

# 一种结合自注意力和门控机制的图像超分辨率重建算法

李颖华, 赵春娜\*, 蒋慕蓉

云南大学信息学院, 云南 昆明  
Email: yinghua\_lee96@163.com, \*chunnazhao@163.com

收稿日期: 2020年11月27日; 录用日期: 2020年12月21日; 发布日期: 2020年12月28日

## 摘要

图像超分辨率重建旨在将低分辨率图像重建为更加清晰的高分辨率图像。超分辨率重建算法有助于提高图像质量, 可以尽可能精确地恢复出原始图像缺失的纹理、细节信息, 在图像处理领域具有重要的科学意义和应用价值。为了进一步提高图像重建质量, 本文将稀疏表示以及深度学习算法相结合, 利用稀疏表示模型得到的重构高分辨率图像作为深度学习模型的输入, 在VDSR网络的基础上减少卷积层并引入自注意力机制以及门控机制, 模型可以在训练过程中动态学习到不同特征的重要性, 从而进一步丰富图像的特征。我们在Set5、set14、B100、Urban100等公开的超分重建数据集上进行了大量的实验, 结果表明, 本文提出的基于自注意力机制和门控机制残差网络图像超分辨率重建算法相较于现有的重建方法, 可以获得更好的重建细节以及更高的PSNR/SSIM值。

## 关键词

图像超分辨率重建, 残差网络, 自注意力机制, 门控机制, 特征提取

# An Image Super-Resolution Reconstruction Algorithm Combining Self-Attention and Gating Mechanism

Yinghua Li, Chunna Zhao\*, Murong Jiang

School of Information, Yunnan University, Kunming Yunnan  
Email: yinghua\_lee96@163.com, \*chunnazhao@163.com

Received: Nov. 27<sup>th</sup>, 2020; accepted: Dec. 21<sup>st</sup>, 2020; published: Dec. 28<sup>th</sup>, 2020

\*通讯作者。

文章引用: 李颖华, 赵春娜, 蒋慕蓉. 一种结合自注意力和门控机制的图像超分辨率重建算法[J]. 计算机科学与应用, 2020, 10(12): 2323-2330. DOI: 10.12677/csa.2020.1012245

## Abstract

Image super resolution reconstruction aims to reconstruct a low resolution image into a clearer high-resolution image. The super resolution reconstruction algorithm is helpful to improve the image quality and can recover the missing texture and detail information as accurately as possible. It has important scientific significance and application value in the field of image processing. In order to further improve the quality of image reconstruction, this paper combines the sparse representation and deep learning algorithm. The reconstruction of the sparse representation model is used to get the high resolution image as input of deep learning model, and on the basis of introducing the VDSR network since attention mechanism and gating mechanism, models can be dynamically in the process of training to learn the importance of different characteristics. Thus, the pixel size and characteristics of granularity further enrich the characteristics of the image. We carried out a large number of experiments on the public super-fractional reconstruction data sets, such as Set5, SET14, B100 and Urban100. The results show that the multi-granularity feature extraction reconstruction algorithm proposed in this paper can obtain better reconstruction details and higher PSNR/SSIM values compared with the existing reconstruction methods.

## Keywords

Image Super Resolution Reconstruction, Residual Network, Self Attention Mechanism, Gating Mechanism, Feature Extraction

Copyright © 2020 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

现如今为止, 图像超分辨率重建主要有三种方法, 分别是基于插值、基于重建和基于学习。基于插值[1] [2]的算法重建速度比较快, 但是重建后的图像太过平滑并且图像边缘会产生锯齿效应。基于重建的算法效果较插值法有所改善; 基于学习的算法通过学习高、低分辨率图像块之间的对应关系来重建高分辨率图像, 较之前的算法获得了更好的图像重建质量。

