



JOURNAL OF IMAGE AND GRAPHICS

主办: 中国科学院空天信息创新研究院
中国图象图形学学会
北京应用物理与计算数学研究所

中国图象图形学报

2021

04

VOL.26

ISSN1006-8961
CN11-3758/TB



细粒度图像分类 P847



第26卷第4期 (总第300期)

2021年4月16日

中国精品科技期刊
中国国际影响力优秀学术期刊
中国科技核心期刊
中文核心期刊

版权声明

凡向《中国图象图形学报》投稿, 均视为同意在本刊网站及CNKI等全文数据库出版, 所刊载论文已获得著作权人的授权。本刊所有图片均为非商业目的使用, 所有内容, 未经许可, 不得转载或以其他方式使用。

Copyright

All rights reserved by Journal of Image and Graphics, Institute of Remote Sensing and Digital Earth, CAS. The content (including but not limited text, photo, etc) published in this journal is for non-commercial use.

主管单位 中国科学院

主办单位 中国科学院空天信息创新研究院

中国图象图形学学会

北京应用物理与计算数学研究所

主 编 吴一戎

编辑出版 《中国图象图形学报》编辑部

通信地址 北京市海淀区北四环西路19号

邮 编 100190

电子信箱 jig@radi.ac.cn

电 话 010-58887035

网 址 www.cjig.cn

广告发布登记号 京朝工商广登字20170218号

总 发 行 北京报刊发行局

订 购 全国各地邮局

海外发行 中国国际图书贸易集团有限公司

(邮政信箱: 北京399信箱 邮编: 100048)

印刷装订 北京科信印刷有限公司

Journal of Image and Graphics

Title inscription: Song Jian

Monthly, Started in 1996

Superintended by Chinese Academy of Sciences

Sponsored by Aerospace Information Research Institute, CAS

China Society of Image and Graphics

Institute of Applied Physics and Computational Mathematics

Editor-in-Chief Wu Yirong

Editor, Publisher Editorial Office of Journal of
Image and Graphics

Address No. 19, North 4th Ring Road West, Haidian District,
Beijing, P. R. China

Zip code 100190

E-mail jig@radi.ac.cn

Telephone 010-58887035

Website www.cjig.cn

Distributed by Beijing Bureau for Distribution of Newspapers and Journals

Domestic All Local Post Offices in China

Overseas China International Book Trading Corporation
(P.O.Box 399, Beijing 100048, P.R.China)

Printed by Beijing Kexin Printing Co., Ltd.

CN 11-3758/TB

ISSN 1006-8961

CODEN ZTTXFZ

国外发行代号 M1406

国内邮发代号 82-831

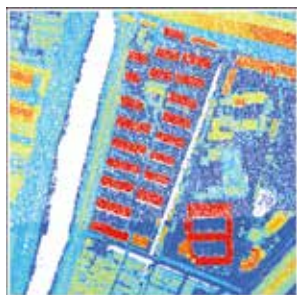
国内定价 60.00元



YOLOv3剪枝模型的多人姿态估计(第0837页)



融入时序和速度信息的自适应更新目标跟踪(第0883页)



可变半径Alpha Shapes提取机载LiDAR点云建筑物轮廓(第0910页)

综述

从图像到语言: 图像标题生成与描述

谭云兰, 汤鹏杰, 张丽, 罗玉盘 0727

CT影像肺结节分割研究进展

董婷, 魏珑, 聂生东 0751

图像处理和编码

全局注意力门控残差记忆网络的图像超分重建

王静, 宋慧慧, 张开华, 刘青山 0766

多层次感知残差卷积网络的单幅图像超分重建

何蕾, 程佳豪, 占志钰, 杨雯博, 刘沛然 0776

采用双尺度图像分解的水下彩色图像增强

李健, 张显斗, 李熹飞, 吴子朝 0787

图像分析和识别

姿态特征结合2维傅里叶变换的步态识别

王新年, 胡丹丹, 张涛, 白桂欣 0796

复杂背景下的手势识别

王银, 陈云龙, 孙前来 0815

结合形变模型与图像修复的人脸姿态矫正

吴从中, 郑荣生, 臧怀娟, 刘明威, 徐甲甲, 詹曙 0828

YOLOv3剪枝模型的多人姿态估计

蔡哲栋, 应娜, 郭春生, 郭锐, 杨鹏 0837

YOLOv3和双线性特征融合的细粒度图像分类

闫子旭, 侯志强, 熊磊, 刘晓义, 余旺盛, 马素刚 0847

深度多模态融合服装风格检索

苏卓, 柯司博, 王若梅, 周凡 0857

超像素条件随机场下的RGB-D视频显著性检测

李贝, 杨铀, 刘琼 0872

融入时序和速度信息的自适应更新目标跟踪

尹宽, 李均利, 胡凯, 李丽 0883

融合逆密度函数与关系形状卷积神经网络的点云分析

邱云飞, 朱梦影 0898

图像理解和计算机视觉

可变半径Alpha Shapes提取机载LiDAR点云建筑物轮廓

伍阳, 王丽妍, 胡春霞, 程亮 0910

3D遮挡模型引导的光场图像深度获取

吴迪, 张旭东, 张骏, 范之国, 孙锐 0924

计算机图形学

基于全正基的非均匀三次加权B样条

姬佩佩, 张贵仓, 汪凯, 孟建军 0939

遥感图像处理

LBP特征分类的极化SAR图像机场跑道检测

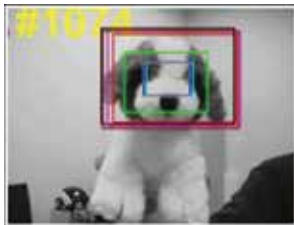
韩萍, 万义爽, 刘亚芳, 韩宾宾 0952

CONTENTS

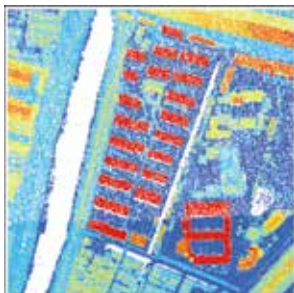
JOURNAL OF IMAGE AND GRAPHICS



Research on multiperson pose estimation combined with YOLOv3 pruning model(P0837)



Adaptive update object tracking algorithm incorporating timing and speed information(P0883)



Extraction of building contours from airborne LiDAR point cloud using variable radius Alpha Shapes method(P0910)

Review

From image to language: image captioning and description

Tan Yunlan, Tang Pengjie, Zhang Li, Luo Yupan 0727

Research progress of lung nodule segmentation based on CT images

Dong Ting, Wei Long, Nie Shengdong 0751

Image Processing and Coding

Learning global attention-gated multi-scale memory residual networks for single-image super-resolution

Wang Jing, Song Huihui, Zhang Kaihua, Liu Qingshan 0766

Single image super-resolution reconstruction based on multi-level perceptual residual convolutional network

