Accepted Manuscript

Single Satellite Imagery Simultaneous Super-resolution and Colorization using Multi-task Deep Neural Networks

Heng Liu, Zilin Fu, Jungong Han, Ling Shao, Hongshen Liu

PII: S1047-3203(18)30048-8
DOI: https://doi.org/10.1016/j.jvcir.2018.02.016
Reference: YJVCI 2147

To appear in: *J. Vis. Commun. Image R.*

Received Date: 24 November 2017
Accepted Date: 28 February 2018

Please cite this article as: H. Liu, Z. Fu, J. Han, L. Shao, H. Liu, Single Satellite Imagery Simultaneous Super-resolution and Colorization using Multi-task Deep Neural Networks, *J. Vis. Commun. Image R.* (2018), doi: https://doi.org/10.1016/j.jvcir.2018.02.016

Single Satellite Imagery Simultaneous Super-resolution and Colorization using Multi-task Deep Neural Networks

Heng Liu[a], Zilin Fu[a], Jungong Han[b], Ling Shao[c], Hongshen Liu[a]

[a] *Anhui University of Technology, China, 243032*
[b] *Lancaster University, U.K., LA1 4YW*
[c] *University of East Anglia, U.K., NR4 7TJ*

Abstract

Satellite imagery is a kind of typical remote sensing data, which holds preponderance in large area imaging and strong macro integrity. However, for most commercial space usages, such as virtual display of urban traffic flow, virtual interaction of environmental resources, one drawback of satellite imagery is its low spatial resolution, failing to provide the clear image details. Moreover, in recent years, synthesizing the color for grayscale satellite imagery or recovering the original color of camouflage sensitive regions becomes an urgent requirement for large spatial objects virtual reality interaction. In this work, unlike existing works which solve these two problems separately, we focus on achieving image super-resolution (SR) and image colorization synchronously. Based on multi-task learning, we provide a novel deep neural network model to fulfill single satellite imagery SR and colorization simultaneously. By feeding back the color feature representations into the SR network and jointly optimizing such two tasks, our deep model successfully achieves the mutual cooperation between imagery reconstruction and image colorization. To avoid color bias, we not only adopt the non-satellite imagery to enrich the color diversity of satellite image, but also recalculate the prior color distribution and the valid color range based on the mixed data. We evaluate the proposed model on satellite images from different data sets, such as RSSCN7 and AID. Both the evaluations and comparisons reveal that the proposed multi-task deep learning approach is superior to the state-of-the-art methods, where image SR and colorization can be accomplished simultaneously and efficiently.

Keywords: Image Super-resolution; Satellite image colorization; Deep neural networks; Multi-task learning

---

1. Introduction

1.1. Motivation

Remote sensing satellite imagery holds the characteristics of extensive coverage, strong macro integrity and consistent imaging scales, which can be widely used for the spatial information related virtual reality applications, such as resource survey virtual interactions, urban traffic virtual analysis, climate change virtual display, and military action virtual deduction. However, due to optical device and imaging sensor limitations coupled with the extreme distance between sensor and sensed object on earth, one natural drawback of satellite imagery is that the spatial resolution is always low, which leads to the inaccurate sensing data due to lack of image details.

On the contrary, high-resolution (HR) satellite imagery, which is very helpful for the realization of large scale spatial information virtual reality (VR), allows extracting the rich details and accurate information at multi-level scales. In order to improve the spatial resolution of the satellite images, the traditional hardware handling method reduces the physical sizes of the charge-coupled device (CCD) or complementary metal oxide semiconductor (CMOS) sensors among sensor fabrication procedure, which will easily generate shot noise that severely degrades the image quality (Yang and Huang, 2010). In addition, manufacturing imaging chips and optical instruments to capture very high-resolution images will cost huge. Thus, it is necessary to exploit signal processing techniques to reconstruct the high-resolution (HR) images from the degraded low-resolution (LR) remote sensing images, which is specifically referred to as satellite imagery SR.

In addition to low resolution, the color of satellite imagery can be easily faded due to inappropriate illumination, exposure and storage. Moreover, satellite imagery sometimes even cannot reflect the actual original color of the observed targets. For example, intentional camouflage is a common means of visually hiding the military facilities or important infrastructure, where the color as well as the appearance of these special targets are always altered and disguised. In order to get clear and accurate knowledge of these objects, recoloring the disguised imagery and enhancing their spatial resolution at the same time become a pressing demand. Therefore, it is necessary to solve these two problems - imagery SR and colorization - simultaneously in one framework. It should be noted that for some military applications, the multi-task simultaneous imagery SR and colorization perhaps is especially significant, for example solider combat VR glasses. Here, to keep consistency with human visual perception, the word ‘colorization’ throughout the whole text only refers to the visible light 3-bands color operations.

1.2. Related work

Over the past five years, a considerable number of image SR works have addressed to reconstruct HR satellite imagery from LR inputs. Usually, these methods are divided into two categories: multiple images reconstruction (Pickup, 2007; Zhang et al., 2014; Hung et al., 2016; Zhu et al., 2016; Brodu, 2016; Alvarez-Ramos et al., 2016) or single image SR (Liebel and Körner, 2016; Patrick, 2016).

Bayesian machine learning method was firstly applied for multi-frame super-resolution (Pickup, 2007), which fully utilizes a prior distribution over the super-resolution image. This Bayesian inference method was improved with variation approximation (Hung et al., 2016) to estimate the distribution of HR satellite imagery, the registration parameter, and some other hyper-parameters. Due to possible resolution differences in multi-angle remote sensing images over the same scene, adaptive weighting schemes (Zhang et al., 2014) are utilized to reconstruct HR satellite imagery. In addition, adaptive multi-scale detail enhancement measures (Zhu et al., 2016) were attempted for multiple LR satellite images SR. Moreover, sparse representation (Alvarez-Ramos et al., 2016) has been employed to deal with overlapping blocks for satellite image SR. Recently, the band-specific information is also applied in resolution enhancement for multi-spectral and multi-resolution satellite images (Brodu, 2016), where the independent reflectance of LR bands is preserved in details reconstruction.

Actually, for satellite imaging, even if it is easily to orbit to acquire multi-frame images of the same scene on a regular basis, the imaging scenes will always keep changes due to many uncontrollable factors, such as clouds or snow coverage, objects moving or seasonal alternation. Thus, if there were no available or reliable multi-frame data, single satellite imagery SR would become a more challenging problem. Fortunately, recent developments in the field of deep learning cast a bright way for single remote sensing image SR. An end-to-end CNN model (Dong et al., 2014, 2016), referred as SRCNN, has been proposed recently and successfully applied in single image SR. Then, Liebel and Körner (2016) retrain the SRCNN model for multi-spectral remote sensing imagery SR with a domain-specific data set to introduce the characteristics of multiple spectral bands. Furthermore, motivated by residual learning (He et al., 2016), Patrick (2016) proposes to construct a deep residual network for single satellite imagery SR.

On the other hand, very recently a few works (Larsson et al., 2016; Zhang et al., 2016) have exploited deep models to address the problem of image colorization which augment color from gray-scale images. These methods manage to learn the corresponding color representation or color distribution by constructing deep networks and training it with ImageNet data set (Russakovsky et al., 2015). It should be noted that such colorization methods actually carry out color remapping and do not consider keeping the reconstruction accuracy of pixels' intensity value between input and output image.