基于学习的方法分为传统方法和基于深度学习的方法。在传统方法中, 稀疏表示则可以取得较好的重建效果, 典型代表是 Yang 等[3]提出的 SCSR 算法, 该算法不仅可以改善图像细节信息, 同时还能够保持图像的几何结构信息, 从而得到更好的重建图像。后来, 许多研究者对 SCSR 进行了改进[4] [5] [6], 通过对字典学习算法或者特征提取算子进行改进, 但由于基于传统学习模型的学习能力有限, 重建图像的高频细节不够丰富。为了弥补传统学习模型的不足, 提出了一些深度学习模型[7] [8] [9], 并在超分辨率图像中得到了广泛的应用。

目前, 一些基于卷积神经网络的深度学习方法应用在图像超分辨重建领域。Dong [10]最早提出了超分辨率 SRCNN 算法, 该算法建立了 CNN 的端到端映射, 该算法虽然结构简单, 但是收敛速度慢、学习信息有限, 难以取得很好的重建结果。之后, Dong [11]等人又提出了 FSRCNN, 此算法是基于 SRCNN 进行改进, 用一些小的卷积层来代替 SRCNN 中大的卷积层, 在速度和重建质量上有了进一步提高。VDSR [12]方法受到残差网络[13]结构的启发, 从而进行显著的改进。VDSR 采用学习经过双三次插值后的 LR 图像与高分辨率图像之间的高频残差图像来进行图像超分辨率重建。这些方法的出现弥补了稀疏表示的

不足，增强了重建图像的高频细节，但是在特征提取方面，VDSR 对提取的特征信息平等对待，没有重点关注图像重要的特征信息。此外，现有工作要么是只利用了稀疏表示，要么是用了深度学习，缺乏将二者进行结合的相关工作。

为了解决上述提出的问题，本文将 SCSR 引入到图像输入阶段，在 VDSR 的基础上引入了自注意力机制与门控机制，提取图像重要特征，提高恢复特征细节的能力，并采用 set5、set14、B100、urban100 等数据集进行重建，验证方法的有效性。

论文结构如下：第二部分，将本文改进的算法进行描述；在第三部分则是对实验结果进行比较与分析；第四部分是对本文的总结。通过实验表明，本文算法的重建质量与重建效率较之前算法都有大幅度提升。

