息 w_i 通过原图像减去低频信息得到。设 p_0 为原始图像, 分解公式为:

$$\begin{aligned} H_0(p_0) &= p_1, w_1 = p_0 - p_1, \\ H_0(p_1) &= p_2, w_2 = p_1 - p_2, \\ H_0(p_2) &= p_3, w_3 = p_2 - p_3, \\ &\vdots \end{aligned} \tag{4}$$

将各级高频信息和最后的低频信息相加, 即可重构原图像。重构公式为:

$$p_0 = \sum_{i=1}^n w_i + p_r \tag{5}$$

其中 p_r 是最后一层分解的低频图像。在小波分解模块中, 将 LR↑、Ref↑↓分别进行一层分解, 输出相对应的 w_1 。

3.3. 纹理提取模块(TE)

利用了迁移学习, 将 VGG19 [17]的前 12 层作为纹理提取模块的网络结构, 并将参数初始化为已经训练好的 VGG19 参数, 在训练过程中, 该模块参数也一起训练。为了减少计算量, 没有将 Q 和 K 像 V 一样分三次输出。该模块结构如下图 3 所示:

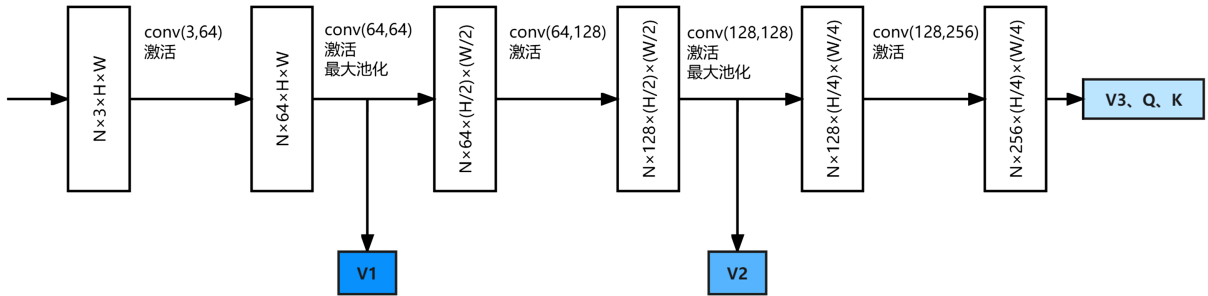


Figure 3. TE network structure
图 3. TE 网络结构

3.4. 浅层特征提取模块(SFE)

受到 ResBlock + CBAM [18]的启发, SFE 网络也采取了通道注意力和空间注意力两者兼顾的方法, 同时和残差网络相结合。由于在 EDSR [7]的文章中, 已经证实 BN 层不仅会拉伸图像本身的色彩、对比度, 使得输出图像产生不可预测的伪影, 而且去掉 BN 层节约下来的空间可用于插入更多残差块来增加模型的表现力, 因此本文删除了 BN 层, 同时多次重复该模块。该模块结构如下图 4 所示:

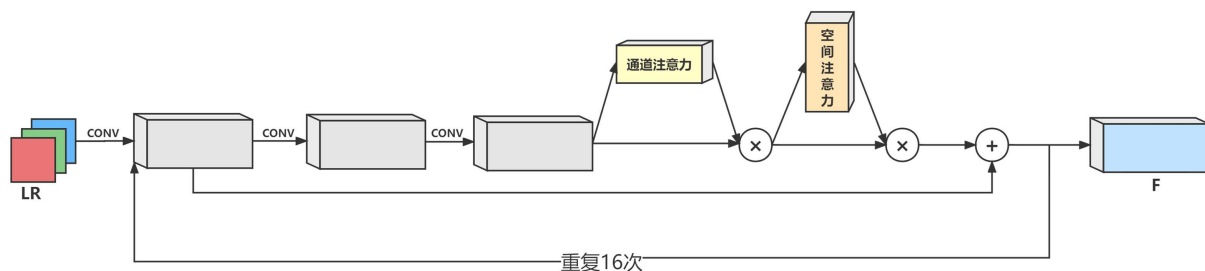


Figure 4. SFE network structure

图 4. SFE 网络结构

3.5. 相关嵌入模块(RE)

