

文章编号:1674-2869(2018)04-0440-06

多尺度残差深度神经网络的卫星图像超分辨率算法

汪家明, 卢涛*

武汉工程大学计算机科学与工程学院, 湖北 武汉 430205

摘要: 卫星图像实现星际对地观测并被广泛的应用到了军事和经济生活领域。受到星载成像设备和星地通讯带宽的限制, 卫星图像的地面分辨率常不能完全满足目标识别与分析的需求。卫星图像的成像幅度宽且范围广, 地面目标的尺度变化大、纹理信息多样化, 给现有图像超分辨率技术带来了新的挑战。针对卫星图像的多尺度特性, 提出了一种多尺度残差深度神经网络, 首先提取低分辨率卫星图像的多尺度特征, 对不同尺度特征建立自适应深度神经网络, 然后使用融合网络进行残差融合, 融合不同尺度高频信息, 最终生成高分辨卫星图像。在SpaceNet卫星图像数据集中的实验结果证明了本文算法的优越性。

关键词: 卫星图像; 超分辨率; 残差网络; 残差学习; 卷积神经网络

中图分类号: TP391.4 文献标识码: A doi: 10.3969/j.issn.1674-2869.2018.04.018

Satellite Imagery Super-Resolution Algorithm via Multi-Scale Residual Deep Neural Network

WANG Jiaming, LU Tao*

School of Computer Science and Engineering, Wuhan Institute of Technology, Wuhan 430205, China

Abstract: Satellite imagery realizes interstellar-earth observations, which is widely used in military and economic fields. Because the performances of satellite-borne imaging equipment and the band width of satellite communications system are limited, the resolution of ground targets in satellite images are often low, thus they cannot fully meet the needs of target identification and analysis. Moreover, satellite images have three features: wide range of imaging, variation of multi-scale of ground targets, and diversification of texture information, which bring new challenges to the existing super-resolution algorithms. Using the multi-scale nature of satellite image, a multi-scale residual neural network was proposed in this paper for accurately reconstructing the multi-scale information. Firstly, different scale features of low-resolution satellite images were extracted, then for each scale-level, an adaptive deep residual neural network was developed for better reconstruction performance. Then a fusion network was used to refine different scales of residual information. The proposed fusion network fuses high-frequency information of different scales to output the target high-resolution satellite image. Experimental results over the SpaceNet satellite image database prove the superiority of the proposed algorithm.

Keywords: satellite image; super-resolution; residual network; multi-scale image; convolutional neural network

收稿日期: 2018-04-16

基金项目: 国家自然科学基金(61502354, 61671332, 41501505); 湖北省自然科学基金(2015CFB451, 2014CFA130, 2012FFA099, 2012FFA134, 2013CF125); 武汉工程大学科研基金(K201713)

作者简介: 汪家明, 硕士研究生。E-mail: 549880890@qq.com

*通讯作者: 卢涛, 博士, 副教授。E-mail: lut@wit.edu.cn

引文格式: 汪家明, 卢涛. 基于多尺度残差深度神经网络的卫星图像超分辨率算法[J]. 武汉工程大学学报, 2018, 40(4): 440-445.

遥感卫星是一种重要的对地观测手段,能够对地面进行大面积的同步观测以获得地面目标的信息。遥感卫星图像在灾害监测预警、资源勘探、和环境检测等领域有着广泛的应用前景。然而,由于硬件设备与成本的限制,遥感图像的分辨率固定。此外,大气湍流、成像过程噪声、设备与景物之间的运动模糊、光学传感器的欠采样,降低了卫星图像的分辨率和质量。为了提高遥感影像的空间分辨率,克服噪声和成像模糊的影响,对遥感图像的超分辨率重建显得尤为迫切。