## 2. 本文算法描述

本文算法结构如图 1 所示，由稀疏表示进行图像输入预处理模块与深度学习进行特征提取模块、重构模块三部分组成，具体步骤如下：

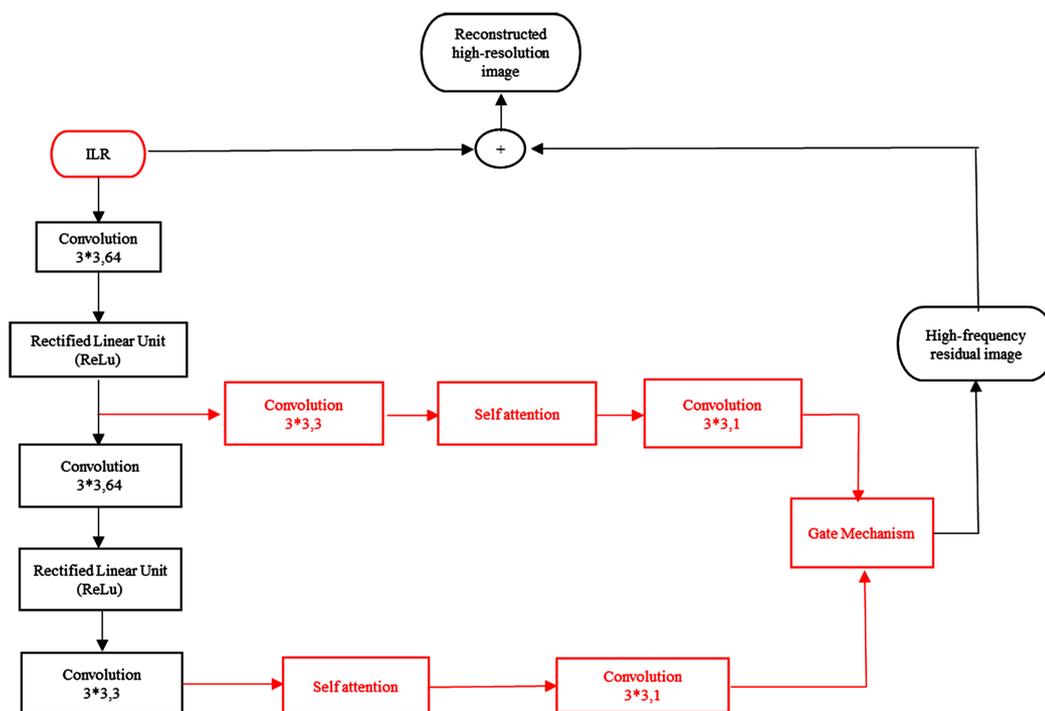


Figure 1. Algorithm structure diagram in this paper

图 1. 本文算法结构图

### 2.1. 图像输入

图像超分辨率重建算法是重建出和原始图像相比清晰度差距最小的图像，尽量减少图像信息的损失。因此图像输入与特征提取都是关键的步骤。SCSR 算法在图像超分辨率重建上应用了稀疏表示理论，通过训练 DIV2K 数据集来生成对应的高低分辨率字典，接着利用高低图像块的稀疏系数一致性来生成最终的高分辨率图像。

经过大量实验表明，SCSR 算法在时间性能与图像重建质量都要优于双三次插值，所以本文采用 SCSR 算法的重建结果作为图像输入，稀疏表示与双三次插值的重建结果如图 2 所示。



**Figure 2.** The image input was compared with the result of bicubic interpolation reconstruction using sparse representation  
**图 2.** 图像输入分别采用稀疏表示与双三次插值重建结果对比

在图 2 中, LR 是低分辨率图像, bicubic 是在图像输入时采用双三次插值后进行重建的图像, SR 是图像输入采用 SCSR 算法重建后的图像, 从图中可以发现, SCSR 的重建性能优于 bicubic, 所以我们将 SCSR 重建后的 SR 图像作为本文结构的输入进行图像超分辨率重建。

## 2.2. 特征提取

由于注意力机制会给每一个特征动态的分配不同的权值, 来代表该特征与相关信息的相关度, 权重越大, 则代表二者越相关, 即该图像较为重要的特征信息。所以本文先采用自注意力机制先进行图像的浅层特征提取, 再利用门控机制进行深度特征提取, 从而为重建图像选取更清晰的特征信息。

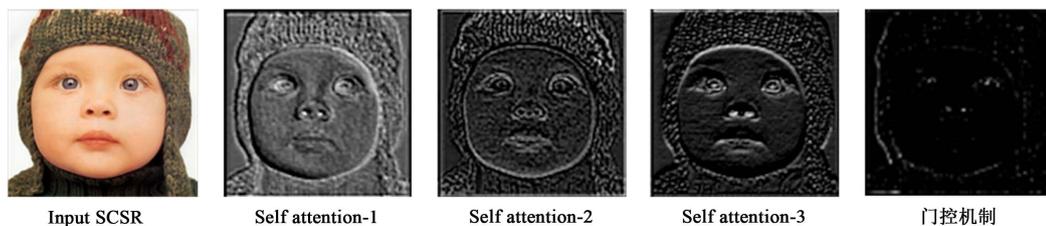
### 2.2.1. 自注意力机制特征提取

本文提出的多残差网络与自注意力机制相结合进行浅特征提取结构是在由只有一个残差结构的 VDSR 基础上增加到多残差结构, 该结构是由 3 层卷积网络和 Relu 激活函数所构成, 利用这一网络结构优化输入 LR 图像与原始 HR 图像之间的高频细节残差图像, 输入 LR 图像(ILR)通过卷积提取其特征图, 然后对这些特征图利用自注意力机制的思想进行特征提取, 对于超分辨率, 自注意力机制[13] [14] [15] 通过学习某一特征和其他所有位置特征之间的关系来生成某一细节信息, 有助于从层次特征中学习局部和非局部信息, 弥补了卷积运算只能学习局部信息的缺点, 使得恢复特征细节的能力提高。