For the accurate understanding and better utilizing of low quality satellite imagery in large spatial related virtual reality applications, in this work we provide an efficient approach that not only can reconstruct HR satellite imagery from single LR input, but also is able to simultaneously colorize the grayscale satellite imagery with appropriate color information. Our contributions can be summarized as:

• We propose a multi-task deep neural model to achieve satellite imagery SR and colorization simultaneously. Our multi-task deep model contains two concurrent but not separated task networks - image features of colorization network are fed back to the beginning of the feature representation parts of the SR network and these two kinds of loss are combined for a joint optimization. To the best of our knowledge, this is the first work which explores to achieve satellite imagery SR and colorization cooperatively.

• In order to avoid color bias in imagery colorization, we incorporate natural images with satellite data to enrich the color diversity and we manage to realize the expectation color distribution learning based on these mixed data.

• We introduce a novel multi-scale deep encoder-decoder symmetrical network for satellite imagery SR, where a residual structure is adopted to improve the imagery reconstruction performance.

2. Methods

2.1. Overall scheme

In order to overcome the processing irrelevancy of existing image SR and colorization methods, our comprehensive consideration is adopting a multi-task optimization strategy which not only can reconstruct the HR satellite imagery but also can colorize the gray scale imagery for proper color information. In the sense of this, we manage to achieve the cooperative learning tasks for satellite imagery by constructing and training a multi-task deep neural network based on satellite imagery data set.

There are several components in our multi-task satellite imagery SR and colorization deep model, including multi-scale SR reconstruction, color distribution prediction based grayscale colorization, features interaction between SR and colorization parts, and multiple tasks synchronous optimization.

Benefiting from the powerful non-linear mapping, SRCNN (Dong et al., 2014, 2016) improves the performance dramatically compared with the traditional SR methods. Since training SRCNN model usually takes a very long time before convergence, Liang et al. (2016) introduce Sobel edge detection so as to capture gradient information to accelerate the training convergence. In fact, the method does reduce the training time but the reconstruction improvement is rather limited. In addition to image gradient priors, in view of the network depth with residual structure (He et al., 2016) is of crucial importance to a remarkable performance improvement, Kim et al. (2016) take twenty convolution layers with residual connection to construct deep network for image SR reconstruction.

The negligence of the above deep SR approaches is that the multiple scales image context in SR reconstruction is not fully utilized at all. Considering the fact that image multi-scale contextual information is essential for the image details reconstruction, in this work, we propose to take a multi-scale symmetrical CNN for image SR. In addition, we also introduce residual structure from the LR input to the end of the network so as to improve the reconstruction accuracy.

For satellite imagery colorization based on grayscale component input, we employ a structure similar to Zhang's network (Zhang et al., 2016) to produce the corresponding color distribution under the fused (satellite imagery and natural images) data set. Since the color diversity of satellite imagery is very

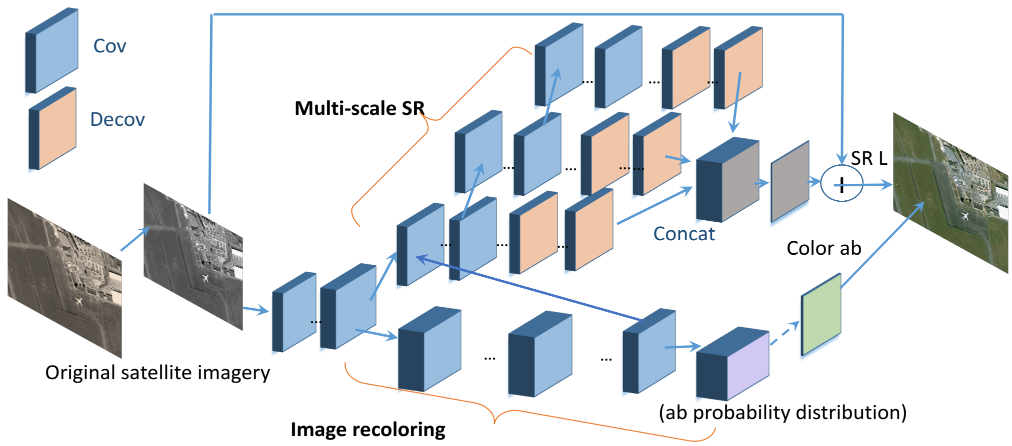


Figure 1: The overall multi-task satellite imagery deep SR and colorization model (in the figure, 'ab' refers to the color components of Lab color space).

different from natural image, we recalculate the color statistics prior, instead of the one available for natural images, and on top of it, we adjust the color class re-balancing coefficient based on fused data.

In addition, for SR and colorization multiple tasks cooperation, we choose a feedback strategy. Specifically, the final convolution features of colorization network are back propagated to the SR network and blended with the input LR imagery representation together for the HR reconstruction cooperatively. Our overall multi-task deep SR and colorization model is shown in Fig. 1.

2.2. Multi-scale learning for satellite imagery SR

With the capability of the hierarchical feature learning, multi-scale deep convolution networks appear in literature, including edge detection (Xie and Tu, 2015), skeleton extraction (Shen et al., 2016) and image dehazing (Ren et al., 2016). In a recent work (Szegedy et al., 2016), convolution filters with variable sizes are carefully designed and applied in multiple residual connections, which will lead to a very wide inception networks with better learning performance. In general, the common characteristics of these multi-scale works are taking different length convolution branches or different sizes filters to achieve different sizes receptive fields so as to extract the image features at different scales.

In addition, without fully connected layers, the fully convolutional networks (FCNs) containing only convolution and deconvolution layers have been successfully applied to semantic segmentation (Hong et al., 2015) and object detection (Yang et al., 2016). Here, a convolution layer can be interpreted as an encoder which serves for features extraction and representation while a deconvolution one, named by the decoder, acts as reconstruction.

For satellite imagery SR, we adopt a multi-scale deep symmetrical encoder-decoder structure. Obviously, the imagery $f(x)$ will be encoded with multiple scales features by different lengths convolution layers (short for coarse scale and long for fine scale). Through symmetrical decoding, the different lengths deconvolution layers will reconstruct the original imagery based on the multi-scale feature representations in a variety of scales. Actually, for an imagery $f(x)$ in L^2 space R, the principle of multi-scale encoding and decoding can be formalized by wavelet multi-resolution analysis (MRA) (Mallat, 1999) as:

$$f(x) = \sum_{k \in Z}^{N} a_k^{j_0} \phi_k^{j_0}(x) + \sum_{j=j_0}^{J} \sum_{k} b_k^j \psi_k^j(x), \tag{1}$$

where j is the scale varying from j_0 to J, k is the index of basis function, and $\{a_k^{j_0}\}$, $\{b_k^j\}$ are coefficients attached to the approximation (scale) function $\phi(x)$ and the detail (wavelet) function $\psi(x)$, respectively. In short, the image $f(x)$ can be viewed as consisting of two components (see Eq. (1)): the low-frequency approximation and the high-frequency detail. When varying the scale j from zero to certain scale, $f(x)$ can be represented as the weighted summation of a series of components at different scales, which contains a low-frequency approximation and several or numerable high-frequency details. From deep learning point of view, Eq. (1) may be treated as a combination of deconvolution (reconstruction) operations at multiple scales. Assuming at each scale, $\tilde{f}_j$, represents a reconstruction of $f(x)$. Thus, according to Eq. (1), if we take a summation function s adding up all encoder-decoder streams, the multi-scale encoder-decoder reconstruction $\tilde{f}(x)$ can be easily represented as:

$$\tilde{f}(x) = s(\tilde{f}_1, \tilde{f}_2, \cdots, \tilde{f}_j, \cdots) \tag{2}$$

Then the optimization target of our multi-scale encoder-decoder learning can be regarded as:

$$\tilde{f} = \arg\min_{\Theta}(\left\| f - \sum_j (F_j^a(y, \Theta_j^a) + F_j^b(y, \Theta_j^b)) \right\|_2^2), \tag{3}$$

where f and y represent the HR image and the corresponding LR image, and $F(\cdot)$ denotes the network reconstruction function. Θ is the learned parameter of the network and symbols j, a, b indicate a specific scale, a low-frequency approximation component and a high-frequency component, separately. By taking into account the components of different scales simultaneously, multi-scale learning will partially overcome the deficiency of only considering the energy amplitude recovering (concentrated in low-frequency components) while ignoring the structural details (in high-frequency components).

