

# Transfer Learning Based on A+ for Image Super-Resolution

Mei Su, Sheng-hua Zhong<sup>(✉)</sup>, Jian-min Jiang

Research Institute for Future Media Computing, College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, 518000 P.R. China

sumeimei@email.szu.edu.cn  
{csshzhong, jianmin.jiang}@szu.edu.cn

**Abstract.** Example learning-based super-resolution (SR) methods are effective to generate a high-resolution (HR) image from a single low-resolution (LR) input. And these SR methods have shown a great potential for many practical applications. Unfortunately, most of popular example learning-based approaches extract features from limited training images. These training images are insufficient for super resolution task. Our work is to transfer some supplemental information from other domains. Therefore, in this paper, a new algorithm Transfer Learning based on A+ (TLA) is proposed for image super-resolution task. First, we transfer supplemental information from other datasets to construct a new dictionary. Then, in sample selection, more training samples are supplemented to the basic training samples. In experiments, we seek to explore what types of images can provide more appropriate information for super-resolution task. Experimental results indicate that our approach is superior to A+ when transferring images containing similar content with original data.

**Keywords:** Image super resolution · Transfer learning · Example learning-based

## 1 Introduction

Image super resolution (SR) is a process to generate a high resolution (HR) image from input low resolution (LR) image and minimize visual artifacts. It has been widely used in many fields including computer vision, video surveillance and remote-sensing images [1]. As a low-cost post-processing technique, the SR technique can break through the limitation of both imaging equipment and the environment to produce an HR image that traditional digital cameras cannot capture from a real scene [2]. SR reconstruction has attracted extensive attention for these reasons. For the last few decades, many image super-resolution methods have been developed to display high quality images and provided a remarkable progress [3]. There is a general agreement that the existing SR methods can be categorized into three groups: interpolation-based methods, reconstruction-based methods, and example learning-based methods [4].

Example learning-based methods utilize a set of external training images to predict the missing high-frequency details in LR input. These example learning-based approaches have gained significant improvement in super-resolution task. And most example learning-based methods construct a dictionary, which contains a large number of low-resolution and high-resolution patch pairs [5]. In the learning procedure, an optimization function is proposed just like other applications [6]. An efficient method named A+ [7] has better performance and relatively lower running time than other example learning-based methods. In training stage, A+ method extracts features from the training set to construct sparse dictionary. Then, A+ method employs samples taken from the training pool to generate the neighborhood used for regression. The training set used in A+ was proposed by Yang *et al.* [8]. This dataset contains only 91 nature images and has been employed in many papers [7], [9]. As we know, the size of the training dataset is less than that used in many other image processing tasks. Thus, we have such a question: are this commonly used training dataset containing sufficient information for super-resolution task?

Transfer learning has the property to transfer information from different domains to a specific domain. Thus, in this paper, we propose a novel algorithm named Transfer Learning based on A+ (TLA). Our goal is to utilize some supplemental information from other domains to achieve better performance for super-resolution task. And we seek to explore what types of images can provide more appropriate information for us to conduct super-resolution task.

The rest of this paper is organized as follows. In section 2, we briefly review the previous works on transfer learning. Section 3 presents the proposed Transfer Learning based on A+ method in detail. We conduct experiments and demonstrate the performance of our method in section 4. Finally, in section 5, we conclude this paper.

## 2 Related Work

As one of the most important research directions in machine learning, transfer learning has been well studied in various fields in recent years. Transfer learning is a new approach to improve the performance of unknown target domain by utilizing previously acquired knowledge learned from a source domain which may be different with the target domain.

Approaches to transfer learning can be divided into four categories based on “what to transfer” [10]: instance-based approaches, feature-based approaches, parameter-based approaches and relational approaches. Instance-based transfer techniques reuse data from the source tasks to augment the target task’s training data to learn in the target domain. Instance-based transfer learning approaches deal with the incomplete knowledge on examples and borrow data from other similar source tasks to improve the performance of target tasks. These approaches have been widely employed in many applications, such as image retrieval, web document classification [11].

As far as we known, there are two approaches used transfer learning for super-resolution task. The difference between these works and ours is the form of the transferred information. Dai *et al.* [12] transferred the manifold structure from the space of HR

patches into LR patches to imitate the metrics computed in HR patches. Dong *et al.* [13] transferred the shallow convolutional neural network learned in a relatively easier task to initialize a deeper or harder network. Different from other methods, our method tries to transfer useful content from other images to construct a better model for super-resolution task.