首先, 由卷积残差网络提取到的特征图  $X$  分别送到 3 个  $1 * 1$  卷积层, 并伴随 Relu 激活函数生成新的特征图  $F(X)$ ,  $G(X)$ ,  $H(X)$ 。然后  $F(X)$  通过转置矩阵与  $G(X)$  相乘, 并用 softmax 层来计算得到注意力特征图。我们称注意力特征图为  $S(X)$ , 其计算过程如式(1)所示。

$$S_{j,i} = \frac{\exp(F_i, G_j)}{\sum_{i=1}^N \exp(F_i, G_j)} \quad (1)$$

此时,  $S(X)$  再与  $H(X)$  进行相乘得到最后的三张自注意力特征图, 如图 3 所示。其计算过程如式(2)所示。



**Figure 3.** The autoattention mechanism and gating mechanism were used to extract the feature map  
**图 3.** 使用自注意力机制和门控机制进行特征提取后的特征图

$$y = S(X) \otimes H(X) \quad (2)$$

### 2.2.2. 门控机制特征提取

由于所提出的自注意力机制特征提取是由两个并行的自注意力机制所构成，所以可以得到这两个模块提取到的不同类别的特征。但是如果我们想要知道哪种特征对图像的影响力更大，使得最终重建出质量更好的图像。这时，我们根据在自然语言处理所看到的一些工作[16]，可以知道有这样一种机制，被称为门控机制可以自动学习两个特征之间的适当权重，并在训练过程中动态分配权重。

$$a = \sigma(\tanh(W_1 * f_1 + W_2 * f_2)) \quad (3)$$

$$F = a * f_1 + (1 - a) * f_2 \quad (4)$$

其中， $f_1$ 表示由第一层卷积相连的自注意力机制获得的图像特征， $f_2$ 表示由第三层卷积相连的自注意力机制获得的图像特征， $w_1$ 和 $w_2$ 分别表示两个自注意力机制对应特征的权重矩阵。 $F$ 是模型的最终特征值，由门控机制计算。利用学习得到的权重值，该结构可以动态分配两个由自注意力机制获得图像特征的比例，其提取结果如图3所示。

### 2.3. 重建图像

在利用门控机制对特征信息提取结束后，得到的高频残差图像并不是最终的重建结果，最终图像重建结果还需要与输入图像 ILR 进行相加。

由于输入原始低分辨率图像 ILR 越接近原始 HR 图像，高频细节残差图像越容易训练。在做了大量相关工作的研究以及实验之后，我们发现，杨建超的基于稀疏表示的图像超分辨率重建算法不仅可以达到我们所要的 ILR 效果，且其过程不复杂，所耗费时间短。所以我们采用杨的算法来得到 ILR，再与网络所得到的高频残差图像相加之后得到我们最终的重建图像。

## 3. 实验结果及分析

本文实验所需要的环境为 Windows10, 64 位操作系统，采用 pytorch 深度学习框架对网络进行训练。

### 3.1. 训练过程

本文在训练时采用基于 DIV2K 数据集，由于要对所提出的网络进行大量参数的训练，所以数据集应进行扩张，以训练出更好的模型。本文将训练集中的 800 张图像进行旋转与缩放得到扩张后的 8000 张图像作为本文的训练集。本文中的所有对比实验都使用统一的测试集，包括 set5、set14、B100、Urban100，这些测试集是由花卉、人脸、建筑等自然场景构成。

本文首先将低分辨率图像和高分辨率图像分成尺寸大小为 5\*5 的图像块，同时使用深度为 3 的网络，batch size 为 16，卷积核大小为 3\*3，步长为 1，填充为 1。本文采用峰值信噪比(PSNR)和结构相似度(SSIM)作为本文算法的评价指标。