Given a set of LR and HR image pairs $\{f_i, y_i\}_{i=1}^N$, if directly treating the input LR image y_i as the low-frequency approximation component of HR image f_i and omitting the high-frequency indicator b, the loss function of the proposed multi-scale encoder and decoder learning can be finally denoted as:

$$Loss(\Theta) = \frac{1}{N}\sum_{i=1}^{N}\left\| f_i - \sum_j (y_i + F_j(y_i, \Theta_j))\right\|^2 \quad (4)$$

2.3. Color distribution prediction based imagery colorization

Although the original meaning of colorization refers to adding color to the gray-scale image, colorization for satellite imageries is more of recoloring, which indicates enhancing or changing the original color of the input satellite images desiired by specific applications, such as camouflage. The reason behind is that satellite imageries in most cases are already 3-bands color data. In practice, recoloring can also be performed in pure colorization way - extracting the intensity channel and colorizing it. In general, There are two different strategies for gray-scale imagery colorization: direct color prediction based on Euclidean color regression loss (Cheng et al., 2015; Iizuka et al., 2016) and multi-modal color distribution prediction based Softmax color classification loss (Zhang et al., 2016; Larsson et al., 2016).

Let x denote a gray-scale channel imagery to be colored, assuming in CIE *Lab* color space its associated two channels (i.e., 'ab' color components; all the following 'ab' items keep the same meaning) color is $y \in \mathbb{R}^{h\times w\times 2}$ (where h, w are image dimensions), the objective of color prediction model is to learn a mapping $\tilde{y} = f(x)$ such that the Euclidean loss $L_2(.,.)$ between predicted and ground truth colors is minimized after training:

$$L_2(\tilde{y}, y) = \frac{1}{2h \times w}\sum_{h,w}\|y_{h,w} - \tilde{y}_{h,w}\|_2^2 \quad (5)$$

Obviously, the Euclidean regression loss will lead the optimal solution $\tilde{y}$ to be the mean of all pixels' color of the ground truth image, which favors unsaturated color prediction results. Moreover, the solution does not consider the problem of color plausibility will in fact give inveracious and implausible color results. Thus, Euclidean loss based color prediction way does not handle the ambiguity and multi-modal color distribution well.

In this work, for satellite imagery colorization, we can use a deep neural network to learn a mapping $m(x)$ to a color distribution $\tilde{z}$ over possible ab color bins ($\tilde{z} \in [0,1]^{h \times w \times q}$, q is the number of color bins) for a given input x. Then we compare the predicted color distribution with the encoded ground truth one and calculate the Softmax cross entropy loss for optimization. We also take color class rebalancing technique to enhance the impact of rare color in the distribution. Finally, we take the $annealed - mean$ technique (Kirkpatrick et al., 1983) to estimate the color of every pixel based on its corresponding color distribution. Our imagery colorization network is similar to the approach of Zhang et al. (2016) but with two main differences: different means of acquiring the color probability density of satellite imagery data and adopting features interacting feeding back structure for multi-task cooperation. More specially, we calculate the color probability distribution under AID satellite data set (Xia et al., 2017) and fuse it with the prior color probability of ImageNet data set (Russakovsky et al., 2015); we expand the actual number of supported ab color bins under satellite data set to be 313 to overcome the problem that the color scope of imagery is not wide enough; finally, two deconvolution layers help to feed back the convolution features of coloriation to the SR network. The original color probability den-

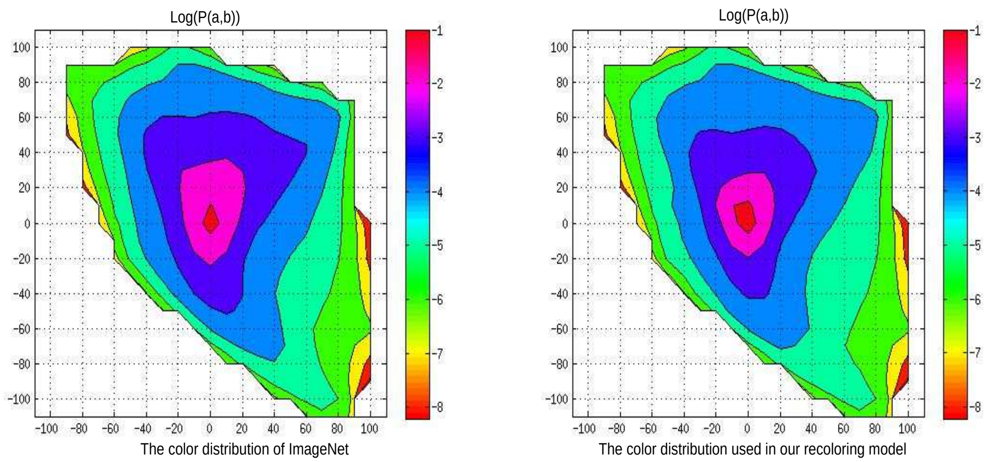


Figure 2: The contrast between ImageNet's color distribution (left) and the color distribution used in our colorization/recoloring model (right).

sity distribution of ImageNet data and the corresponding color distribution used in our colorization/recoloring model are illustrated in Fig. 2 (shown in log scale). From the figure, we can see the high probability region of our color probability distribution slightly shrinks, compared to the corresponding color distribution one of ImageNet data set, which perhaps is due to the color scope of satellite imagery is much narrower than that of natural images.

With the fused color probability and the expanded ab color bins, each ground truth color y can be easily encoded to a color vector presentation $z(\in [0,1]^{h\times w\times q})$ with its nearest neighbor color bins. For whole imagery color prediction, we define the cross-entropy loss of such color encoding prediction $L_{ce}(.,.)$ as following:

$$L_{ce}(\tilde{z}, z) = \sum_{h,w} c(z_{h,w}) \sum_{q} z_{h,w,q} log(\tilde{z}_{h,w,q}) \tag{6}$$

Here, c is a loss weighting factor used to consider the effect of the color-class rarity. At last, we estimate the final color values $\tilde{y}$ by mapping the probability distribution $\tilde{z}$ through simulated annealing way. The detailed techniques on color rebalance and color estimation can be referred to Zhang et al. (2016).

2.4. Joint multi-task learning for satellite imagery SR and colorization

Our satellite imagery multi-task deep model actually combines the proposed SR network and colorization network for concurrent execution by the convolutional features sharing (see the two front convolution layers in Fig. 1) and the features interaction (see the feedback from colorization network to SR network in Fig. 1). Based on Eq. (4) and Eq. (6), for any low resolution and gray-scale input image x_i (h, w are its height and width), if assuming its HR label image in SR model is f_i and its corresponding ground truth color distribution is $z(x_i)_{h,w}$, then the loss of multi-task joint learning for satellite imagery SR and colorization can be formalized as :

$$\begin{aligned} Loss_{total}(\Theta, \tilde{z}) = \frac{1}{N}\sum_{i=1}^{N}(\left\| f_i - \sum_{j}(x_i + F_j(x_i, \Theta_j)) \right\|^2 \\ + \eta \sum_{h,w} c(z(x_i)_{h,w}) \sum_{q} z(x_i)_{h,w,q} log(\tilde{z}(x_i)_{h,w,q})), \end{aligned} \tag{7}$$

where η is a regularization factor which controls the effects of SR reconstruction loss and color distribution loss in the whole optimization. Obviously, through such joint learning, the procedures of SR reconstruction and multi-modal color prediction will constantly regularize each other and be optimized simultaneously. When the multi-task model is trained to converge, the acquired solution (the parameters of the deep model) will be an optimal trade-off which not only can reconstruct the low resolution image well but also can map it to a color image with strong sense of reality. Through such joint learning, a satellite imagery with high resolution and visual realistic color can be obtained directly.