### 3 Proposed Method

We propose an algorithm called Transfer Learning based on A+. Our TLA method is closest related to A+, while at the same time improving over its performance. We will explain the general formulation for TLA in this section.

#### 3.1 Feature Extraction and Representation

The number of the original training dataset used in A+ is denoted as  $T_o$ . At this stage, we transfer images from other domain to provide more information for training. We randomly selected  $T_i$  images from other datasets as the transferred images. The new training dataset is formed by the original training images and the transferred images. These images are taken as HR images.

We use the same feature extraction method as Zeyde *et al.* [9] and A+ [7]. First, we obtain LR images from the HR images by down-sampling. For these given HR images, the corresponding LR images are downscaled from HR images for a given upscaling factor  $u$ . Then, the LR images is bicubically interpolated to an interpolated HR images by the same factor. Since the high-frequency details are missing, the interpolated HR images are also called LR images.

In order to extract local features that correspond to their high-frequency content, all LR images (interpolated HR images) are filtered using  $R$  high-pass filters. Thus, each interpolated HR image  $X_i$  leads to  $R$  filtered images from  $f_r * X_i$ , for  $r = 1, 2, \dots, R$  (where  $*$  stand for a convolution) [9].

Then, we extract small patches from these images. In this paper, the size of LR patches is  $3 \times 3$  pixels, so we work with patches of  $(3 \times 3) \times u = 3u \times 3u$  pixels. We extract low-resolution features from the filtered images  $f_r * X_i$ . The feature size is  $3u \times 3u$  and every feature is vectorized to the length of  $9u^2$ . Then, every corresponding  $R$  such low-resolution features are merged into one vector of length  $9u^2R$ .

For all the HR images, we remove their low-frequency information. Then, the corresponding HR features  $y$  are extracted from the same locations in LR images. Thus, all LR features  $x$  and their corresponding HR features  $y$  are obtained  $\mathcal{D} = \{(x_1, y_1), (x_2, y_2), \dots, (x_F, y_F)\} \in R^M \times R^N$ .

#### 3.2 Learning Phase

Since the representation of LR features is quite high-dimensional, we apply the Principal Component Analysis (PCA) dimensionality reduction to project features to a low-

dimensional subspace while preserving 99.9% of the energy. The same PCA projection matrix produced when reducing the dimension of training features is also applied during testing. Then, we use the dictionary training method proposed by Zeyde *et al.* [9] to obtain a low-resolution sparse dictionary  $\mathbf{D}_L = \{\mathbf{d}_{L,k}\}_{k=1}^K$  using the following formulation:

$$\min_{\boldsymbol{\delta}} \|\mathbf{D}_L \boldsymbol{\delta} - \mathbf{x}\|_2^2 + \lambda \|\boldsymbol{\delta}\|_1 \quad (1)$$

where  $\boldsymbol{\delta}$  is the sparse representation,  $\mathbf{x}$  are low resolution features and  $\lambda$  is a weighting factor. Then, we reconstruct the corresponding high resolution dictionary  $\mathbf{D}_H$  by enforcing the coefficients in the HR and LR patch decompositions to be the same.

A+ employs samples taken from the training pool to generate the neighborhood used for regression. This training pool is formulated by LR features, which are extracted from the original training dataset. The number of these training samples is  $S_o$ . In order to make training samples more sufficient, we increase a fixed number of  $S_i$  samples transferred from other domain. Our new training samples are defined as  $S$  and it consists of original  $S_o$  samples and the new  $S_i$  samples. The optimization problem becomes:

$$\min_{\boldsymbol{\gamma}} \|\mathbf{x}_i - \mathbf{N}_{L,k} \boldsymbol{\gamma}\|_2^2 + \lambda \|\boldsymbol{\gamma}\|_2 \quad (2)$$

where  $\mathbf{N}_{L,k}$  containing  $m$  training samples from the  $S$  training samples. And these  $m$  training samples lie closest to the dictionary atom to which the input feature  $\mathbf{x}_i$  is matched. The closed-form solution is given in equation (3).

$$\boldsymbol{\gamma} = (\mathbf{N}_{L,k}^T \mathbf{N}_{L,k} + \lambda \mathbf{I})^{-1} \mathbf{N}_{L,k}^T \mathbf{x}_i \quad (3)$$