相关嵌入模块用 Q 、 K 的归一化内积来表示 Q 、 K 之间的相关性, 如式(6)所示, $R_{i,j}$ 大小为 $[N, (H_{Ref}/4) \times (W_{Ref}/4), (H_{LR}/4) \times (W_{LR}/4)]$ 。

$$r_{i,j} = \left\langle \frac{q_i}{\|q_i\|}, \frac{k_j}{\|k_j\|} \right\rangle, i \in \left[1, \frac{H_{LR}}{4} \times \frac{W_{LR}}{4} \right], j \in \left[1, \frac{H_{Ref}}{4} \times \frac{W_{Ref}}{4} \right] \quad (6)$$

取相关性最大的值和对应下标, 如式(7)、(8)所示, 分别组成软注意力图 S 和硬注意力图 H , 大小均为 $[N, (H_{LR}/4) \times (W_{LR}/4)]$ 。

$$h_i = \arg \max_j (r_{i,j}) \quad (7)$$

$$s_i = \max_j (r_{i,j}) \quad (8)$$

3.6. 纹理传输模块(TT)

纹理传输模块用于纹理传输, 得到 Ref 图像的不同尺度纹理特征图 H 位置的纹理特征, 如式(9)所示, 其中 t_i 是 T 的第 i 个位置, 由 V 的第 h_i 个位置得到。

$$t_i = v_{h_i} \quad (9)$$

3.7. 跨尺度集成模块(CSI)

跨尺度特征模块首先将 Ref 图像的纹理特征融合进 LR 图像中, 然后将不同尺度的图像集成在一起, 该模块结构如下图 5 所示。

融合过程如式(10)、(11)、(12)所示, 其中 concat 表示串联, conv 表示卷积, \otimes 表示特征图的元素相乘。

$$x_{11} = F + \text{conv}(\text{concat}(F, T_3)) \otimes S_{1X} \quad (10)$$

$$x_{22} = x_{21} + \text{conv}(\text{concat}(x_{21}, T_2)) \otimes S_{2X} \quad (11)$$

$$x_{33} = x_{32} + \text{conv}(\text{concat}(x_{32}, T_1)) \otimes S_{4X} \quad (12)$$

之后通过亚像素卷积[19]来进行小尺度到大尺度的上采样, 并和大尺度的特征在通道维度上进行串联, 然后通过卷积映射成原始的通道数。整个过程中配合着残差块, 加强原始信息。

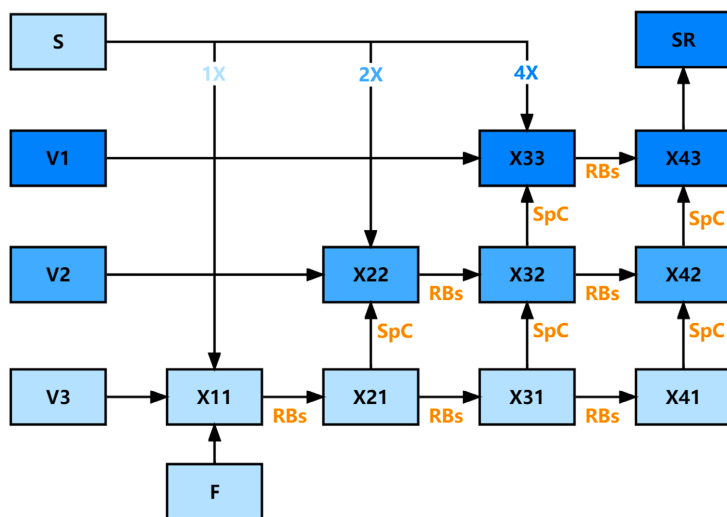


Figure 5. CSI network structure
图 5. CSI 网络结构

3.8. 损失函数

本文采用重建损失作为损失函数:

$$L_{rec} = \frac{1}{CHW} \|I^{HR} - I^{SR}\|_1 \quad (13)$$

其中 C 、 H 、 W 分别是 HR 图像的通道数、高、宽。在 L1 损失和 L2 损失中, 由于 L1 损失性能更高, 容易收敛[12], 因此本文选择了 L1 损失。

4. 实验与结果

本文使用一块 16GB 的 GPU (NVIDIA Quadro TRX 5000), 并在基于 GPU 版本的 Pytorch1.11.0 框架下训练模型。在训练过程中, 采用 Adam [20]对网络进行优化, 设置批数量大小是 16, 初始学习率为 10^{-4} , 训练 200 次后学习率减半, 共训练 300 次。