### 3.2. 实验结果比较与分析

为比较本文算法重建图像的优势，我们将本文算法与其他重建算法在不同尺度下的重建效果进行了对比，比较算法包括：SRCNN、VDSR、LapSRN [17]、DRRN [18]，将这些算法在 set5、set14、B100、Urban100 这些测试集上分别进行上采用因子为\*2、\*3、\*4 的重建。

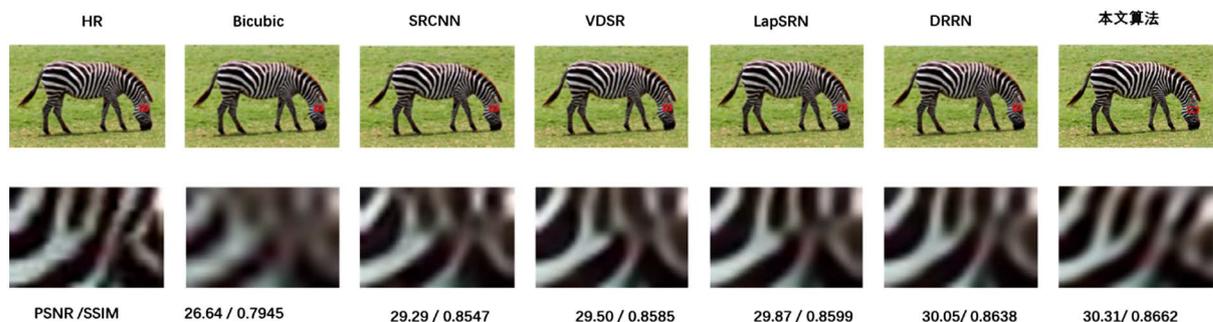
表 1 对比了不同重建算法在不同尺度下的 PSNR 与 SSIM，从表中我们可以看出当上采用因子为 2 倍、3 倍、4 倍时，本文算法与其他算法相比较，PSNR 与 SSIM 均明显提高。当上采样因子为 4 倍时，

本文算法在 set5、set14、B100、Urban100 数据集上与 DRRN 相比分别提高了。由此可见，本文算法具有较好的重建效果。为了进一步证明本文算法的重建性能，本文在图像重建整体和图像重建细节分别选择了具有代表性的几组数据在不同算法下进行对比。

**Table 1.** Different SR algorithms use different amplification factor evaluation indexes on four benchmark data sets  
**表 1.** 不同 SR 算法在四个基准测试数据集上采用不同放大因子的评估指标