We define the HR neighborhood corresponding to  $\mathbf{N}_{L,k}$  as  $\mathbf{N}_{H,k}$ . Therefore, the super-resolution problem can be solved by calculating for each LR feature  $\mathbf{x}_i$  its nearest neighbor atom in the dictionary, and then reconstructed HR patch  $\mathbf{y}_{H,i}$  as show in equation (4).

$$\mathbf{y}_{H,i} = \mathbf{N}_{H,k} (\mathbf{N}_{L,k}^T \mathbf{N}_{L,k} + \lambda \mathbf{I})^{-1} \mathbf{N}_{L,k}^T \mathbf{x}_i, k = 1, 2, \dots, K \quad (4)$$

The projection matrix  $\mathbf{P}_k$  can be defined as show in equation (5). Thus, every dictionary atom  $\mathbf{d}_k$  has its corresponding projection matrix  $\mathbf{P}_k$  and it can be computed offline.

$$\mathbf{P}_k = \mathbf{N}_{H,k} (\mathbf{N}_{L,k}^T \mathbf{N}_{L,k} + \lambda \mathbf{I})^{-1} \mathbf{N}_{L,k}^T \quad (5)$$

### 3.3 Reconstruction Phase

In the reconstruction phase, for these given test images, we first extract LR features from them by the same operations as used in training. For each LR feature  $\mathbf{x}_{T,j}$ , we

calculate its nearest dictionary atom  $d_k$  from the LR dictionary  $D_L$ . And then, the projection matrix  $P_k$  of  $d_k$  is obtained. Therefore, the reconstructed HR patch  $y_{T,j}$  for test images is computed using the following equation:

$$y_{T,j} = P_k x_{T,j} \quad (6)$$

These reconstructed HR patches  $y_{T,j}$  are combined by the same way they decomposed and form the final reconstructed HR image.

## 4 Experiments

This section includes three experiments, which are conducted to investigate the performance of the proposed TLA and compare it with A+ method. The first experiment is to explore the effect when transferring similar information on TLA method. In the second experiment, we try to explore whether the improvement also exist when supplemental information comes from texture images. In the third experiment, we seek to analyze the performance of transfer learning based on different types of nature scene images.

### 4.1 Experiment Setting

We use the standard training set proposed by Yang *et al.* [8] which contains 91 images for training (original training images number  $T_o$  is 91). In each experiment, we newly increase one training dataset as transfer learning materials, including: face images, texture images and nature scene images, respectively. The standard nature scene set Urban and Natural Scene dataset [14] used in the third experiment is composed of 2688 images with eight categories and each one with an average of 336 images. Thus, we also use 336 images for face and texture images to conduct Experiment 1 and Experiment 2. For test, we employ two datasets Set14 and BD100. Set14 was proposed by Zeyde *et al.* [9] including 14 commonly used images for super-resolution evaluation, while BD100 contains the 100 testing images from the Berkeley Segmentation Data Set 300 (BSDS300) [15]. This Dataset is widely used for various computer vision tasks including super-resolution.

In our experiment, we compare our method with A+ [7] under the same condition, since it is an efficient dictionary-based SR method and the closest related method with our TLA. And our method is based on A+. The implementations are from the publicly available codes provided by the authors. For performance evaluation, methods are evaluated in terms of Peak Signal-to-Noise Ratio (PSNR) and structural similarity (SSIM) [16]. PSNR usually correlates well with the visual quality and SSIM is used for measuring the similarity between two images.

We conduct most of our experiments related to the internal parameters of A+ method in order to compare with A+ as fairly as possible. For the parameters used both in our method and A+, we follow the setting of them in A+. For instance, we set the upscaling factor  $u$  equals to 3 since it is the most widely used value in super resolution task, the original training samples number  $S_o$  is 5 million, the dictionary size  $K$  is 1024,

the neighborhood size  $m$  is 2048 and  $\lambda$  is 0.1 [17]. These are the best or common parameters as reported in its respective original work. The high-pass filter number  $R$  is set to 4 including the first and second order gradients filters. We also set the transferred images number  $T$ , equals to 100 and new samples number  $S$ , is 1 million because a larger number of  $S$ , does not improve significantly by experimental verification. All the experiments in this paper are repeated five times.