4.1. 数据集

本文在 RefSR 基准数据集 CUFED5 [11]上训练并测试了模型。CUFED5 的训练集有 11,871 对, 每对由输入图像和参考图像组成, 测试集有 126 幅测试图像, 每幅图像都有从高到低的多个相似度级别的参考图像。为了评估在 CUFED5 上训练的 SRTW 的泛化性能, 本文另外在 Sun80 和 Manga109 上进行了测试。对于 Sun80 数据集, 将其 LR 图像视为参考图像。对于同样缺少参考图像的 Manga109 数据集, 随机抽取该数据集中的 HR 图像作为参考图像。

在训练过程中, 通过随机水平和垂直翻转, 随机旋转 90° 、 180° 、 270° 来增强训练图像。每个小批量包含 16 个大小为 40×40 的 LR 图像, 以及 16 个大小为 160×160 的 HR 和 Ref 图像。

4.2. 结果分析

为了评估本文模型的有效性, 将本文的 SRTW 模型与其他 SISR 和 RefSR 方法进行比较。SISR 方法包括 Nearest、Bicubic、SRCNN [1]、SRGAN [5]。RefSR 方法包括 SRNTT [11]。所有实验均按 4 倍比例放大系数来对 LR 和 HR 图像进行定量评估。为了公平比较每个模型性能的优劣, 按照 SRNTT 中的设置

来训练所有方法。对 SRTW 的定量评估与定性评估结果如表 1 所示。通过表 1 可以看出, SRTW 在 CUFED5、Sun80 和 Manga109 数据集上都具有最佳性能。

Table 1. PSNR/SSIM comparison of different algorithms on three different datasets
表 1. 在三个不同数据集上不同算法的 PSNR/SSIM 比较

方法	CUFED5	Sun80	Manga109
Nearest	20.93/0.667	25.71/0.758	22.31/0.793
Bicubic	24.18/0.684	27.24/0.739	23.36/0.790
SRCNN	25.33/0.745	28.26/0.781	27.12/0.850
SRGAN	24.40/0.702	26.76/0.725	25.12/0.802
SRNTT	25.61/0.764	27.59/0.756	27.54/0.862
SRNTT-rec	<u>26.24/0.784</u>	<u>28.54/0.793</u>	<u>28.95/0.885</u>
SRTW	26.84/0.795	29.68/0.805	29.28/0.900

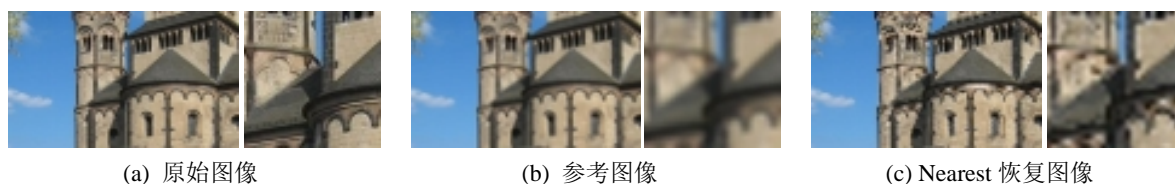
注: 加粗部分代表第一高分, 下划线部分代表第二高分。

图 6~8 分别展示了对 CUFED5、Sun80、Manga109 数据集所选图片使用 Nearest、Bicubic、SRCNN、SRGAN、SRNTT、SRNTT-rec 和 SRTW 方法进行超分辨重建的结果。(SRNTT 使用了可以提高视觉质量但降低 PSNR 的对抗性损失, 在 PSNR/SSIM 上会弱于 SRNTT-rec, 但会获得更高的视觉质量[11].)