现有的图像超分辨率重建算法(super-resolution, SR)主要分为2类:基于重建的算法和基于学习的算法。基于重建的算法通过多个低分辨率图像来融合亚像素精度的多帧信息^[1],重建出更高分辨率的图像。但基于重建的方法仅能利用低分辨率图像之间的互补信息,不能增加新的高频信息,因此这类重建方法的放大倍数有限。然而后续的认识任务更加关注高频信息中包含的特征^[2]。近年来,受到机器学习理论发展的影响,基于学习的方法受到高度的重视日益成为超分辨率算法研究的主流方向。基于学习的算法通过训练样本提供的先验信息,学习低分辨率图像块与高分辨率图像块之间的映射模型。从训练样本的使用角度来看,基于学习的算法可以分为3类:基于表达的超分辨率算法,基于回归的超分辨率算法和基于深度学习的超分辨率算法。

基于表达的算法根据高分辨图像在不同分辨率空间表达系数不变这一性质,使用低分辨字典表示输入图像块,并找寻低分辨率字典对应的高分辨率字典进行重建^[3-4]。基于回归的算法通过回归模型对超分重建进行求解^[5-6]。考虑到卷积神经网络(convolutional neural network, CNN)的优秀学习能力,通过卷积神经网络^[7-8]来学习低分辨率和高分辨图像之间的复杂非线性映射关系,通过多个隐藏层来扩展传统网络的深度。Shi等^[9]构建了亚像素空间的卷积神经网络,直接而高效地学习了从低分辨率图像到高分辨率图像的映射函数。Kim等^[10]通过深度残差网络证明了深度残差网络层数对于超分辨率性能有提升。生成对抗网络^[11]由生成网络和判别网络组成,通过生成网络生成细节,取得了良好的视觉效果,但是客观指标较低。Luo等^[12]在Kim^[10]的基础上根据自相似性填充图像取代零填充,避免增加无用的信息,在卫星图像超分辨率任务中表现出了优异的重建效果。

卫星图像超分辨率是特定领域的超分辨率重

建问题,准确的边缘和纹理等高频信息对于卫星图像检测和识别任务是至关重要的^[13]。待识别目标的全局高频信息在检测任务中占据主导地位,而局部的高频信息在识别任务中不可或缺。但是遥感卫星图像成像范围广,目标的尺度差异较大,直接使用同构同质的网络预测重建不同尺度的目标,导致重建精度不满足要求或者效率低下。而现有的基于多尺度的超分辨率算法中,多尺度信息融合是在像素域或特征域进行,重建结果注重的是图像的内容和局部高频信息,却忽略了全局的高频信息。因此需要一个网络结构,可以灵活处理不同尺度的目标和不同精度的超分辨率任务,预测出不同尺度目标的边缘和纹理等高频信息,并在更加快速和有效的残差空间融合多尺度信息。综合考虑上述问题,本文提出了一个多尺度残差深度神经网络的卫星图像超分辨率算法。

1 基于多尺度残差深度神经网络的卫星图像超分辨率

结合多尺度图像分析和深度残差网络,将同一卫星图像以多尺度的形式放入不同的深度残差神经网络中,建立不同尺度特征的自适应深度神经网络,学习从底层到高层的特征,预测不同尺度空间的高频残差信息。然后通过残差融合网络将预测所得的不同尺度空间的残差图像在残差空间进行加权,进行残差信息之间的互补和融合,使得不同目标能够在最合适的空间进行表达。本文中的多尺度残差网络以两个不同尺度空间子网络为例,分别为尺度空间1和尺度空间2,网络结构图如图1所示。

对于训练样本为 $\{\bar{x}_i, y_i\}_{i=1}^M$,其中低分辨图像 $\bar{x}_i \in \mathbb{R}^{m \times n}$,高分辨率图像 $y_i \in \mathbb{R}^{mt \times nt}$, t 为下采样倍数,本文将低分辨率图像插值到高分辨率图像大小,则训练样本为 $\{x_i, y_i\}_{i=1}^M$ 。本文中使用上标表示网络索引,下标表示网络中的层数索引。其中上标 $S1, S2, F$ 分别代表尺度空间1的残差重建网络、尺度空间2的残差重建网络和残差融合网络。