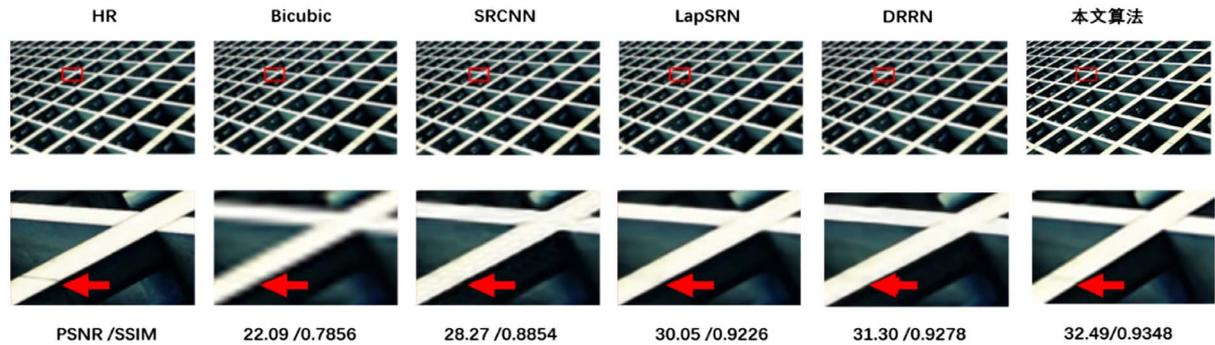
算法	放大因子	Set5 PSNR/SSIM	Set14 PSNR/SSIM	B100 PSNR/SSIM	Urban100 PSNR/SSIM
Bicubic	X4	28.42/0.810	26.10/0.704	25.96/0.669	23.15/0.659
SRCNN	X4	30.49/0.862	27.61/0.754	26.91/0.712	24.53/0.724
VDSR	X4	31.35/0.882	28.03/0.770	27.29/0.726	25.18/0.753
LapSRN	X4	31.54/0.885	28.19/0.772	27.32/0.728	25.21/0.756
DRRN	X4	31.68/0.889	28.21/0.772	27.38/0.728	25.44/0.764
<b>本文算法</b>	X4	<b>32.07/0.893</b>	<b>28.78/0.786</b>	<b>27.71/0.739</b>	<b>26.53/0.802</b>
Bicubic	X3	30.39/0.868	27.55/0.774	27.21/0.739	24.46/0.736
SRCNN	X3	32.75/0.909	29.28/0.821	28.41/0.786	26.24/0.799
VDSR	X3	33.66/0.921	29.77/0.831	28.82/0.798	27.14/0.828
LapSRN	X3	33.78/0.921	29.87/0.833	28.81/0.797	27.06/0.827
DRRN	X3	34.03/0.924	29.96/0.835	28.95/0.800	27.53/0.838
<b>本文算法</b>	X3	<b>34.45/0.937</b>	<b>30.47/0.846</b>	<b>29.15/0.814</b>	<b>28.64/0.864</b>
Bicubic	X2	33.65/0.930	30.34/0.870	29.56/0.844	26.88/0.841
SRCNN	X2	36.65/0.954	32.29/0.903	31.36/0.888	29.52/0.895
VDSR	X2	37.53/0.958	32.97/0.913	31.90/0.896	30.77/0.914
LapSRN	X2	37.52/0.959	33.08/0.913	31.80/0.895	30.41/0.910
DRRN	X2	37.74/0.959	33.23/0.914	32.05/0.897	31.23/0.919
<b>本文算法</b>	X2	<b>37.93/0.961</b>	<b>33.79/0.921</b>	<b>32.25/0.904</b>	<b>32.58/0.936</b>

图 4 和图 5 分别显示了使用 set14 和 Urban100 中的数据进行重建之后的效果对比，两组对比均使用 4 倍的放大因子。如图 5，第一行表示使用不同算法对 set14 中的“zebra”进行重建后的结果对比，第二行表示第一行每个图红框中的详细细节。我们发现大多数进行比较的算法都会产生模糊和伪影，而本文算法较所比较的算法有更清晰的细节，更接近于高分辨率图像。