He Lei, Cheng Jiahao, Zhan Zhiyu, Yang Wenbo, Liu Peiran 0776

Underwater color image enhancement based on two-scale image decomposition

Li Jian, Zhang Xiandou, Li Shangfei, Wu Zizhao 0787

Image Analysis and Recognition

Gait recognition using pose features and 2D Fourier transform

Wang Xinnian, Hu Dandan, Zhang Tao, Bai Guixin 0796

Hand gesture recognition in complex background

Wang Yin, Chen Yunlong, Sun Qianlai 0815

Face pose correction based on morphable model and image inpainting

Wu Congzhong, Zheng Rongsheng, Zang Huaijuan, Liu Mingwei, Xu Jiajia, Zhan Shu 0828

Research on multiperson pose estimation combined with YOLOv3 pruning model

Cai Zhedong, Ying Na, Guo Chunsheng, Guo Rui, Yang Peng 0837

Fine-grained classification based on bilinear feature fusion and YOLOv3

Yan Zixu, Hou Zhiqiang, Xiong Lei, Liu Xiaoyi, Yu Wangsheng, Ma Sugang 0847

Fashion style retrieval based on deep multimodal fusion

Su Zhuo, Ke Sibao, Wang Ruomei, Zhou Fan 0857

RGB-D video saliency detection via superpixel-level conditional random field

Li Bei, Yang You, Liu Qiong 0872

Adaptive update object tracking algorithm incorporating timing and speed information

Yin Kuan, Li Junli, Hu Kai, Li Li 0883

Point cloud analysis model combining inverse density function and relation-shape convolution neural network

Qiu Yunfei, Zhu Mengying 0898

Image Understanding and Computer Vision

Extraction of building contours from airborne LiDAR point cloud using variable radius Alpha Shapes method

Wu Yang, Wang Liyan, Hu Chunxia, Cheng Liang 0910

Light field depth estimation guided by 3D occlusion model

Wu Di, Zhang Xudong, Zhang Jun, Fan Zhiguo, Sun Rui 0924

Computer Graphics

Non uniform weighted cubic B-spline based on all positive basis

Ji Peipei, Zhang Guicang, Wang Kai, Meng Jianjun 0939

Remote Sensing Image Processing

Airport runway detection based on LBP feature classification in PolSAR images

Han Ping, Wan Yishuang, Liu Yafang, Han Binbin 0952

中图法分类号: TP391.4 文献标识码: A 文章编号: 1006-8961(2021)04-0766-10

论文引用格式: Wang J, Song H H, Zhang K H and Liu Q S. 2021. Learning global attention-gated multi-scale memory residual networks for single-image super-resolution. Journal of Image and Graphics, 26(04): 0766-0775 (王静, 宋慧慧, 张开华, 刘青山. 2021. 全局注意力门控残差记忆网络的图像超分辨率重建. 中国图象图形学报, 26(04): 0766-0775) [DOI: 10.11834/jig.200174]

全局注意力门控残差记忆网络的图像超分重建

王静, 宋慧慧*, 张开华, 刘青山

南京信息工程大学大气环境与装备技术协同创新中心, 江苏省大数据分析技术重点实验室, 南京 210044

摘要: **目的** 随着深度卷积神经网络的兴起, 图像超分重建算法在精度与速度方面均取得长足进展。然而, 目前多数超分重建方法需要较深的网络才能取得良好性能, 不仅训练难度大, 而且到网络末端浅层特征信息容易丢失, 难以充分捕获对超分重建起关键作用的高频细节信息。为此, 本文融合多尺度特征充分挖掘超分重建所需的高频细节信息, 提出了一种全局注意力门控残差记忆网络。**方法** 在网络前端特征提取部分, 利用单层卷积提取浅层特征信息。在网络主体非线性映射部分, 级联一组递归的残差记忆模块, 每个模块融合多个递归的多尺度残差单元和一个全局注意力门控模块来输出具备多层次信息的特征表征。在网络末端, 并联多尺度特征并通过像素重组机制实现高质量的图像放大。**结果** 本文分别在图像超分重建的5个基准测试数据集(Set5、Set14、B100、Urban100和Manga109)上进行评估, 在评估指标峰值信噪比(peak signal to noise ratio, PSNR)和结构相似性(structural similarity, SSIM)上相比当前先进的网络模型均获得更优性能, 尤其在Manga109测试数据集上本文算法取得的PSNR结果达到39.19 dB, 相比当前先进的轻量型算法AWSRN(adaptive weighted super-resolution network)提高0.32 dB。**结论** 本文网络模型在对低分图像进行超分重建时, 能够联合学习网络多层次、多尺度特征, 充分挖掘图像高频信息, 获得高质量的重建结果。

关键词: 单幅图像超分辨率(SISR); 深度卷积神经网络(DCNN); 注意力门控机制; 多尺度残差单元(MRUs); 递归学习

Learning global attention-gated multi-scale memory residual networks for single-image super-resolution

Wang Jing, Song Huihui*, Zhang Kaihua, Liu Qingshan

Collaborative Innovation Center on Atmospheric Environment and Equipment Technology, Jiangsu Key Laboratory of Big Data Analysis Technology, Nanjing University of Information Science and Technology, Nanjing 210044, China

Abstract: **Objective** With the rapid development of deep convolutional neural networks (CNNs), great progress has been made in the research of single-image super-resolution (SISR) in terms of accuracy and efficiency. However, existing methods often resort to deeper CNNs that are not only difficult to train, but also have limited feature resolutions to capture the rich high-frequency detail information that is essential for accurate SR prediction. To address these issues, this letter presents a global attention-gated multi-scale memory network (GAMMNet) for SISR. **Method** GAMMNet mainly consists of three key components: feature extraction, nonlinear mapping (NM), and high-resolution (HR) reconstruction. In the fea-

收稿日期: 2020-05-18; 修回日期: 2020-07-10; 预印本日期: 2020-07-17

* 通信作者: 宋慧慧 songhuihui@nuist.edu.cn

基金项目: 国家新一代人工智能重大项目(2018AAA0100400); 国家自然科学基金项目(61872189, 61876088, 61532009); 江苏省自然科学基金项目(BK20191397, BK20170040)