1.1 残差学习

Kim^[10]证明了随着网络层数的增加,大量的超参数促进了网络的非线性表达能力。然而,网络层数的单一级联带了一个不可避免的问题:梯度消失/梯度爆炸导致在反向传播中网络的参数无法优化^[14]。由于输入的低分辨率图像和原始高分辨率图像非常相似,本文定义残差图像为 $r_i = y_i - x_i$ 。残

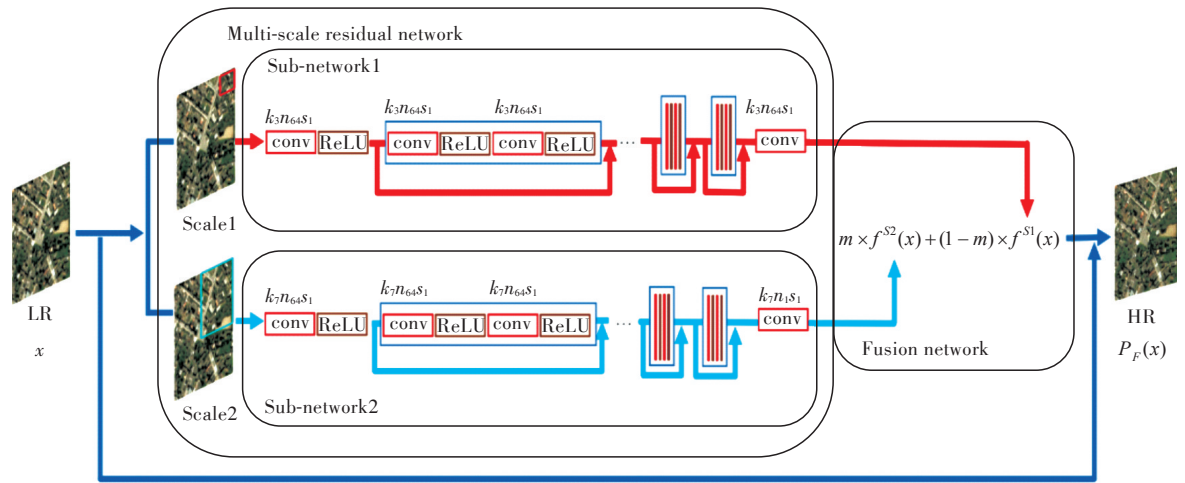


图 1 多尺度残差网络的网络结构。网络包括多尺度残差学习网络和残差融合网络两部分
(k 表示卷积核大小, n 表示卷积核个数, s 表示步长)

Fig. 1 Network architecture of multi-scale residual deep neural network. The network includes two subnetworks: multi-scale residual network and residual fusion network. (k is the convolution kernel size, n is the number of convolution kernel and s is stride size)

差图像的像素值大多数为零或者较小值。低分辨率图像 \bar{x} 的插值结果 x_i 属于原始图像 y_i 的低频分量,因此残差图像 r_i 为 y_i 的高频分量。相比直接预测像素域的图像,使用残差学习不仅会大量减少网络携带的信息,可加快网络收敛速度,还可以避免深度网络的不收敛问题^[10]。

1.2 多尺度特征提取

在尺度空间 1 残差学习网络中,本文使用大小为 41×41 像素的重叠图像块进行训练。本文定义卷积核大小为 3×3 。网络输出的残差为:

$$f^{S1}(x) = W_{20}^{S1} * H_{19}^{S1}(x) + b_{20}^{S1} \tag{1}$$

其中 $f^{S1}(x)$ 为尺度空间 1 残差学习网络预测所得的残差。 $H_{19}^{S1}(x)$ 为尺度空间 1 残差学习网络第 19 层输出的特征图。卷积的权值 W_{20}^{S1} 大小为 $64 \times 3 \times 3 \times 1$,偏置项 b_{20}^{S1} 大小为 1×1 。