**Figure 4.** A visual comparison of models with magnification factor of 4 in Set14 “zebra”

**图 4.** 放大因子为 4 倍的不同模型在 set14 “zebra” 中的可视化对比

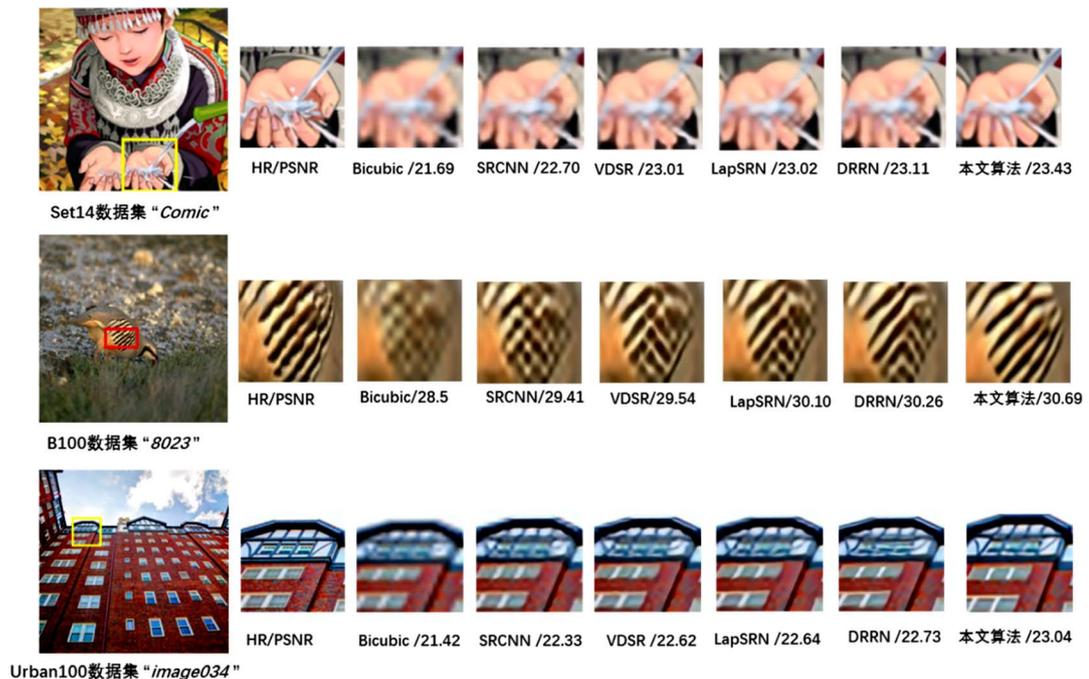


**Figure 5.** Visual Comparison of different models with magnification factor of 4 in Urban “img043”

**图 5.** 放大因子为 4 倍的不同模型在 Urban “img043” 中的可视化对比

对于 Urban “img043” 图像来说，如图 5 所示，第二行中的红色箭头所指的横线所比较算法均无法恢复，但我们的算法可以清晰地恢复出这一细节，主要是由于本文算法使用自注意力机制与门控机制相结合进行特征提取的结果。

我们在图 6 中显示了放大因子为 4 的 set14、B100、Urban100 数据集经过不同算法的重建效果可视化比较。我们观察到 set14 的 “comic” 中的手指细节算法存在模糊、B100 的 “8023” 中小鸟身上的花纹等算法存在伪影、Urban100 的 “img034” 中窗户上的窗栏等算法存在伪影，LapSRN、DRRN 等算法仍然存在明显伪影，而我们的算法可以有效的抑制这些伪影和模糊，恢复细节。



**Figure 6.** Visual comparison of SET14 “Comic”, B100 “8023” and Urban100 “Image034” with magnification factor of 4

**图 6.** 放大因子为 4 倍的 set14 “comic”、B100 “8023”、Urban100 “image034” 的可视化对比

#### 4. 结论

本文基于 VDSR 网络，为了进一步提高图像重建质量，提出了首先将 SCSR 算法重建结果作为图像输入，其次将自注意力机制、门控机制与残差网络相结合，其中残差网络作为基本的构建模块，在每层

残差网络中, 都有一个自注意力机制来提取特征。自注意力机制不仅可以一步到位的捕捉全局与局部的联系, 而且可以高效的进行特征提取, 节省计算时间。在自注意力机制后加入门控机制进一步对特征进行提取, 门控机制可以更好的提取对图像影响力更大的特征, 使算法性能得到显著提高。通过充分利用 SCSR 算法、残差网络、自注意力机制与门控机制, 本文算法取得了更好的重建效果。