Supported by: National Major Project of China for New Generation of AI (2018AAA0100400); National Natural Science Foundation of China (61872189, 61876088, 61532009); Natural Science Foundation of Jiangsu Province, China (BK20191397, BK20170040)

ture extraction part, the input low-resolution (LR) image passes through a 3×3 convolution layer to learn the low-level features. Then, we utilize a recursive structure to perform NM to learn the deeper-layer features. At the end of the network, we arrange four kernels with different sizes followed by a global attention gate (GAG) to achieve the HR reconstruction. Specifically, the NM module consists of four recursive residual memory blocks (RMBs). Each RMB outputs a multi-level representation by fusing the output of the top multi-scale residual unit (MRU) and the ones from the previous MRUs, followed by a GAG module, which serves as a gate to control how much of the previous states should be memorized and decides how much of the current state should be kept. The details of how to design the two novel modules (MRU and GAG) will be explained as follows. MRU takes advantage of the wide activation architecture inspired by the wide activation super-resolution (WDSR) method. As the nonlinear ReLU (rectified linear units) layers in residual blocks hinder the transmission of information flow from shallow layers to deeper layers, the wide activation architecture increases the channels of feature map before ReLU layers to help transmit information flows. Moreover, as the depth of the network increases, problems such as underutilization of features and gradual disappearance of features during transmission occur. Making full use of features is the key to network reconstruction of high-quality images. We parallelly stack two different convolution kernels with sizes of 3×3 and 5×5 to extract multi-scale features, which are further enhanced by the proposed GAG module, yielding the residual output. Finally, the output of MRU is the addition of the input and the residual output. GAG serves as a gate to enhance the input feature channels with different weights. First, different from the residual channel attention network (RCAN) that mainly considers the correlation between feature channels, our GAG takes into account the useful spatial holistic statistic information of the feature maps. We utilize a special spatial pooling operation to improve the global average pooling used in the original channel attention to obtain global context features, and take a weighted average of all location features to establish a more efficient long-range dependency. Then, we aggregate the global context information on each location feature. We first leverage a 1×1 convolution to reduce the feature channel number to 1. Then, we use a Softmax function to capture the global context information for each pixel, yielding the pixel-wise attention weight map. Afterward, we introduce a learning parameter λ_1 to adaptively rescale the weight map. Now the weight map is correlated with the input feature maps, outputting the channel-wise weight vector that encodes the global holistic statistic information of the feature maps. Second, in order to reduce the amount of calculations under the premise of establishing effective long-distance dependencies, we learn to use a bottleneck layer used in SE block to implement feature transform, which not only significantly reduces the amount of calculations but also captures the correlation between feature channels. We feed the channel-wise attention weight vector into a feature transform module that consists of one 1×1 convolution layer, one normalization layer, one ReLU layer, and one 1×1 convolution layer and multiply an adaptively learning parameter λ_2 , yielding the enhanced channel-wise attention weights that capture channel-wise dependencies. Finally, we channel-wisely multiply the input feature maps and the enhanced channel-wise attention weights to aggregate the global context information onto each local feature. In the image magnification stage, we design an efficient reconstruction structure to combine local multi-scale features and global features to achieve image magnification. We first leverage three different convolutions followed by a GAG to adaptively adjust the reconstruction feature weights, which make full use of the local multi-scale features of the reconstructed part. Then, a pixel-shuffle module is added behind each branch to perform image magnification. At last, all the reconstructed outputs are added together with the top network branch to combine local and global feature information, outputting the final SR image. **Result** We adopt the DIV2K (DIVERse 2K resolution image) dataset, the most widely used training dataset for deep CNN-based SISR, to train our model. This dataset contains 1 000 HR images, 800 of which are used for training. We preprocess the HR images by bicubic down-sampling to obtain the LR images. Then, we use several commonly used benchmark datasets, including Set5, Set14, B100, Urban100, and Manga100 to test our model. The evaluation metrics are the peak signal-to-noise ratio (PSNR) and the structural similarity index measurement (SSIM) on the Y channel of the transformed YCbCr space. The input image patches are cropped with a size of 48×48 pixels, and the mini training batch size is set to 16. The hyperparameters of input, internal, and output channel numbers are set to 32, 128, and 32 for the MRUs, respectively. We arrange four RMBs in the non-linear mapping module, among which each block has four MRUs. For the upscale module, we use four different kernel sizes (3×3 , 5×5 , 7×7 , and 9×9), followed by a GAG to generate the HR outputs. We compare our GAMMNet with several state-of-the-art deep CNN-based SISR meth-

ods, including super-resolution convolutional neural network (SRCNN), deeply-recursive convolutional network (DRCN), deep recursive residual network (DRRN), memory network (MemNet), cascading residual network (CARN), multi-scale residual network (MSRN), and adaptive weighted super-resolution network (AWSRN). Obviously, our GAMMNet achieves the best performance in terms of both PSNR and SSIM among all compared methods in almost all benchmark datasets, except for the SSIM on the Urban 100, where our GAMMNet achieves the second best performing SSIM of 0.792 6, which is slightly lower than the best AWSRN with SSIM of 0.793 0. Finally, we conduct ablation experiments on the important components of GAMMNet. We use $2 \times \text{Set5}$ as the test set. The experiment first replaces the MRUs in GAMMNet with the smallest residual unit in WDSR and removes the GAG as the benchmark. Then, it adds the MRU and GAG and finally trains GAMMNet (simultaneously using the two proposed modules). The results show that the MRU and GAG modules improve PSNR by 0.1 and 0.08 points, respectively, and GAMMNet achieves the best performance on PSNR and SSIM, demonstrating the effectiveness of both module designs. **Conclusion** In this study, the shallow features of the network are first extracted by the feature extraction module. Then, a nested recursive structure is used to realize the NM and learn the deeper features. This structure combines the features of different scales to effectively learn the context information of the feature maps at each level, and solves the problem of feature disappearing during information transmission by fusing the output of different levels. Finally, in the reconstruction part, the features of different scales are parallelized, and pixel shuffle is used to achieve high-quality magnification of images.

Key words: single-image super-resolution (SISR); deep convolutional neural networks (DCNN); attention gate mechanism; multi-scale residual units (MRUs); recursive learning

0 引言

图像超分重建目标是将模糊的低分辨率输入图像转变为对应清晰的高分辨率图像。重建后的高分辨率图像获得目标更丰富的细节信息,因此该技术广泛应用于安全监控(Zou 和 Yuen, 2012)、医学图像(Shi 等, 2016)和卫星图像(Song 等, 2018)等领域。图像超分重建作为不适定的问题,即单个低分图像通常对应多个不同的高分图像,如何选择最优高分图像仍是一个未解难题,尽管目前该方向取得一些突破性进展,但以上难题仍未完全解决。

深度卷积神经网络具有强大的特征学习能力,可以直接从大规模训练数据中学习低分到高分图像之间复杂的非线性映射关系,基于深度卷积神经网络的图像超分重建研究方法引起广泛关注。Dong 等人(2014)首次将深度卷积神经网络引入图像超分重建,提出开创性的工作 SRCNN(super-resolution convolutional neural network)。SRCNN 包含 3 部分:特征提取、非线性映射与高分图像重构。之后,基于 SRCNN 的图像超分重建研究迅速发展。Kim 等人(2016)提出基于极深卷积神经网络的图像超分重建算法 VDSR(very deep super-resolution),通过残差学习与梯度剪裁改善极深网络结构导致的梯度消失问题,同时通过提高学习率加快收敛速度。Tai 等