在尺度空间 2 残差学习网络中使用 100×100 的非重叠图像块。由于输入图像块大小的增加,所以需要更大的感受野。本文定义卷积核大小为 7×7 ,尺度空间 1 和尺度空间 2 残差空间增强网络相似。网络输出的残差为:

$$f^{S2}(x) = W_{20}^{S2} * H_{19}^{S2}(x) + b_{20}^{S2} \tag{2}$$

其中 $f^{S2}(x)$ 为尺度空间 2 残差学习网络预测所得的残差。 $H_{19}^{S2}(x)$ 为残差学习网络第 19 层输出的特征图。卷积的权值 W_{20}^{S2} 大小为 $64 \times 7 \times 7 \times 1$,偏置项 b_{20}^{S2} 大小为 1×1 。

1.3 残差融合网络

相同尺度的目标在不同的尺度空间具有不同的表达形式^[15],本文通过不同尺度空间找寻不同尺度的目标的最适合的表达形式,并通过多尺度的表达形式进行信息互补,保证了目标表达的准确性。本文试图通过融合模型组合尺度空间 1 的残差信息和尺度空间 2 的残差信息。融合后的残差为:

$$f^F(x) = m \times f^{S2}(x) + (1 - m) \times f^{S1}(x) \tag{3}$$

其中 m 表示尺度空间 2 预测残差的权重,在本文中 $m = 0.9$ 。最终的重建图像为:

$$P^F(x) = f^F(x) + x \tag{4}$$

1.4 损失函数

使用均方误差(mean squared error, MSE)作为网络的损失函数,整体的损失函数为:

$$L = \alpha \left\| y - x - f^{S1}(x) \right\|_2^2 + \beta \left\| y - x - f^{S2}(x) \right\|_2^2 \tag{5}$$

其中 α, β 控制不同尺度残差学习网络之间的平衡,在本文中 $\alpha = \beta = 1$ 。网络优化过程中使用梯度下降法来进行反向传播。每次卷积操作都会缩小特征图的大小,通过填充零来保留更多的边缘像素,以更准确地推断中心像素,同时也确保了所有的特征图大小相同。网络损失如图 2 所示,由于 2 个子网络训练使用的图像块大小差异,所以对应的网络训练时的损失在量级有差异,但收敛趋势相同,证明了使用不同尺度的残差网络恢复残差的合理性。