Figure 6. Comparison of reconstruction results of different models on CUFED5: 007 image

图 6. 在 CUFED5: 007 图像上不同模型重建结果对比



(a) 原始图像

(b) 参考图像

(c) Nearest 恢复图像

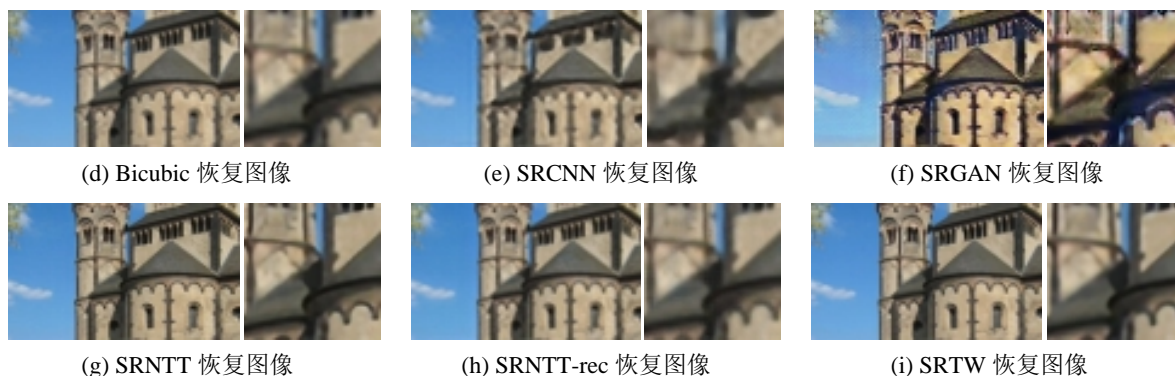


Figure 7. Comparison of reconstruction results of different models on SUN80: 001 image

图 7. 在 SUN80: 001 图像上不同模型重建结果对比



Figure 8. Comparison of reconstruction results of different models on MANGA109: Jyovolley image

图 8. 在 MANGA109: Jyovolley 图像上不同模型重建结果对比

图 6 中, Nearest 恢复图像很模糊, 有一根蜡烛只恢复了一部分, Bicubic 恢复图像同样模糊, SRGAN 恢复图像不仅模糊, 还有许多伪影。SRNTT 恢复图像和 SRNTT-rec 恢复图像里都有一根蜡烛存在伪影, 且 SRNTT-rec 恢复图像里花纹更加模糊。SRCNN 恢复图像整体不错, 但是 SRTW 恢复图像更加清晰, 尤其是蜡烛部分。

图 7 中, Nearest 恢复图像有明显的锯齿感, SRGAN 恢复图像颜色失真, SRCNN 恢复图像十分模糊, SRNTT 恢复图像左侧线条有些扭曲, Bicubic 恢复图像和 SRNTT-rec 恢复图像整体不错, 但 SRTW 恢复图像的线条更加深刻。

图 8 中, Nearest 恢复图像也有明显的锯齿感, SRGAN 恢复图像颜色失真, SRCNN 恢复图像和 Bicubic 恢复图像有不同程度的模糊。SRNTT 恢复图像和 SRNTT-rec 恢复图像整体不错, 但是 SRTW 恢复图像的线条更加深刻。

5. 结论

本文提出了一种新的基于参考的图像超分辨率重建模型来生成高分辨率图像。结合不可分加性小波

提升 Transformer 对高频信息的注意力, 对细节进一步强调, 浅层特征提取模块加强了原图像信息, 跨尺度集成模块使得不同尺度的特征可以交换信息, 获得更强大的特征表示能力。实验结果表明, 所提 SRTW 在主观视觉与客观指标上都有所优化, 即使在参考图像就是 LR 图像(变成单图像超分辨率重建)或参考图像与真实图像相关性很低的情况下, SRTW 也能获得更清晰、准确的重建图像。