人(2017a,b)提出基于递归卷积神经网络的图像超分重建算法 DRRN(deep recursive residual network)和 MemNet(memory network),有效加深网络并减少网络参数。针对极深卷积神经网络训练难的问题, Li 等人(2018)设计出 MSRN(multi-scale residual network),通过拓宽网络来提取多尺度特征,融合不同层输出来充分学习输入图像的高频信息。Zhang 等人(2018)提出 RCAN(residual channel attention network),首次将注意力机制应用于图像超分重建,利用通道注意力机制学习不同通道特征之间的相关性,取得当时最佳性能。

然而,现在很多算法实现精确的超分重建常常过度依赖极深的卷积神经网络,参数量过大,消耗大量内存。此外,随着网络的加深,高层特征更倾向于表征低频的语义信息,导致对提高超分性能至关重要的高频信息缺失。为此,本文提出一种基于全局注意力门控残差记忆网络(global attention-gated multi-scale memory residual network, GAMMNet)的图像超分重建算法,设计递归网络结构降低参数量,并自适应地融合多层特征学习丰富的高频细节信息(沈明玉等, 2019),实现高质量超分重建。本文主要贡献如下:

1) 提出一个图像超分重建网络架构,采用递归残差记忆模块实现高效的非线性映射。

2) 设计的多尺度残差单元能有效扩大特征图

感受野,充分捕获特征蕴含的丰富上下文信息。

3)设计的全局注意力门控模块能自适应融合多层次、多尺度卷积特征,增强所学特征表征力。

4)与其他较先进的算法相比,GAMMNet在5个图像超分重建标准测试数据集上都取得较好的性能。

1 全局注意力门控残差记忆网络

图1展示了本文提出的全局注意力门控残差记忆网络结构,包含浅层特征提取、深层非线性映射和高分重建3部分。特征提取阶段,网络输入的低分辨率图像经过一个 3×3 卷积层学习浅层特征信息(李宇翔等,2018)。然后,利用一个嵌套的递归结构设

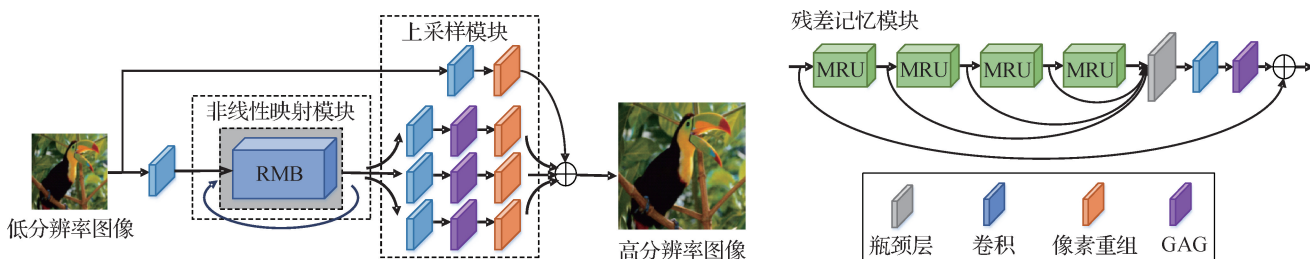


图1 全局注意力门控残差记忆网络结构

Fig. 1 Architecture of GAMMNet

以下将详细阐述 GAMMNet 的 3 个主要模块：多尺度残差单元 MRU、全局注意力门控模块 GAG 和上采样模块。

1.1 多尺度残差单元 MRU

目前超分辨率研究通常倾向于使用深层卷积神经网络来提高性能。然而,随着网络深度的提高会出现特征利用不足和特征逐渐消失的问题。对此,Li 等人(2018)提出一种多尺度结构充分利用网络的特征以缩减网络深度。但是并联不同尺度的卷积核使得网络参数量陡增。本文在设计多尺度残差单元时考虑到这一问题,通过在残差模块内 ReLU (rectified linear unit) 层之后添加瓶颈层、在残差模块末端使用注意力调整模块,在保证网络性能前提下,有效降低参数量。

图2展示了 MRU 结构,本文受 WDSR (wide activation super-resolution) (Yu 等,2018) 方法启发,设计一种宽激活残差结构。由于传统残差模块中非线性激活层 ReLU 阻碍了信息流从浅层到深层的传输,本文拓宽 ReLU 层之前特征图通道数量,实现信

息流的高效传输。此外,因为在神经网络浅层特征图中的像素点只是对原始输入图像进行局部细节特征信息提取,图像全局信息较少。但是深层网络特征图中的像素点映射到原始输入图像的范围逐渐变大,其更倾向于表征全局的语义信息。因而随着网络的加深,将出现特征所蕴含高频细节信息逐渐丢失的问题,而这些高频信息是实现高质量图像超分重建的关键。为此,本文设计多尺度特征融合和残差学习来补偿高层特征细节信息的损失。

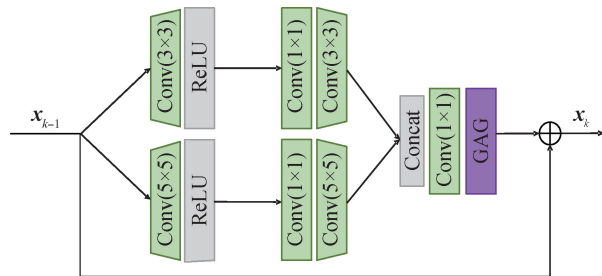


图2 多尺度残差单元

Fig. 2 Multi-scale residual unit

首先,本文通过并联两个大小分别为 3×3 和

5×5 的卷积核来提取多尺度特征。不同大小的卷积核使得网络传输过程中特征图像素点对应到原始输入图像上的感受野进一步增强。然后,通过一个 1×1 的卷积层来融合不同尺度的信息并将融合后的结果送入 GAG 模块,进一步增强特征得到残差输出。最后,MRU 的输出是其初始输入与残差输出之和。本文设 $\mathbf{x}_{k-1}$ 和 $\mathbf{x}_k$ 为每个 MRU 模块输入和输出的特征图,MRU 表示为

$$\mathbf{x}_k = H_{\text{MRU}}^k(\mathbf{x}_{k-1}) + \mathbf{x}_{k-1} \quad (1)$$

式中, H_{MRU}^k 表示第 k 个 MRU 的残差函数,针对多个 MRU 组成的 RMB 结构,本文分别定义 $\mathbf{x}_{m-1}$ 和 $\mathbf{x}_m$ 作为其输入和输出,模块 RMB 表示为

$$\mathbf{x}_m = H_{\text{RMB}}^m([\mathbf{x}_m^0, \mathbf{x}_m^1, \dots, \mathbf{x}_m^n]) + \mathbf{x}_{m-1} \quad (2)$$