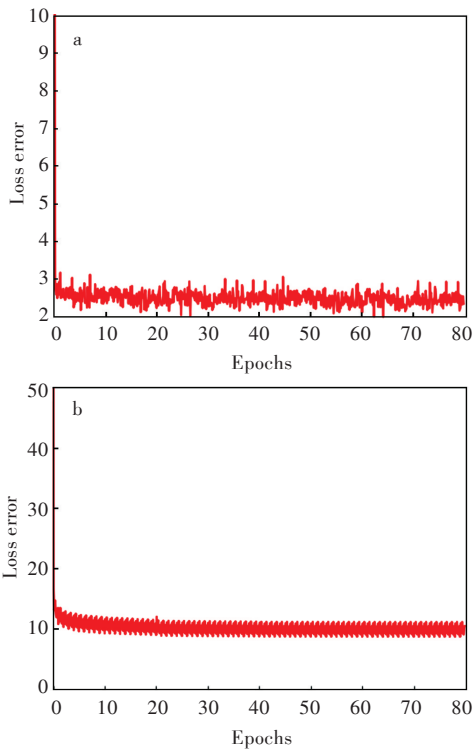


图2 不同尺度残差网络的训练损失图:
(a) 尺度1, (b) 尺度2

Fig. 2 Training loss of multi-scale residual networks:
(a) scale 1, (b) scale 2

2 结果与讨论

2.1 实验数据与参数

本文在 SpaceNet 卫星图像数据集中进行实验,并通过峰值信噪比(peak signal to noise ratio, PSNR)本文中定义为 R_{PSNR} 、结构相似性(structural similarity, SSIM)和互信息(mutual information, MI)作为重建图片质量的客观评估标准。PSNR 和 SSIM 从图像的角度刻画了重建图像与原始图像之间的相似性,而互信息可从信息的角度表示两幅图像之间的信息的依赖程度,互信息评分越高,图像之间的依赖越高,图像之间的相似性也越高。图像 x 与 y 之间的互信息定义为:

$$H(x, y) = \sum_{m \in x} \sum_{n \in y} p(m, n) \log \frac{p(m, n)}{P(m)P(n)} \quad (6)$$

式(6)中, m 和 n 表示灰度值, $P(m)$ 表示灰度值为 m 的像素数量占整幅图像的比例。

SpaceNet 卫星图像数据集中包括里约热内卢、巴黎、拉斯维加斯、上海和喀土穆 5 个可用的区域,每个地图图像都由完整的高分辨率卫星图像分割为 438×406 像素的 RGB 图像,图像的兴趣内

容主要为建筑和道路等。从里约热内卢地图中随机选取 240 张建筑群图像,其中的 160 张作为训练集,40 张作为验证集,余下 40 张作为测试样本。

本文算法网络为并行网络结构,每个子网络都需要训练 80 个时期。由于网络层数较深,因此本文算法使用学习率衰减,学习率初始化为 0.1,每隔 20 个时期学习率便会下降至原来的 1/10,而网络的动量为 0.9。为了避免过拟合,本文使用 ℓ_2 正则化,其权值衰减为 0.000 1。在尺度空间 1 残差学习网络中,本文设置步长为 1,填充为 1,而尺度空间 2 残差学习网络中,设置步长为 1,填充为 3。本文使用 msra 来初始化权重,即满足均值为 0 方差为 $\frac{2}{n}$ 的高斯分布(n 输入训练样本的个数),使用 constant 来初始化偏置项,即初始值为 0。实验先将 RGB 图像转换到 YCbCr 颜色空间,在 Y 通道中进行重建,重建完成后再将 Y 通道的图像还原到 RGB 颜色空间。

2.2 实验结果与分析

对结果进行定量和定性的分析,并使用 PSNR、SSIM 和 MI 用于评估重建的结果进行定量分析。此外为了验证本文算法网络结构的有效性,比较了 CNN 中前沿的算法:加速的基于卷积网络的图像超分辨率算法(aueclerating the super_resolution convolutional neural network, FSRCNN)^[16],经典深度学习算法基于深度卷积网络的图像超分辨率算法(image super-resolution using deep convolutional networks, SRCNN)^[7]的加速优化版本,拉普拉斯金字塔超分辨率网络(laplacian pyramid super-resolution networks, LapSRN)^[17],视频卫星图像超分辨率算法(video satellite imayery super resolution, VISR)^[12]。本文中使用的对比算法 FSRCNN 的源代码由论文作者 Chao Dong 提供, LapSRN 的源代码由论文作者 Wei-Sheng Lai 提供, VISR 的源代码由论文作者 Yimin Luo 提供。所有测试图像的 PSNR 和 SSIM 如图 3 所示(由于 FSRCNN 的客观评分过低,并未在图 3 中展示),本文算法重建图像的 PSNR 和 SSIM 都是最优的,如表 1 所示,本文算法无论是 PSNR、SSIM 还是 MI 都获得了最高的客观评分。与性能排名第二的算法 VISR 相比,分别提升了 0.25 dB, 0.01 和 0.03,表明了多尺度的残差网络通过不同尺度空间更有效地学习图像内容和结构,以及通过更深网络恢复图像的信息的有效性。