3. Experiments and discussions

3.1. Data sets and evaluation measures

The imageries from AID data set (30 different scene classes with about 200 to 400 samples of size 600×600 in each class) may be used for SR training while other images from RSSCN7 (Zou et al., 2015) (7 scene categories with 400 samples of size 400×400 in each class) may be utilized for testing. For satellite imagery colorization, actually the combination of 10000 random selected ImageNet images with 10000 AID satellite imageries is preferred to be applied for colorization training. Also some imageries from RSSCN7 can be regarded as the test data.

As for the quality measurements, for HR reconstruction, well-known PSNR metrics are adopted. For colorization evaluation, visual results are shown in contrast. We notice that in a down-sampled satellite imagery different regions may hold different PSNR values - smooth areas (such as plain or grassland) will get higher PSNR scores than the uneven locations (such as cross connection or zebra line). In general, we take the whole imagery's PSNR for quality evaluation which is a mean of all local regions' PSNR values. Fig. 3 gives an example.

3.2. Model training

There are two ways to train the proposed multi-task model: training it from scratch or finetuning it from the colorization model (Zhang et al.

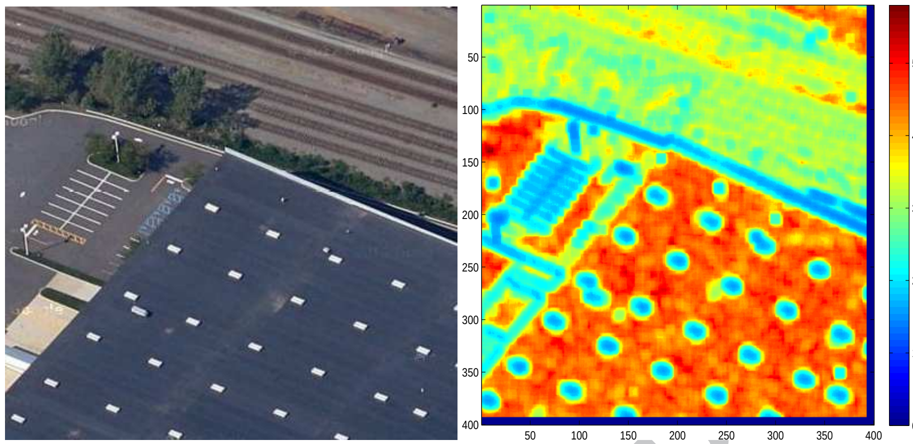


Figure 3: A satellite imagery (left) and the PSNR visualization for its down-sampled version (right); different image regions hold different PSNR values.

(2016)) for our multiple tasks. Actually, we tend to train the proposed multi-task model from scratch with a unified data strategy. In this case, about 20,000 images coming from AID satellite data and ImageNet are selected and confused for multi-task model training. Each image is augmented to 8 images by rotation which yields about 160,000 images as training set. Images are cropped into small overlapped patches with a size 96×96 and a stride of 27. For SR part, the cropped ground truth patches are used as the HR labels and the corresponding LR pairs are acquired by imposing the bi-cubic interpolation twice on the ground truth. For colorization part, the LR color images will be converted to Lab color space and keep the intensity component. The labels of this part are the encoded color distributions in *ab* color bins of the ground truth.

In the training procedure, we follow the proposal from He et al. (2016) to initialize the weights of all layers. We initially set the learning rate to 0.001 and reduce it by multiplying 0.316 every 100 thousand iterations. Momentum and weight decay parameters are set to 0.9 and 0.0001, respectively. The regularization coefficient *eta* is set to be 1 at the beginning and can be manually adjusted it to 1.5, emphasizing the impact of image color recover once the gradients of the model become relatively small. The whole deep network training is implemented using Adam solver from the Caffe package (Jia et al., 2014) with a batch size of 32. For 4× down-scaling and the confused grayscale satellite data, the model training takes about 1,300,000 iterations before convergence.

Our multi-task model can also be trained by finetuning way: training it using ImageNet then finetuning with satellite data. In this second case, twenty thousands images are randomly selected from ImageNet and used to train the multi-task model, then some AID satellite images are taken for model finetuning. All settings and parameters are the same as the first training strategy. However, we found that the second finetuning way is easily inclined to lead to the color deviation (see Fig. 6; more examples can be refereed to Fig .10). The detail configuration of our multi-task SR and recoloring model is given in Table. 1.

Table 1: Multi-task satellite imagery SR and colorization deep network configuration.

Imagery SR	(Conv3-32)×3	(Conv3-64)×3	(Conv3-128)×3	(BatchNorm)×36
	(Deconv3-32)×9	(Deconv3-64)×6	(Deconv3-128)×3	(Prelu)×36
Imagery Colorization	(Conv3-64)×2	(Conv3-128)×2	(Conv3-256)×5	(Conv3-512)×12
	(Conv1-313)×1	(Deconv3-256)×2	(Deconv4-64)×2	(BatchNorm)×7

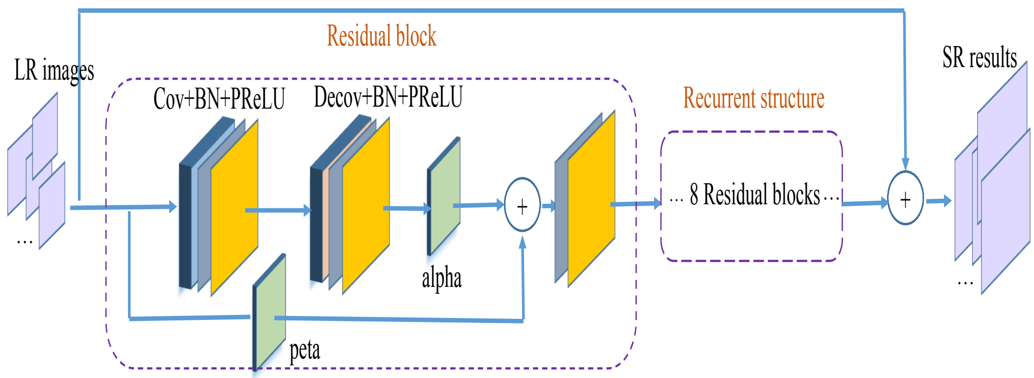


Figure 4: Another satellite image SR network - deep recurrent residual network.

3.3. Another SR network and color distribution effect

In addition to the proposed multi-scale structure for satellite image SR, other deep structures can also be utilized to achieve the same target, such as residual skip or recurrent connection. Here, we combine residual skip with recurrent connection to form a deep recurrent residual model for satellite image SR, which is similar to Patrick's one (Patrick, 2016) but with two differences : 1) we add Batch Normalization layer and replace ReLU with PReLU after each convolution layer; 2) we add a direct skip from the input to the end of the network and fix the scale parameters in every residual block (both in bypass connection part and convolution route) while not learning it from training. The architecture of our deep recurrent residual network is illustrated in Fig. 4. According to our observations (a comparison example is shown in Fig. 5), this kind of SR network sometimes can get smoother edges, but many other imagery details will be lost after reconstruction. Thus, we finally opt for the multi-scale deep structure for imagery SR of multiple tasks.