式中, $\mathbf{x}_m^n$ 表示第 m 个 RMB 内各个 MRU 的输出, H_{RMB}^m 表示第 m 个 RMB 映射函数。

1.2 全局注意力门控模块 GAG

针对深度神经网络中残差块的堆叠不能很好地

利用其特征表征能力的问题,Zhang 等人(2018)提出一种通道注意力结构。由于特征图不同通道的信息对图像放大具有不同作用,因此为通道间依赖关系建模,利用平均池化将每个通道上信息压缩为一个通道描述符,然后使用有效的全连接结构对压缩过的信息建模。但是,由于特征图的同一通道不同位置特征信息重要度不同,通过平均池化对其归一化会导致很多空间信息丢失,本文就这种权重信息的获取进行扩展学习。

如图 3 所示,GAG 模块首先对输入特征空间进行注意力求解,然后自适应地调整通道之间重要度关系,GAG 模块表示为

$$\mathbf{z}_i = \mathbf{X}_i \otimes \text{tf}(\sum_j \text{spl}(\mathbf{x}_j)) \quad (3)$$

式中, $\sum_j \text{spl}(\mathbf{x}_j)$ 表示空间池化操作,用来捕获特征丰富的全局上下文信息; tf 表示学习特征通道之间相关性的转换操作,如图 3 中的瓶颈层的设置; $\otimes$ 表示将全局上下文信息重新聚合到局部特征上的逐像素相乘。

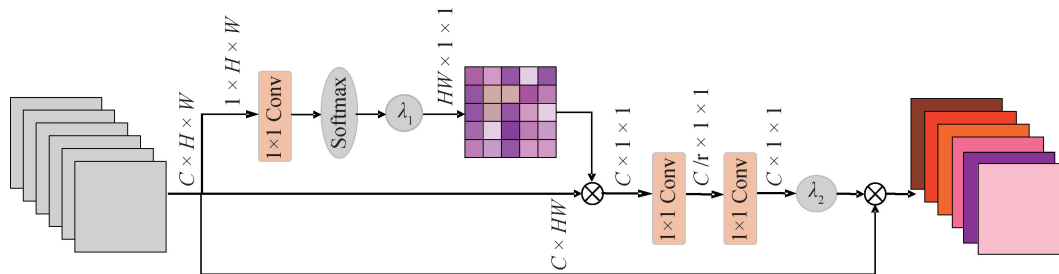


图 3 全局注意力门控模块

Fig. 3 Global attention gate module

GAG 模块考虑了特征图空间全局的统计信息。本文利用空间池化操作来改善原始通道注意力中使用的平均池化(Hu 等,2018)。 spl 对所有位置特征进行加权平均,建立更加有效的长范围依赖关系,可获得更加丰富的全局上下文信息。如图 3 所示,首先利用 1×1 卷积将特征通道数减少为 1。然后,通过 Softmax 函数捕获像素的全局上下文信息,从而生成逐像素注意力权重图。之后,引入一个学习参数 λ_1 来自适应地缩放权重图的比例。至此,生成的权重图与输入的特征图之间产生相互依赖的关系,输出对特征图全局统计信息重新编码的权重向量。

为了在建立有效的长范围依赖关系前提下进一步减少计算量,学习经典的 SE(squeeze and excitation)(Cao 等,2019;Hu 等,2018)模块中的瓶颈层来实现特征转换 tf 。不但可以有效减少计算量,而且

有助于捕获特征通道之间的相关性。将前端生成的注意力权重向量送入特征转换模块,该模块由一个 1×1 卷积层、一个归一化层、一个 ReLU 层和一个 1×1 卷积层组成,然后增加可变参数 λ_2 对其输出比例进行缩放,最终得到强化的注意力权重图。

最后,本文逐通道地将输入特征图和增强后的注意力权重图相乘,将获得的全局上下文信息重新聚合到每个局部特征之中。

1.3 上采样模块(up-sampling model)

在图像放大阶段,如图 1 所示,本文设计一种有效的超分重建结构,结合多尺度特征并通过像素重组方法实现高质量的图像放大。

首先,在网络非线性映射末端并联多个不同大小的卷积核,通过此方法可以增强感受野,获取网络学习到的特征图不同尺度的信息。接着,在每个卷

积核后连接一个 GAG 模块分别自适应地调整各尺度重建特征的权重,进一步充分利用重建部分不同尺度的特征信息。然后,在每个分支末端通过像素重组模块(Shi 等,2016)实现图像的放大,像素重组模块通过学习一组尺度扩展滤波器,将网络输出的低分辨率特征图直接映射到高分辨率图中,同时降低超分操作的计算复杂度。最后,所有重建的输出特征融合并通过跳跃连接与网络初始输入放大后的分支相加,输出最终的高质量超分重建图像。

2 实验设置和结果分析

2.1 实验设置

采用图像超分重建算法中使用最广泛的训练数据集 DIV2K (DIVERse 2K resolution image dataset) (Timofte 等,2017)训练模型。该数据集包含 1 000 幅高分图像,其中,800 幅用于训练。首先,通过双三次下采样对高分图像进行预处理获得低分图像,然后,使用常用的基准测试数据集: Set5 (Bevilacqua 等,2012), Set14 (Yang 等,2010), B100 (Martin 等,2001), Urban100 (Huang 等,2015) 和 Manga109 (Matsui 等,2017) 来测试模型。评估指标采用图像转换到 YCbCr (Y 为亮度,Cb 为蓝色色度,Cr 为红色色度)空间的 Y 通道上的峰值信噪比 (peak signal to noise ratio, PSNR) 和结构相似性 (structural similarity, SSIM) (Wang 等,2004)。

输入图像块的裁剪尺寸设置为 48×48 像素,最

小训练批次大小设置为 16。MRU 的输入、内部和输出通道数的超参数分别设置为 32、128、32。在非线性映射模块部分设置 4 个 RMB,其中,每个 RMB 设置 4 个 MRU。在图像放大阶段,分别设置 4 个卷积核,大小分别为 3×3 、 5×5 、 7×7 和 9×9 。本文使用 Adam (adaptive moment estimation) 优化器 (Kingma 和 Ba, 2015) 和 L1 损失 (Timofte 等, 2016),并以端到端的方式训练模型。初始学习率设置为 10^{-3} ,然后利用余弦衰减方式改变学习率。利用 PyTorch (Paszke 等,2017) 框架在 1 块 GeForce RTX 2080 Ti GPU 上进行模型训练与测试。

2.2 定性评估

图 4 是将 GAMMNet 与性能先进的模型 CARN (cascading residual network) (Ahn 等,2018)、MSRN (Li 等,2018) 和 AWSRN (adaptive weighted super-resolution network) (Wang 等,2019) 结果图像在视觉效果方面进行比较。从图中放大的局部区域可以比较出 CARN、MSRN 和 AWSRN 算法生成的结果相对模糊很多,并且 AWSRN 还生成了与原图相差较大的斑马条纹。此外,GAMMNet 中 GAG 模块可以利用图像全局信息增强当前特征,有助于局部细节的还原。例如多数算法在学习图像中斑马面部时,只是对局部特征进行学习,还原出的结果相对模糊,而本文算法可以从全局信息学习图片特征,对图像中更大范围区域的语义信息有更好的理解,进而还原出更加清晰的斑马面部细节。因此,相比其他算法,本文算法可以生成视觉效果更佳