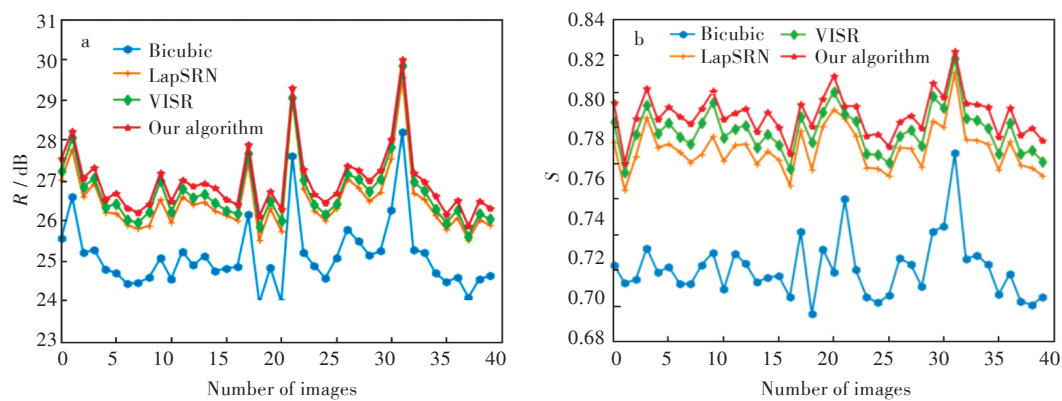


图3 SpaceNet卫星图像超分辨率结果的客观对比: (a)PSNR, (b)SSIM

Fig. 3 Objective results of super-resolution algorithms over SpaceNet satellite images: (a)PSNR, (b)SSIM

表 1 SpaceNet数据集上 PSNR,SSIM 和 MI 的平均结果

Tab. 1 Average results of PSNR, SSIM and MI on SpaceNet dataset

	双三次插值	FSRCNN	LapSRN	VISR	本文算法
R_{PSNR}/dB	25.13	14.26	26.52	26.74	26.99
S_{SSIM}	0.701 3	0.193 0	0.768 5	0.779 3	0.788 3
H_{MI}	1.11	0.20	1.31	1.33	1.36

注: R_{PSNR} 为峰值信噪比, S_{SSIM} 为结构相似性, H_{MI} 为互信息

2.2.1 本文算法与FSRCNN对比 FSRCNN使用亚像素空间来取代SRCNN在像素域的计算,但是FSRCNN中输入的低分辨率图像块大小为 11×11 像素,对应的高分辨率图像块大小为 19×19 像素,因此并不适用于尺度差异较大的卫星图像,如图4所示,FSRCNN有较好的重建效果,但是由于尺度问题,导致重建图4(b)中图像块拼接处不平滑,客观评分较低。



图4 SpaceNet卫星图像超分辨率结果的主观对比: (a)双三次插值, (b)FSRCNN, (c)LapSRN, (d)VISR, (e)本文算法, (f)高分辨图像

Fig. 4 Subjective results of different super-resolution algorithms over SpaceNet satellite images: (a)bicubic, (b)FSRCNN, (c)LapSRN, (d)VISR, (e) our algorithm, (f) high resolution image

2.2.2 本文算法与 LapSRN 对比 LapSRN 为多尺度的深度学习算法,但是 LapSRN 的多尺度信息融合主要是在训练网络中的特征域进行。而本文算法则是通过构建并行网络,在残差域中进行融合。残差空间与图像的高频信息之间联系更为紧密,且残差域中数值较小,携带的信息量较小,能够有更高的计算效率。

2.2.3 本文算法与 VISR 对比 VISR 通过对低分辨率图像进行自相似填充,来避免网络中的零填充带来无用的信息,但是也因此将网络改变为像素域。像素域中的重建会更多的关注于图像中的低频信息内容,而卫星图像超分辨率任务中更希望恢复的是高频信息,且在残差域网络中更为高效。