Moreover, different color distributions coming from different training data will lead to diverse colorization effect. For example, using the partial data of ImageNet for color probability distribution calculation will get very different colorization results. Specifically, for satellite image, there exists a trade-off: whether using the satellite image data or using the ImageNet data to derive the color distribution used for colorization. According to our observations, only using the satellite imagery for color distribution acquisition will inevitably cause color bias effect. Fig. 6 gives an typical illustration. Thus, in practice we get the final color distribution by data fusing strategy which is stated in Section 2.3.

3.4. Results of Imagery super-resolution and colorization

The proposed multi-task SR and colorization network accepts the LR grayscale satellite imagery as the input and reconstructs it to be a HR version and at the same time maps it to a colorized one. When given a low-resolution color imagery, it should be converted from RGB color space to Lab firstly, then the luminance component is pipelined into the multi-task network and a reconstructed HR and recoloring imagery will be output. Some simultane-

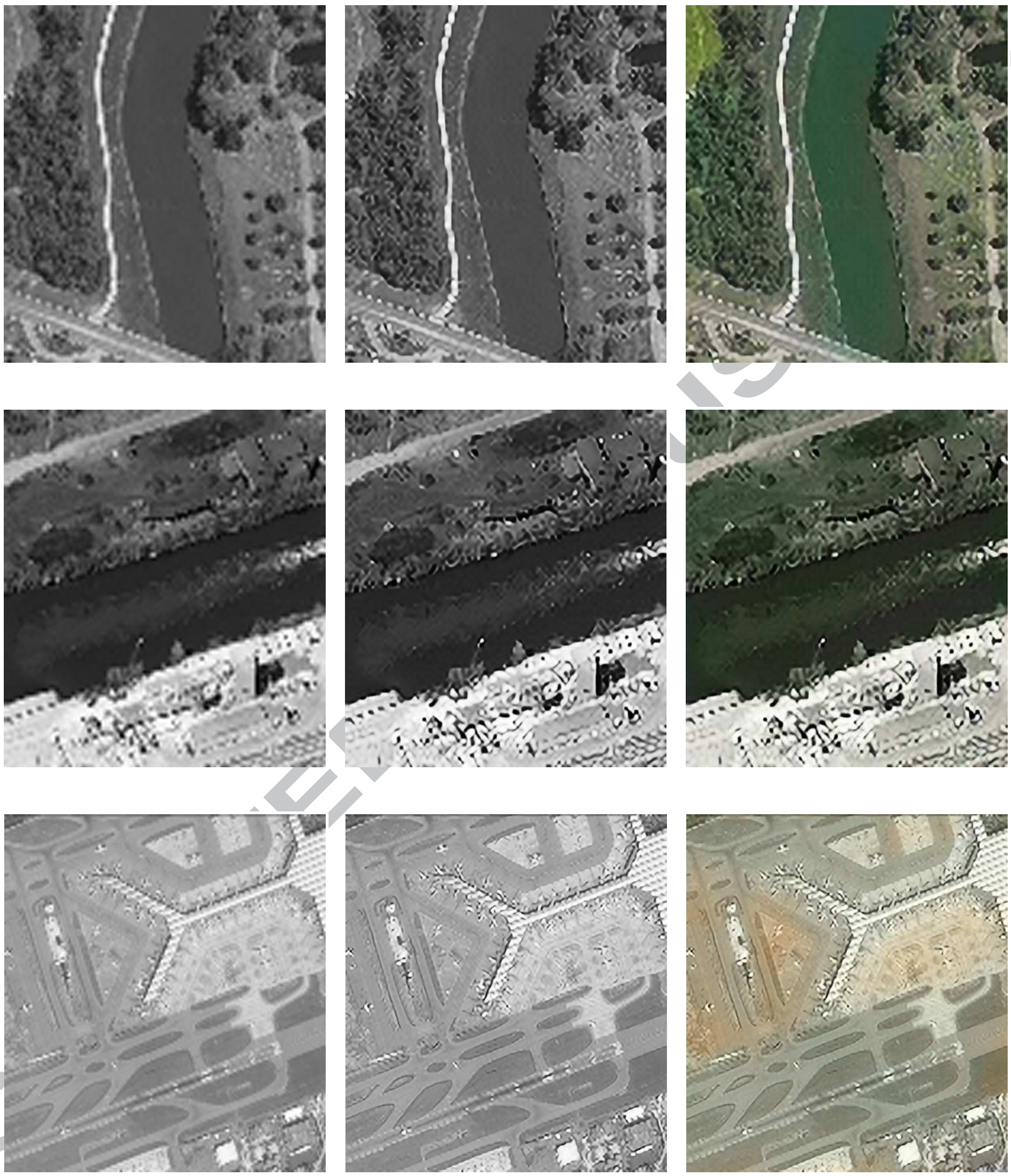
Figure 5: Imagery SR comparison: the recurrent residual network(middle) vs. the multi-scale network(right) with LR satellite image input(left).

Figure 6: Color bias example in imagery colorization when only using satellite data to acquire color probability distribution: (left)LR satellite lake imagery; (middle)the grayscale input; (right)Color biased colorization result.

ous super-resolution and colorization results of the proposed multi-task deep learning approach are shown in Fig. 7.

3.5. Comparisons and discussions

Since there is no other related work which pursues single satellite imagery simultaneous SR and colorization, we choose to compare our approach with the state-of-the art methods of two aspects: SRCNN (Dong et al., 2016) and Patrick's method (Patrick, 2016) (the model is realized by ourselves and trained with some images of SpaceNet AOI1 (SpaceNet, 2016)) for single imagery super-resolution; Zhang's method (Zhang et al., 2016), Iizuka's method (Iizuka et al., 2016) and Larsson's one (Larsson et al., 2016) for single imagery colorization. We compare and evaluate the effect of super-resolved and colorized imagery not only by subjective visual effect but also with objective PSNR(db) value. For satellite imagery super-resolution, visual results involve the subjective clarity inception of imagery details. As for imagery colorization, visual results mainly refer to the color consistency and the realism of the objects. These comparisons and experimental results are illustrated



(a) LR grayscale (b) HR reconstruction (c) SR and colorization

Figure 7: Some results of satellite imagery ('Riverlake' from RSSCN7 and 'Airport' from AID) simultaneous SR and colorization: (a)LR grayscale imagery; (b)Reconstructed HR imagery; (c)Super-resolved and colorized Imagery.

in Table 2, Fig. 8 and Fig. 9.

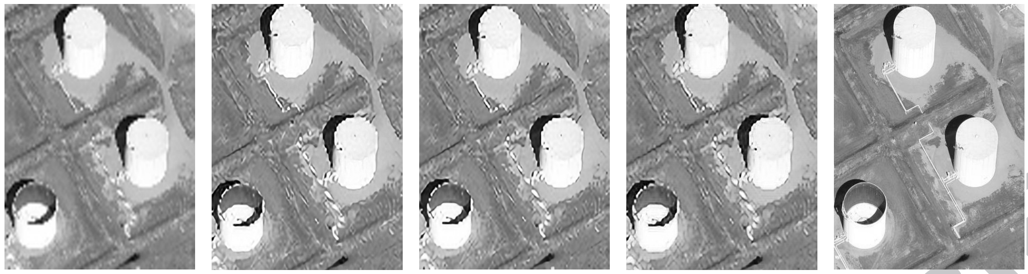
Table 2: The average PSNR (db) comparisions of imagery SR on RSSCN7.