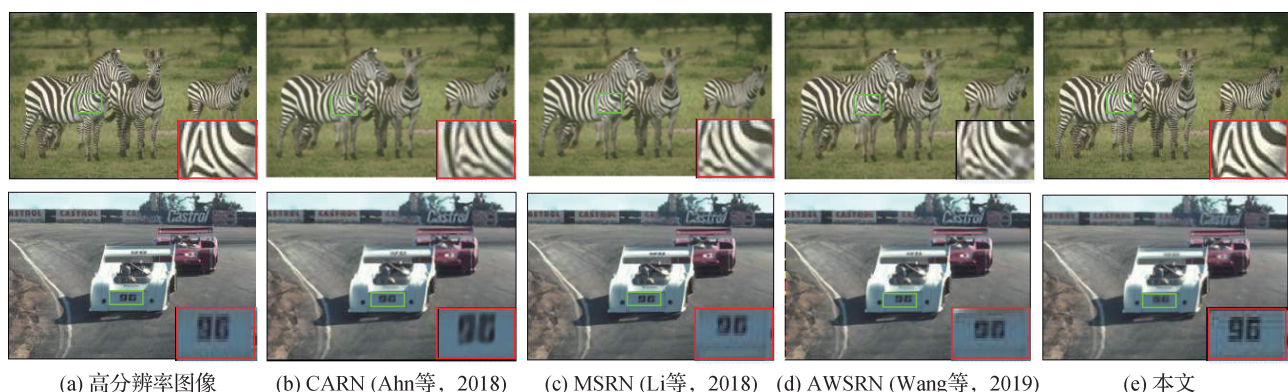


图 4 图像“253027”、“21077”在放大倍数为 4 时超分辨率结果

Fig. 4 SISR results on image “253027”, “21077” for the upscaling factor 4

((a) HR images; (b) CARN (Ahn et al., 2018); (c) MSRN (Li et al., 2018); (d) AWSRN (Wang et al., 2019); (e) ours)

在图5中,本文通过热力图将GAG模块中经过空间池化后生成的注意力图进行了可视化。图中按序选取了GAMMNet包含的非线性映射部分单个RMB模块中不同深度的GAG。由图5可见,该模块中前端网络层相对更关注于局部复杂的细节信息,

例如,错杂的树枝、鸟的嘴巴,而后端网络层逐渐注重于图像全局的整体信息,例如,图像中的鸟、骑在马上运动员。总的来说,GAG模块使得网络关注于图像中更加显著的部分,更利于网络提升图片中显著目标的重建效果。



图5 GAG可视化效果图

Fig. 5 Visual effects of GAG((a) GAG-1; (b) GAG-2; (c) GAG-3; (d) GAG-4; (e) GAG output)

2.3 定量评估

将本文算法实验结果与几种基于深度卷积神经网络的先进的轻量型图像超分方法进行比较,包括SRCNN(Dong等,2014),DRCN(deeply-recursive convolutional network)(Kim等,2016),DRRN(Tai

等,2017a),MemNet(Tai等,2017b),CARN(cascading residual network)(Ahn等,2018),MSRN(multi-scale residual network)(Li等,2018)和AWSRN(adaptive weighted super-resolution network)(Wang等,2019)。表1和表2列举了各类方法的PSNR和SSIM结

表1 评估方法的定量结果(PSNR)

Table 1 Quantitative results(PSNR) of the evaluated methods

数据集	倍数	PSNR/dB							
		SRCNN	DRCN	DRRN	MemNet	CARN	MSRN	AWSRN	本文
Set5	2	36.66	37.63	37.74	37.78	37.76	38.08	<u>38.11</u>	38.16
	3	32.75	33.82	34.03	34.09	34.29	34.46	<u>34.52</u>	34.61
	4	30.48	31.53	31.68	31.74	32.13	32.26	<u>32.27</u>	32.40
Set14	2	32.45	33.04	33.23	33.28	33.52	33.70	<u>33.78</u>	33.85
	3	29.30	29.76	29.96	30.00	30.29	<u>30.41</u>	30.38	30.49
	4	27.50	28.02	28.21	28.26	28.60	28.63	<u>28.69</u>	28.74
B100	2	31.36	31.85	—	32.08	32.09	32.23	<u>32.26</u>	32.28
	3	28.41	28.80	—	28.96	29.06	29.15	<u>29.16</u>	29.21
	4	26.90	27.23	—	27.40	27.58	27.61	<u>27.64</u>	27.64
Urban100	2	29.50	30.75	31.23	31.31	31.92	32.29	<u>32.49</u>	32.64
	3	26.24	27.15	27.53	27.56	28.06	28.33	<u>28.42</u>	28.54
	4	24.52	25.14	25.44	25.50	26.07	26.22	<u>26.29</u>	26.34
Manga109	2	35.60	—	—	—	—	38.69	<u>38.87</u>	39.19
	3	30.48	—	—	—	—	33.67	<u>33.85</u>	34.07
	4	27.58	—	—	—	—	30.57	<u>30.72</u>	30.86

注:加粗字体为每行最优结果,下划线字体为每行次优结果,“—”表示原文中未给出实验结果。

表 2 评估方法的定量结果(SSIM)
Table 2 Quantitative results(SSIM) of the evaluated methods

数据集	倍数	SSIM							
		SRCNN	DRCN	DRRN	MemNet	CARN	MSRN	AWSRN	GAMMNet
Set5	2	0.929 9	0.958 8	0.959 1	0.959 7	0.959 0	0.960 7	<u>0.960 8</u>	0.961 0
	3	0.909 0	0.922 6	0.924 4	0.924 8	0.925 5	0.927 8	<u>0.928 1</u>	0.928 8
	4	0.862 8	0.885 4	0.888 8	0.889 3	0.893 7	0.896 0	<u>0.896 0</u>	0.897 0
Set14	2	0.906 7	0.911 8	0.913 6	0.914 2	0.916 6	0.918 6	<u>0.918 9</u>	0.919 7
	3	0.821 5	0.831 1	0.834 9	0.835 0	0.840 7	<u>0.843 7</u>	0.842 6	0.844 5
	4	0.751 3	0.767 0	0.772 1	0.772 3	0.780 6	0.783 6	<u>0.784 3</u>	0.784 4
B100	2	0.887 9	0.894 2	—	0.897 8	0.897 8	0.900 2	<u>0.900 6</u>	0.900 9
	3	0.786 3	0.796 3	—	0.800 1	0.803 4	0.806 4	<u>0.806 9</u>	0.807 6
	4	0.710 7	0.723 3	—	0.728 1	0.734 9	0.738 0	<u>0.738 5</u>	0.738 6
Urban100	2	0.894 6	0.913 3	0.918 8	0.919 5	0.925 6	0.930 3	<u>0.931 6</u>	0.932 8
	3	0.798 9	0.827 6	0.837 8	0.837 6	0.849 3	0.856 1	<u>0.858 0</u>	0.859 6
	4	0.722 1	0.751 0	0.763 8	0.763 0	0.783 7	0.791 1	0.793 0	<u>0.792 6</u>
Manga109	2	0.966 3	—	—	—	—	0.977 2	<u>0.977 6</u>	0.977 9
	3	0.911 7	—	—	—	—	0.945 6	<u>0.946 3</u>	0.947 4
	4	0.855 5	—	—	—	—	0.910 3	<u>0.910 9</u>	0.912 3