由图4图像重建客观结果来看,该算法的重建图像比其他算法的结果包含更多的细节信息。使用本文的算法重建的高分辨率图像拥有更多的高频信息。例如,对于小目标建筑,本文的算法使房屋的轮廓更清晰;对于大目标建筑,本文的算法在屋顶的纹理方面有更好的表现;对于建筑间的道路,本文的方法更清晰的重建出了道路中的细纹。总之,由本文的结果证明,使用多尺度的残差网络有助于卫星图像的超分辨复原。

3 结 语

考虑到卫星图像的特殊性,使用多尺度学习来提高卷积神经网络对遥感卫星图像不同尺度先验信息的适应能力,通过融合多尺度残差得到更准确的边缘和纹理等高频信息。在 SpaceNet 数据集上的实验结果表明:提出的多尺度残差深度神经网络能有效地增强重建图像中的高频信息,相比前沿的卫星图像而言具有更好的主客观重建质量。

参考文献:

[1] 杨欣,费树岷,周大可. 基于 MAP 的自适应图像配准及超分辨率重建[J]. 仪器仪表学报, 2011, 32(8): 1771-1775.

[2] 夏平平,吕太之. 动态人脸识别系统的设计与实现[J]. 武汉工程大学学报, 2011, 33(10):107-110.

[3] 赵娜,赵彤洲,邹冲,等. 稀疏表示中字典学习的影响因子研究[J]. 武汉工程大学学报, 2017, 39(3): 267-272.

[4] 肖杰,卢涛,万永静,等. 基于稀疏表达的图像超分辨率算法实现[J]. 计算机与数字工程, 2014(11): 2160-2163.

[5] 王宏,卢芳芳,李建武. 结合支持向量回归和图像自

相似的单幅图像超分辨率算法[J]. 中国图象图形学报, 2016, 21(8):986-992.

[6] 吴炜,杨晓敏,陈默,等. 基于偏最小二乘算法的人脸图像超分辨率技术[J]. 光子学报, 2009, 38(11): 3025-3033.

[7] DONG C, LOY C C, HE K, et al. Image super-resolution using deep convolutional networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 38(2): 295-307.

[8] LU T, WANG H, XIONG Z, et al. Face hallucination using region-based deep convolutional networks [C]// IEEE International Conference on Image Processing. Beijing: IEEE, 2017:1657-1661.

[9] SHI W, CABALLERO J, HUSZÁR F, et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016:1874-1883.

[10] KIM J, KWON L J, MU L. Accurate image super-resolution using very deep convolutional networks[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 1646-1654.

[11] LEDIG C, THEIS L, HUSZÁR F, et al. Photo-realistic single image super-resolution using a generative adversarial network [J]. Computer Vision and Pattern Recognition, 2017(2):105-114.

[12] LUO Y, ZHOU L, WANG S, et al. Video satellite imagery super resolution via convolutional neural networks [J]. IEEE Geoscience & Remote Sensing Letters, 2017, 14(12):2398-2402.

[13] 宋晓宇,单新建. 高分辨率卫星影像在城市建筑物识别中的初步应用[J]. 遥感信息, 2002(1):27-31.

[14] HE K, ZHANG X, REN S, et al. Deep Residual Learning for Image Recognition [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016:770-778.

[15] 卢涛,章瑾,陈白帆,等. 多尺度自适应配准的视频超分辨率算法[J]. 武汉工程大学学报, 2016, 38(2):178-184.

[16] DONG C, LOY C C, TANG X. Accelerating the super-resolution convolutional neural network [C]// European Conference on Computer Vision. Amsterder: Springer, 2016: 391-407.

[17] LAI W S, HUANG J B, AHUJA N, et al. Deep laplacian pyramid networks for fast and accurate super-resolution [C]//Proc. IEEE Conf. Comput. Vis. Pattern Recognition. 2017: 624-632.