Bicubic	SRCNN	Patric	Our multi-scale SR
27.85	28.63	28.86	29.07

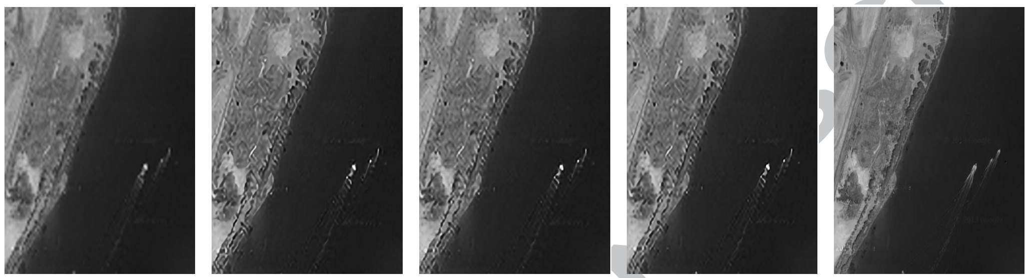
From Tabel 2 and Fig. 8, we can easily see that, our multi-scale SR approach can get superior super-resolved results even at different imagery scenario compared to SRCNN and Patrick's one not only in visual effect but also in PSNR value. Meanwhile, from the Fig. 9, it shows that the colorization effect of Zhang's method is too saturated and unnatural whereas Iizuka's and Larsson's are too light and almost equivalent to without colorization. Obviously, compared to these colorization methods, our colorization approach can get more natural and appropriate colorization effects on the whole, though which may be different from the groundtruth ones.

In addition, for fair play we also finetune Zhang's method (its visual performance ranks second in Fig. 9) with satellite data and compare the corresponding colorization results with ours. Some comparisons are shown in Fig. 10. From the figure, it is clear that even the fine-tuned Zhang's model still fails to provide acceptable colorization effect (color is monotonous or biased), whereas the proposed multi-task approach is always able to get satisfactory results.

To sum up, our multi-task imagery SR and colorization approach can not only provide subtle imagery details but also make the overall color style be coordinated and natural to visual sensation. For many applications, such as those in image synthesis, the ultimate test of colorization and super-resolution is how compelling the colors and the resolution look to a human observer. Thus, from the perspective of human perception, we also introduce