基金项目

国家自然科学基金面上项目(No. 61471160)。

参考文献

- [1] Dong, C., Loy, C.C., He, K., *et al.* (2014) Learning a Deep Convolutional Network for Image Super-Resolution. *European Conference on Computer Vision*, Zurich, 6-12 September 2014, 184-199. https://doi.org/10.1007/978-3-319-10593-2_13
- [2] Dong, C., Loy, C.C. and Tang, X. (2016) Accelerating the Super-Resolution Convolutional Neural Network. *European Conference on Computer Vision*, Amsterdam, 11-14 October 2016, 391-407. https://doi.org/10.1007/978-3-319-46475-6_25
- [3] Kim, J., Lee, J.K. and Lee, K.M. (2016) Accurate Image Super-Resolution Using Very Deep Convolutional Networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 27-30 June 2016, 1646-1654. <https://doi.org/10.1109/CVPR.2016.182>
- [4] Kim, J., Lee, J.K. and Lee, K.M. (2016) Deeply-Recursive Convolutional Network for Image Super-Resolution. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 27-30 June 2016, 1637-1645. <https://doi.org/10.1109/CVPR.2016.181>
- [5] Ledig, C., Theis, L., Huszár, F., *et al.* (2017) Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, 21-26 July 2017, 4681-4690. <https://doi.org/10.1109/CVPR.2017.19>
- [6] Johnson, J., Alahi, A., *et al.* (2016) Perceptual Losses for Real-Time Style Transfer and Super-Resolution. *European Conference on Computer Vision*, Amsterdam, 11-14 October 2016, 694-711. https://doi.org/10.1007/978-3-319-46475-6_43
- [7] Lim, B., Son, S., Kim, H., *et al.* (2017) Enhanced Deep Residual Networks for Single Image Super-Resolution. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, Honolulu, 21-26 July 2017, 136-144. <https://doi.org/10.1109/CVPRW.2017.151>
- [8] Wang, X., Yu, K., Wu, S., *et al.* (2018) Esrgan: Enhanced Super-Resolution Generative Adversarial Networks. *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, Munich, 8-14 September 2018, 63-79. https://doi.org/10.1007/978-3-030-11021-5_5
- [9] Zheng, H., Ji, M., Han, L., *et al.* (2017) Learning Cross-Scale Correspondence and Patch-Based Synthesis for Reference-Based Super-Resolution. *Proceedings of the British Machine Vision Conference*, London, 4-7 September 2017, Article No. 138. <https://doi.org/10.5244/C.31.138>
- [10] Zheng, H., Ji, M., Wang, H., *et al.* (2018) Crossnet: An End-to-End Reference-Based Super Resolution Network Using Cross-Scale Warping. *Proceedings of the European Conference on Computer Vision (ECCV)*, Munich, 8-14 September 2018, 88-104. https://doi.org/10.1007/978-3-030-01231-1_6
- [11] Zhang, Z., Wang, Z., Lin, Z., *et al.* (2019) Image Super-Resolution by Neural Texture Transfer. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, 15-20 June 2019, 7982-7991. <https://doi.org/10.1109/CVPR.2019.00817>
- [12] Yang, F., Yang, H., Fu, J., *et al.* (2020) Learning Texture Transformer Network for Image Super-Resolution. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, 13-19 June 2020, 5791-5800. <https://doi.org/10.1109/CVPR42600.2020.00583>
- [13] Vaswani, A., Shazeer, N., Parmar, N., *et al.* (2017) Attention Is All You Need. *31st Conference on Neural Information Processing Systems (NIPS 2017)*, Long Beach, 4-9 December 2017, 30.
- [14] Chen, Q.H., Micchelli, C.A., Peng, S.L., *et al.* (2003) Multivariate Filter Banks Having Matrix Factorizations. *SIAM Journal on Matrix Analysis and Applications*, **25**, 517-531.
- [15] 刘斌, 彭嘉雄. 基于二通道不可分加性小波的多光谱图像融合[J]. 光学学报, 2007(8): 1419-1424.
- [16] Nunez, J., Otazu, X., Fors, O., *et al.* (1999) Multiresolution-Based Image Fusion with Additive Wavelet Decomposi-

- tion. *IEEE Transactions on Geoscience and Remote Sensing*, **37**, 1204-1211. <https://doi.org/10.1109/36.763274>
- [17] Simonyan, K. and Zisserman, A. (2014) Very Deep Convolutional Networks for Large-Scale Image Recognition. *3rd International Conference on Learning Representations (ICLR 2015)*, San Diego, 7-9 May 2015, 1-14.
- [18] Woo, S., Park, J., Lee, J.Y., *et al.* (2018) Cbam: Convolutional Block Attention Module. *Proceedings of the European Conference on Computer Vision (ECCV)*, Munich, 8-14 September 2018, 3-19. https://doi.org/10.1007/978-3-030-01234-2_1
- [19] Shi, W., Caballero, J., Huszár, F., *et al.* (2016) Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 27-30 June 2016, 1874-1883. <https://doi.org/10.1109/CVPR.2016.207>
- [20] Kingma, D.P. and Ba, J. (2014) Adam: A Method for Stochastic Optimization. arXiv: 1412.6980.