注:加粗字体为每行最优结果,下划线字体为每行次优结果,“—”表示原文中未给出实验结果。

果。显然,本文方法在 5 个基准测试集上 PSNR 和 SSIM 指标均达到最优。本文算法在 Urban100 上的 SSIM 值为 0.792 6,在所有对比算法中排名第 2,仅低于结果最佳的 AWSRN 0.793 0。另外,表 3 中列出 GAMMNet 与几种其他轻量型网络模型参数量与计算量的比较结果。可以看出 MemNet 相对参数量比较小,但是计算量很大且性能一般。CARN 和 AWSRN 参数量与计算量近似 GAMMNet,但是 GAMMNet 的性能提升显著,PSNR 达到 38.16 dB。

表 3 网络模型的参数量与计算量比较
Table 3 Comparison of parameters and calculation of network models

	模型				
	MemNet	CARN	MSRN	AWSRN	GAMMNet
参数量/K	677	1 592	5 930	1 397	1 570
计算量/G	2 662.4	222.8	1 365.4	320.5	576.5
PSNR/dB	37.78	37.76	38.08	38.11	38.16

注:加粗字体为最优结果。

2.4 消融实验

本文对所提模型中各重要组成部分所起作用进行消融实验,测试集为 Set5,放大倍数设置为 2。如

表 4 所示,实验首先用 WDSR 中最小残差单元代替 GAMMNet 中的 MRU,且去除 GAG 模块作为基准训练 2 倍超分模型,然后分别添加 MRU 模块、GAG 模块进行训练,最后训练 GAMMNet(同时使用所提出的两种模块)。表 4 列出实验结果,可以看出 MRU 和 GAG 模块分别使得网络 PSNR 指标提升 0.1 dB 和 0.08 dB,且同时使用两种模块可以在 PSNR 和 SSIM 上取得最佳性能,从而验证了本文所提两种模块的有效性。

表 4 GAMMNet 的消融实验
Table 4 Ablative study of GAMMNet

Baseline	MRU	GAG	PSNR/dB	SSIM
✓			38.02	0.950 5
✓	✓		38.12	0.960 5
✓		✓	38.10	0.960 2
✓	✓	✓	38.16	0.961 0

注:“✓”分别表示 Baseline、MRU 和 GAG 的选用,加粗字体为最优结果。

3 结 论

本文提出一种全局注意力门控残差记忆网络来

实现精确的图像超分重建。首先通过特征提取模块提取网络浅层特征。然后,设计一种嵌套的递归结构实现网络非线性映射来学习鲁棒的深层特征。该结构联合网络不同尺度的特征,有效地学习其中各级特征图的上下文信息,同时自适应融合不同层级的输出,充分挖掘超分重建所需的丰富的高频细节信息,缓解了网络高层特征高频细节丢失严重的问题。最后,在重建模块部分,并联不同尺度特征并利用像素重组的方法实现图像放大。本文设计的 GAMMNet 在图像超分重建的 5 个常用基准测试数据集 (Set5, Set14, B100, Urban100 和 Manga109) 上进行了评估,实验表明 GAMMNet 在 PSNR 和 SSIM 上相比现有的系列先进方法更具优越性。

此外,现有的图像超分重建技术虽然取得了较高的精度,但是通常是以牺牲速度为代价,难以成功应用于一些对速度要求较高的实际任务中,因此,如何权衡好超分网络精度与速度之间的关系,从而构建精确且快速的轻量型网络模型是本文后续进一步研究的方向。

参考文献 (References)

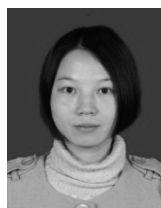
- Ahn N, Kang B and Sohn K A. 2018. Fast, accurate, and lightweight super-resolution with cascading residual network//Proceedings of the 15th European Conference on Computer Vision. Munich, Germany: Springer; 256-272 [DOI: 10.1007/978-3-030-01249-6_16]
- Bevilacqua M, Roumy A, Guillemot C and Albirola-Morel M L. 2012. Low-complexity single-image super-resolution based on nonnegative neighbor embedding//Proceedings of British Machine Vision Conference. Surrey, UK: BMVC; 221-231 [DOI: 10.5244/C.26.135]
- Cao Y, Xu J R, Lin S, Wei F Y and Hu H. 2019. GCNet: non-local networks meet squeeze-excitation networks and beyond//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision Workshop. Seoul, Korea (South): IEEE; 1971-1980 [DOI: 10.1109/ICCVW.2019.00246]
- Dong C, Loy C C, He K M and Tang X O. 2014. Learning a deep convolutional network for image super-resolution//Proceedings of the 13th European Conference on Computer Vision. Zurich, Switzerland: Springer; 184-199 [DOI: 10.1007/978-3-319-10593-2_13]
- Hu J, Shen L and Sun G. 2018. Squeeze-and-excitation networks//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE; 7132-7141 [DOI: 10.1109/CVPR.2018.00745]
- Huang J B, Singh A and Ahuja N. 2015. Single image super-resolution from transformed self-exemplars//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA: IEEE; 5197-5206 [DOI: 10.1109/CVPR.2015.7299156]
- Kim J, Lee J K and Lee K M. 2016. Deeply-recursive convolutional network for image super-resolution//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE; 1637-1645 [DOI: 10.1109/CVPR.2016.181]
- Kingma D P and Ba J. 2015. Adam, the Netherlands: a method for stochastic optimization//Proceedings of the 3rd International Conference on Learning Representations. Amsterdam, the Netherlands: Amsterdam Machine Learning lab; 1-15
- Li J C, Fang F M, Mei K F and Zhang G X. 2018. Multi-scale residual network for image super-resolution//Proceedings of the 15th European Conference on Computer Vision. Munich, Germany: Springer; 527-542 [DOI: 10.1007/978-3-030-01237-3_32]
- Li X G, Sun Y M, Yang Y L and Miao C Y. 2018. Image super-resolution reconstruction based on intermediate supervision convolutional neural networks. *Journal of Image and Graphics*, 23(7): 984-993 (李现国, 孙叶美, 杨彦利, 苗长云. 2018. 基于中间层监督卷积神经网络的图像超分辨率重建. *中国图象图形学报*, 23(7): 984-993) [DOI: 10.11834/jig.170538]
- Li Y X, Deng H P, Xiang S, Wu J and Zhu L. 2018. Depth map super-resolution reconstruction based on the texture edge-guided approach. *Journal of Image and Graphics*, 23(10): 1508-1517 (李宇翔, 邓慧萍, 向森, 吴谨, 朱磊. 2018. 纹理边缘引导的深度图像超分辨率重建. *中国图象图形学报*, 23(10): 1508-1517) [DOI: 10.11834/jig.180127]
- Martin D, Fowlkes C, Tal D and Malik J. 2001. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics//Proceedings of the 8th IEEE International Conference on Computer Vision. Vancouver, Canada: IEEE; 416-423 [DOI: 10.1109/ICCV.2001.937655]
- Matsui Y, Ito K, Aramaki Y, Fujimoto A, Ogawa T, Yamasaki T and Aizawa K. 2017. Sketch-based manga retrieval using Manga109 dataset. *Multimedia Tools and Applications*, 76(20): 21811-21838 [DOI: 10.1007/s11042-016-4020-z]
- Paszke A, Gross S, Chintala S, Chanan G, Yang E, DeVito Z, Lin Z M, Desmaison A, Antiga L and Lerer A. 2017. Automatic differentiation in pytorch//Proceedings of the 31st Conference on Neural Information Processing Systems. Long Beach, USA: NIPS; 1-4
- Shen M Y, Yu P F, Wang R G, Yang J and Xue L X. 2019. Image super-resolution reconstruction via deep network based on multi-staged fusion. *Journal of Image and Graphics*, 24(8): 1258-1269 (沈明玉, 俞鹏飞, 汪荣贵, 杨娟, 薛丽霞. 2019. 多阶段融合网络的图像超分辨率重建. *中国图象图形学报*, 24(8): 1258-1269) [DOI: 10.11834/jig.180619]
- Shi W Z, Caballero J, Huszar F, Totz J, Aitken A P, Bishop R, Rueckert D and Wang Z H. 2016. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural net-