(a) Bicubic (b) SRCNN:24.20 (c) Patric's:24.33 (d) Ours:24.42 (e) Industry

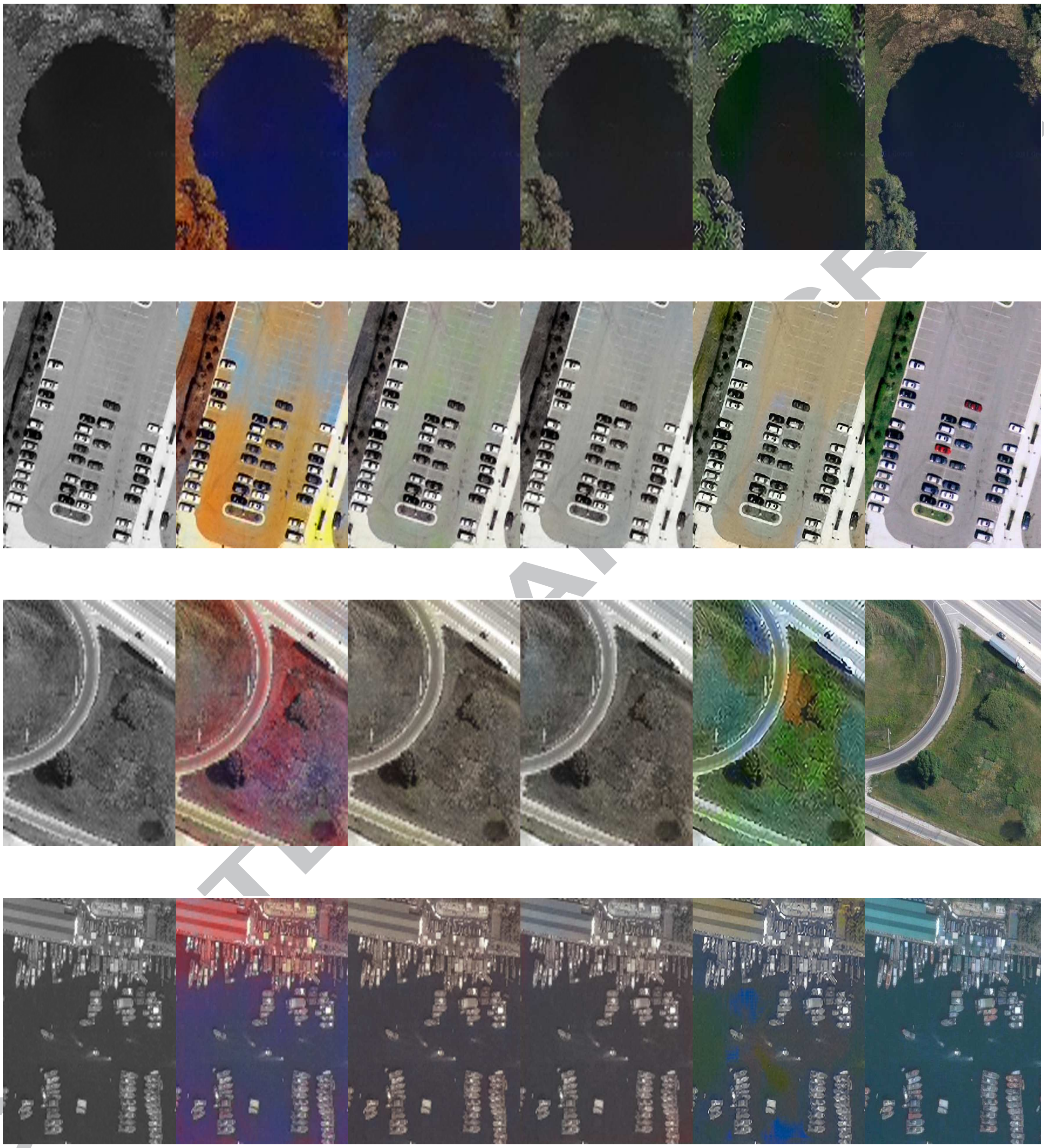


(a) Bicubic (b) SRCNN:28.35 (c) Patric's:29.06 (d) Ours:29.30 (e) Riverlake

(a) Bicubic (b) SRCNN:24.19 (c) Patric's:24.75 (d) Ours:24.95 (e) Grass

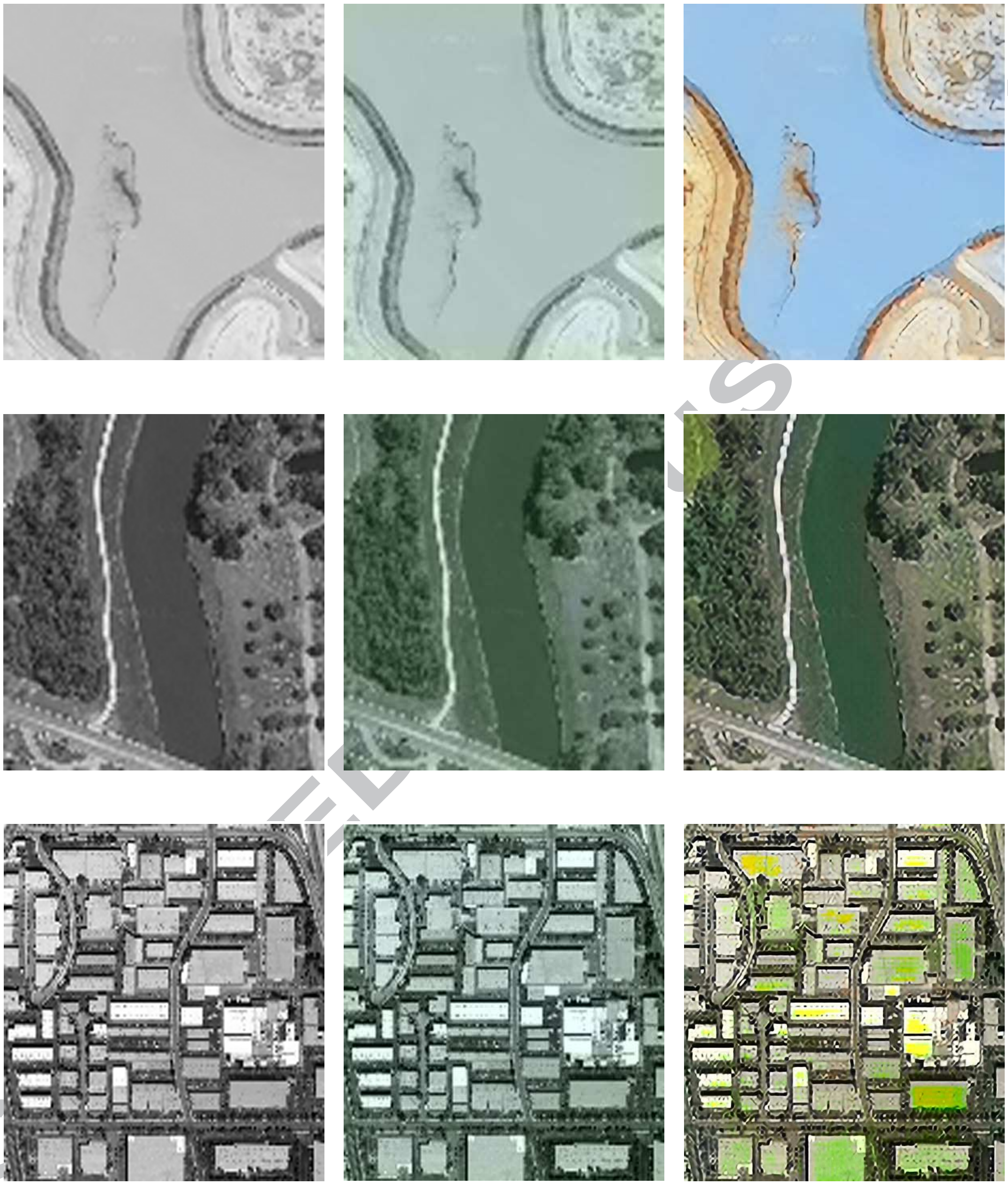
(a) Bicubic (b) SRCNN:21.17 (c) Patric's:21.55 (d) Ours:22.07 (e) Airplane

Figure 8: Visual and PSNR (db) comparisons of super-resolved (4×) images for 'Industry', 'Riverlake', 'Grass', and 'Airplane' grayscale satellite imagery from RSSCN7 by (a)Bicubic, (b)SRCNN, (c)Patric's method, and (d)Our multi-scale SR method, respectively.



(a) Input (b) Zhang's (c) Iizuka's (d) Larsson's (e) Ours (f) Groundtruth

Figure 9: Visual comparisons of satellite imagery colorization for 'Riverlake' from RSSCN7, 'Parking' from AID, 'Grass' from RSSCN7, 'Port' from AID: (a)Input LR grayscale; (b)Colorization by Zhang's; (c)Colorization by Iizuka's; (d)Colorization by Larrsson's; (e)Colorization by the proposed multi-task approach; (f)Groundtruth imagery.



(a) LR grayscale (b) Finetued Zhang's model (c) Our method

Figure 10: Colorization comparisons between ours and the finetuned Zhang's model for 'Riverlak' from RSSCN7, 'Industrial' from AID: (a) Input LR grayscale; (b) Imagery colorization with finetuned Zhang's model; (c) Imagery colorization using the proposed multi-task approach.

subjective evaluation measure to show the performance of our multi-task approach. We ran a real vs. fake two-alternative forced choice experiment on campus. Totally 30 people participated in such survey and they were shown eight pairs of satellite imageries from RSSCN7 and AID, which contain natural scene -river or lake and military sensitive images - airport or parking lots. Each pair consisted of a satellite imagery next to a re-colorized and super-resolved version, produced by either our algorithm or others. Participants were asked to discriminate the imageries and choose the one they believed contained fake colors or resolution generated by a computer program and the comparisons. Each experimental session contains eight tests (each test for only one algorithm besides ground truth: four tests for colorization and three for super-resolution) and the result of each choice is recorded and no feedback was given during all eight test pairs. To ensure that all algorithms were tested in equivalent conditions (i.e. time of day, demographics, etc.), all experiment sessions were posted simultaneously and distributed to campus in an i.i.d. fashion. These satellite imagery subjective results are shown in Table 3.

To check that participants do understand the connotation of the task, additional experimental tests were carried out - the two images of each pair were both derived from random baseline described above. Participants successfully identified these random synthesis as fake 91% of the time, indicating that they understood the task and were paying attention. The ground truth satellite imagerires are ‘d162’,‘d164’,‘d023’,‘d294’,‘a007’ from RSSCN7 and ‘ariport228’,‘airport108’,‘parking176’ from AID. We also compare the average PSNR value of such eight super-resolved imageries in Table 3.

Table 3: Satellite imagery colorization and SR subjective results.

Method		**Model**		**PSNR(db)**	**Labeled Real(%)**
SR	Colorization	Params(MB)	Runtime(ms)		
Ground Truth		–	–	–	47
Random		–	–	–	9.0
SRCNN(Dong et al., 2016)		0.3(mat file)	115	24.31	18.1
Patric's(Patrick, 2016)*		13.6	242	24.78	19.2
Our multi-scale SR		1.1	141	**25.05**	**20.8**
	Zhang's(Zhang et al., 2016)	128.9	570	–	26.6
	Iizuka's(Iizuka et al., 2016)	694.7	360	–	24.5
	Larsson'sLarsson et al. (2016)	516.0	440	–	25.2
	Our imagery clorization	129.0	570	–	**28.4**
Our multi-task SR and Colorization		131.6	390	**25.05**	**29.7**

*: We realize and train its caffe version.

From the table, it is clear that our multi-task approach fooled participants on about 30% of tests, which is significantly higher than all compared imagery colorization or SR algorithms. These results validate the effectiveness and applicability of the proposed multi-task model for satellite imagery simultaneous colorization and SR. In addition, it is interesting to catch that image color perhaps plays more important role than the resolution when we try to perceive satellite imageries visually.

4. Conclusions

In this work, for satellite imagery virtual reality applications, by presenting a novel multi-task deep learning model, we have achieved simultaneous satellite imagery SR and colorization. The proposed multi-scale SR deep structure can reconstruct LR imagery with high-frequency details and the given imagery colorization engine can efficiently recover realistic color imagery for a grayscale input. Through features interaction of different task networks and simultaneous optimization, the experimental results and comparisons based on the satellite imagery data sets show that the proposed multi-task approach outperforms the state-of-the-art methods and will get better imagery SR and colorization effect.

Future work will focus on two aspects: introducing satellite image classification (Gong et al., 2017) structure for multi-task learning; investigating the possibility of applying our multi-task deep neural model to other applications, such as saliency detection (Zhang et al., 2017b,c), image retrieval (Guo et al., 2017; Lin et al., 2017) and activity recognition (Zhang et al., 2017a; Han et al., 2012).