## 参考文献

- [1] Dai, S., Han, M., Xu, W., Wu, Y. and Gong, Y. (2007) Soft Edge Smoothness Prior for Alpha Channel Super Resolution. *Proceedings of the IEEE Conference on Computer Vision and Pattern Classification (CVPR)*, Minneapolis, 17-22 June 2007, 1-8. <https://doi.org/10.1109/CVPR.2007.383028>
- [2] Sun, J., Xu, Z. and Shum, H. (2008) Image Super-Resolution Using Gradient Profile Prior. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Anchorage, 24-26 June 2008, 1-8.
- [3] Yang, J.C., Wright, J., Huang, T., et al. (2010) Image Super-Resolution via Sparse Representation. *IEEE Transactions on Image Processing*, **19**, 2861-2873. <https://doi.org/10.1109/TIP.2010.2050625>
- [4] Yeganli, F., Nazzal, M., Unal, M. and Ozkaramanli, H. (2014) Image Super Resolution Viasparse Representation over Coupled Dictionary Learning Based on Patch Sharpness. *European Modelling Symposium*, Prague, 21-23 October 2014, 203-208. <https://doi.org/10.1109/EMS.2014.67>
- [5] Zhang, Y., Wu, W., Dai, Y., Yang, X., Yan, B. and Lu, W. (2013) Remote Sensing Images Super-Resolution Based on Sparse Dictionaries and Residual Dictionaries. *IEEE 11th International Conference on Dependable, Autonomic and Secure Computing*, Chengdu, 21-22 December 2013, 318-323. <https://doi.org/10.1109/DASC.2013.82>
- [6] Fu, C.-H., Chen, H., Zhang, H. and Chan, Y.-L. (2014) Single Image Super Resolution Based on Sparse Representation and Adaptive Dictionary Selection. *19th International Conference on Digital Signal Processing*, Hong Kong, 20-23 August 2014, 449-453.
- [7] Timofte, R., De Smet V. and Van Gool, L. (2014) A+: Adjusted Anchored Neighborhood Regression for Fast Super-Resolution. In: *Proceedings of the Asian Conference on Computer Vision (ACCV)*, Springer, Berlin, 111-126. [https://doi.org/10.1007/978-3-319-16817-3\\_8](https://doi.org/10.1007/978-3-319-16817-3_8)
- [8] Zeiler, M.D. and Fergus, R. (2014) Visualizing and Understanding Convolutional Networks. In: *Proceedings of the European Conference on Computer Vision (ECCV)*, Springer, Berlin, 818-833. [https://doi.org/10.1007/978-3-319-10590-1\\_53](https://doi.org/10.1007/978-3-319-10590-1_53)
- [9] Radford, A., Metz, L. and Chintala, S. (2015) Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. *ICLR 2016*, Computer Science.
- [10] Dong, C., Loy, C.C., He, K. and Tang, X. (2015) Image Super Resolution Using Deep Convolutional Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **38**, 295-307. <https://doi.org/10.1109/TPAMI.2015.2439281>
- [11] Dong, C., Chen, C.L. and Tang, X. (2016) Accelerating the Super Resolution Convolutional Neural Network. In: *European Conference on Computer Vision*, Springer, Cham, 391-407. [https://doi.org/10.1007/978-3-319-46475-6\\_25](https://doi.org/10.1007/978-3-319-46475-6_25)
- [12] Kim, J., Kwon, L.J. and Mu, L.K. (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>
- [13] He, K., Zhang, X., Ren, S. and Sun, J. (2016) Deep Residual Learning for Image Recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, 27-30 June 2016, 770-778. <https://doi.org/10.1109/CVPR.2016.90>
- [14] Zhang, H., Goodfellow, I., Metaxas, D. and Odena, A. (2018) Self-Attention Generative Adversarial Networks.
- [15] Zhang, Y., Li, K., Li, K., Zhong, B. and Fu, Y. (2019) Residual Non-Local Attention Networks for Image Restoration.
- [16] Yi, P., Wang, Z.Y., Jiang, K., Jiang, J.J. and Ma, J.Y. (2019) Progressive Fusion Video Super Resolution Network via Exploiting Non Local Spatiotemporal Correlations. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Seoul, 27 October-2 November 2019, 3106-3115. <https://doi.org/10.1109/ICCV.2019.00320>
- [17] Luong, M.T., Pham, H. and Manning, C.D. (2015) Effective Approaches to Attention Based Neural Machine Translation. *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, Lisbon, September 2015, 1412-1421. <https://doi.org/10.18653/v1/D15-1166>
- [18] Lai, W.-S., Huang, J.-B., Ahuja, N. and Yang, M.-H. (2017) Deep Laplacian Pyramid Networks for Fast and Accurate Super Resolution. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, 21-26 July 2017, 5835-5843. <https://doi.org/10.1109/CVPR.2017.618>