- work//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 1874-1883 [DOI: 10.1109/CVPR.2016.207]
- Song H H, Liu Q S, Wang G J, Hang R L and Huang B. 2018. Spatio-temporal satellite image fusion using deep convolutional neural networks. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11 (3): 821-829 [DOI: 10.1109/JSTARS.2018.2797894]
- Tai Y, Yang J and Liu X M. 2017a. Image super-resolution via deep recursive residual network//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 2790-2798 [DOI: 10.1109/CVPR.2017.298]
- Tai Y, Yang J, Liu X M and Xu C Y. 2017b. Memnet: a persistent memory network for image restoration//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE: 4549-4557 [DOI: 10.1109/ICCV.2017.486]
- Timofte R, Agustsson E, Van Gool L, Yang M H, Zhang L, Lim B, Son S, Kim H, Nah S, Lee K M, Wang X T, Tian Y P, Yu K, Zhang S X, Dong C, Lin L, Qiao Y, Loy C C, Bae W, Yoo J, Han Y, Ye J C, Choi J S, Kim M, Fan Y C, Yu J H, Han W, Liu D, Yu H C, Wang Z Y, Shi H H, Wang X C, Huang T S, Chen Y J, Zhang K, Zuo W M, Tang Z M, Luo L K, Li S H, Fu M, Cao L, Heng W, Bui G, Le T, Duan Y, Tao D C, Wang R X, Lin X, Pang J X, Xu J C, Zhao Y, Xu X Y, Pan J S, Sun D Q, Zhang Y J, Song X B, Dai Y C, Qin X Y, Huynh X P, Guo T T, Mousavi H S, Vu T H, Monga V, Cruz C, Egiazarian K, Katkovnik V, Mehta R, Jain A K, Agarwalla A, Praveen C V S, Zhou R F, Wen H D, Zhu C, Xia Z Q, Wang Z T and Guo Q. 2017. NTIRE 2017 challenge on single image super-resolution: methods and results//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops. Honolulu, USA: IEEE: 1110-1121 [DOI: 10.1109/CVPRW.2017.149]
- Timofte R, Rothe R and Van Gool L. 2016. Seven ways to improve example-based single image super resolution//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 1865-1873 [DOI: 10.1109/CVPR.2016.206]
- Tong J C, Fei J L, Chen J S, Li H and Ding D D. 2019. Multi-level feature fusion image super-resolution algorithm with recursive neural network. *Journal of Image and Graphics*, 24 (2): 302-312 (佟骏超, 费加罗, 陈靖森, 李恒, 丁丹丹. 2019. 递归式多阶特征融合图像超分辨率算法. *中国图象图形学报*, 24 (2): 302-312) [DOI: 10.11834/jig.180410]
- Wang C F, Li Z and Shi J. 2019. Lightweight image super-resolution with adaptive weighted learning network[EB/OL]. [2020-04-18]. <https://arxiv.org/pdf/1904.02358.pdf>
- Wang Z, Bovik A C, Sheikh H R and Simoncelli E P. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13 (4): 600-612 [DOI: 10.1109/TIP.2003.819861]
- Yang J C, Wright J, Huang T S and Ma Y. 2010. Image super-resolution via sparse representation. *IEEE Transactions on Image Processing*, 19 (11): 2861-2873 [DOI: 10.1109/TIP.2010.2050625]
- Yu J, Fan Y, Yang J, Xu N, Wang Z, Wang X and Huang T. 2018. Wide activation for efficient and accurate image super-resolution. [EB/OL]. [2020-04-18]. <https://arxiv.org/pdf/1808.08718v1.pdf>
- Zhang Y L, Li K P, Li K, Wang L C, Zhong B E and Fu Y. 2018. Image super-resolution using very deep residual channel attention networks//Proceedings of the 15th European Conference on Computer Vision. Munich, Germany: Springer: 294-310 [DOI: 10.1007/978-3-030-01234-2_18]
- Zou W W and Yuen P C. 2012. Very low resolution face recognition problem. *IEEE Transactions on Image Processing*, 21 (1): 327-340 [DOI: 10.1109/TIP.2011.2162423]

作者简介



王静, 1995年生, 女, 硕士研究生, 主要研究方向为超分辨率重建。
E-mail: 1604557385@qq.com



宋慧慧, 通信作者, 女, 教授, 主要研究方向为遥感图像处理。
E-mail: songhuihuis@nuist.edu.cn

张开华, 男, 教授, 主要研究方向为图像分割与目标跟踪。
E-mail: zhkhua@gmail.com
刘青山, 男, 教授, 主要研究方向为视频内容分析与理解。
E-mail: qslu@nuist.edu.cn