Acknowledgment

This work is partly supported by the Major Project of Natural Science of Anhui Provincial Department of Education (Grant No. KJ2015ZD09). It is also supported by Anhui Provincial Natural Science Foundation (Grant No. 1608085MF129).

Conflict of Interest

The authors declare that there is no conflict of interest.

Reference

Alvarez-Ramos, V., Ponomaryov, V., Reyes-Reyes, R., Gallegos-Funes, F., 2016. Satellite image super-resolution using overlapping blocks via sparse representation. In: Physics and Engineering of Microwaves, Millimeter and Submillimeter Waves (MSMW), 2016 9th International Kharkiv Symposium on. IEEE, pp. 1–4.

Brodu, N., 2016. Super-resolving multiresolution images with band-independant geometry of multispectral pixels. arXiv preprint arXiv:1609.07986.

Cheng, Z., Yang, Q., Sheng, B., 2015. Deep colorization. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 415–423.

Dong, C., Loy, C. C., He, K., Tang, X., 2014. Learning a deep convolutional network for image super-resolution. In: European Conference on Computer Vision. Springer, pp. 184–199.

Dong, C., Loy, C. C., He, K., Tang, X., 2016. Image super-resolution using deep convolutional networks. IEEE Transactions on Pattern Analysis and Machine Intelligence 38 (2), 295–307.

Gong, C., Han, J., Lu, X., 2017. Remote sensing image scene classification: Benchmark and state of the art. Proceedings of the IEEE 105 (10), 1865–1883.

Guo, Y., Ding, G., Liu, L., Han, J., Shao, L., 2017. Learning to hash with optimized anchor embedding for scalable retrieval. IEEE Transactions on Image Processing 26 (3), 1344–1354.

Han, J., Pauwels, E., de Zeeuw, P., de With, P., 2012. Employing a rgb-d sensor for real-time tracking of humans across multiple re-entries in a smart environment. IEEE Transactions on Consumer Electronics 58 (2), 255–263.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 770–778.

Hong, S., Noh, H., Han, B., 2015. Decoupled deep neural network for semi-supervised semantic segmentation. In: Advances in Neural Information Processing Systems. pp. 1495–1503.

Hung, L. Q., Giang, D. T., Quang, N. N., 2016. Superresolution method approach for vietnam remote sensing imagery. International Journal of Remote Sensing Applications 6 (0), 118–126.

Iizuka, S., Simo-Serra, E., Ishikawa, H., 2016. Let there be color!: joint end-to-end learning of global and local image priors for automatic image colorization with simultaneous classification. ACM Transactions on Graphics (TOG) 35 (4), 110.

Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T., 2014. Caffe: Convolutional architecture for fast feature embedding. In: Proceedings of the 22nd ACM international conference on Multimedia. ACM, pp. 675–678.

Kim, J., Kwon Lee, J., Mu Lee, K., 2016. Accurate image super-resolution using very deep convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1646–1654.

Kirkpatrick, S., Gelatt, C. D., Vecchi, M. P., et al., 1983. Optimization by simulated annealing. science 220 (4598), 671–680.

Larsson, G., Maire, M., Shakhnarovich, G., 2016. Learning representations for automatic colorization. In: European Conference on Computer Vision. Springer, pp. 577–593.

Liang, Y., Wang, J., Zhou, S., Gong, Y., Zheng, N., 2016. Incorporating image priors with deep convolutional neural networks for image super-resolution. Neurocomputing 194, 340–347.

Liebel, L., Körner, M., 2016. Single-image super resolution for multispectral remote sensing data using convolutional neural networks. In: XXIII ISPRS Congress proceedings. pp. 883–890.

Lin, Z., Ding, G., Han, J., Wang, J., 2017. Cross-view retrieval via probability-based semantics-preserving hashing. IEEE Transactions on Cybernetics 47 (12), 4342–4355.

Mallat, S., 1999. A wavelet tour of signal processing. Academic press.

Patrick, H., 2016. Super-resolution on satellite imagery using deep learning part 1. https://medium.com/the-downlinq/super-resolution-on-satellite-imagery-using-deep-learning-part-1, accessed April 29, 2017.

Pickup, L. C., 2007. Machine learning in multi-frame image super-resolution. Oxford University.

Ren, W., Liu, S., Zhang, H., Pan, J., Cao, X., Yang, M.-H., 2016. Single image dehazing via multi-scale convolutional neural networks. In: European Conference on Computer Vision. pp. 154–169.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al., 2015. Imagenet large scale visual recognition challenge. International Journal of Computer Vision 115 (3), 211–252.

Shen, W., Zhao, K., Jiang, Y., Wang, Y., Zhang, Z., Bai, X., 2016. Object skeleton extraction in natural images by fusing scale-associated deep side outputs. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 222–230.

SpaceNet, 2016. https://amazonaws-china.com/cn/public-datasets/spacenet/, accessed April 30, 2017.

Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A., 2016. Inception-v4, inception-resnet and the impact of residual connections on learning. arXiv preprint arXiv:1602.07261.

Xia, G.-S., Hu, J., Hu, F., Shi, B., Bai, X., Zhong, Y., Zhang, L., Lu, X., 2017. Aid: A benchmark data set for performance evaluation of aerial scene classification. IEEE Transactions on Geoscience and Remote Sensing.

Xie, S., Tu, Z., 2015. Holistically-nested edge detection. In: European Conference on Computer Vision. pp. 1395–1403.

Yang, J., Huang, T., 2010. Image super-resolution: Historical overview and future challenges. In: Milanfar, P. (Ed.), Super-resolution imaging. CRC Press, pp. 1–33.

Yang, J., Price, B., Cohen, S., Lee, H., Yang, M.-H., 2016. Object contour detection with a fully convolutional encoder-decoder network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 193–202.

Zhang, B., Yang, Y., Chen, C., Yang, L., Han, J., Shao, L., 2017a. Action recognition using 3d histograms of texture and a multi-class boosting. IEEE Transactions on Image Processing 26 (10), 4648–4660.

Zhang, D., Han, J., Jiang, L., Ye, S., Chang, X., 2017b. Revealing event saliency in unconstrained video collection. IEEE Transactions on Image Processing 26 (4), 1746–1758.

Zhang, D., Meng, D., Han, J., 2017c. Co-saliency detection via a self-paced multiple-instance learning framework. IEEE Trans. on Pattern Analysis and Machine Intelligence 39 (5), 865–878.

Zhang, H., Yang, Z., Zhang, L., Shen, H., 2014. Super-resolution reconstruction for multi-angle remote sensing images considering resolution differences. Remote Sensing 6 (1), 637–657.

Zhang, R., Isola, P., Efros, A. A., 2016. Colorful image colorization. In: European Conference on Computer Vision. Springer, pp. 649–666.

Zhu, H., Song, W., Tan, H., Wang, J., Jia, D., 2016. Super resolution reconstruction based on adaptive detail enhancement for zy-3 satellite images. In: ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences. Vol. III-7. pp. 213–217.

Zou, Q., Ni, L., Zhang, T., Wang, Q., 2015. Deep learning based feature selection for remote sensing scene classification. IEEE Geoscience and Remote Sensing Letters 12 (11), 2321–2325.

ACCEPTED MANUSCRIPT

Highlights

- We propose a multi-task deep neural model to achieve satellite imagery SR and colorization simultaneously. To the best of our knowledge, this is the first work which explores to achieve satellite imagery SR and colorization cooperatively.
- We incorporate natural images with satellite data to enrich the color diversity in imagery colorization and we manage to realize the expectation color distribution learning to avoid color bias in colorization.
- We introduce a novel multi-scale deep encoder-decoder symmetrical network for satellite imagery SR, where a residual structure is adopted to improve the imagery reconstruction performance.