#### 4.2 Experiment 1: What is The Effect of Adding Similar Information

In this subsection, we seek to explore what is the effect on the performance when we try to transfer some information similar with the original information in training and test datasets. Faces are important in human perception and the images of face are widely used in image processing since they convey a wealth of information. We also find that face images are often selected as materials in super-resolution task. In order to make the samples of face images abundant, we selected 336 face images from the training data of ImageNet dataset [18] in our experiment. We decided to select images from ImageNet because it is large in scale and diversity in types. Some samples of these face images are shown in Fig.1.

For the images in the test datasets, we find that there are six human face images, and one animal face image in Set14 dataset. For dataset BD100, there are 17 human face images and eight animal face images. We regard all of these images as face images in our experiment. However, the original training dataset only has five face images, and most of them just contain some parts of human face. Therefore, we believe that sufficient face images do provide some useful information for transfer learning.



Fig. 1. 10 face images selected from ImageNet dataset.

Table 1. Average PSNR(dB) and SSIM results comparison on two test datasets and the face images in these datasets.

Benchmark		PSNR(dB)		SSIM	
		Set14	BD100	Set14	BD100
All test images	A+	29.130	28.180	0.89330	0.77970
	TLA	<b>29.138</b>	<b>28.192</b>	<b>0.89412</b>	<b>0.78090</b>
Face images	A+	29.200	28.140	0.88651	0.80174
	TLA	<b>29.203</b>	<b>28.158</b>	<b>0.88748</b>	<b>0.80301</b>

Table 1 shows two groups of results for PSNR and SSIM on two test datasets, Set14 and BD100. The first group is conducted on all images in test datasets. And the second one is the results of face images in test datasets. From these two groups of results, we find that our proposed method TLA shows performance improvement than A+ method, especially on BD100 dataset. Meanwhile, the improvement of face images is better than

other images on BD100 dataset. It indicates that in super-resolution, the supplemental information could be useful. Therefore, the transfer learning based on face images is beneficial to the performance of super-resolution task.

### 4.3 Experiment 2: What is The Effect of Increasing Texture Images

The first experiment has proved that supplemental images have similar content with original data are beneficial to improve the super-resolution performance. But others may have such a question: whether this improvement comes from more related information or just more images? To rule out the second possibility, in Experiment 2, we use transfer learning based on texture images in our experiment.

In this part, a standard dataset Describable Textures Dataset (DTD) [19] is utilized in our experiment. As a texture database, DTD consist of 5640 images. And it is organized according to a list of 47 categories (e.g. banded, dotted, grid, stratified) and 120 images for each category. For a fair comparison with other categories in Experiment 1, we also select 336 texture images from DTD and conduct our experiment based on these texture images. Fig.2 shows some sample images selected from the dataset.

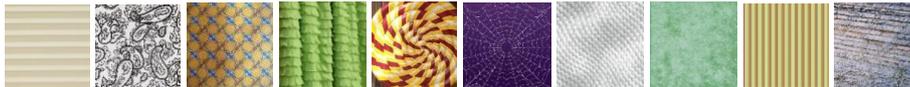


Fig. 2. 10 texture sample images selected from DTD.

Table 2. Average PSNR(dB) and SSIM results comparison on Set14 and BD100.

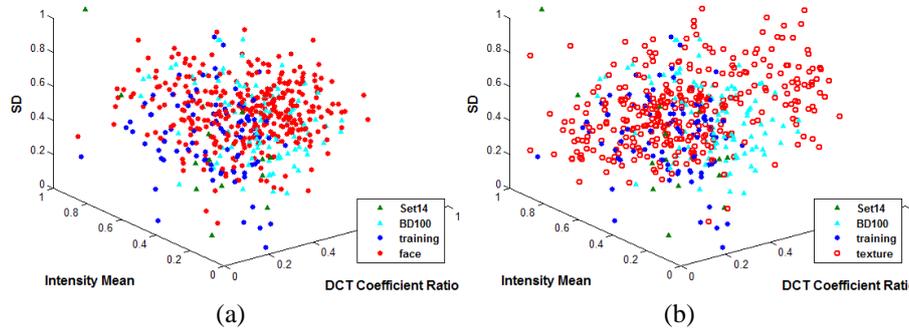
Benchmark		PSNR		SSIM	
		Set14	BD100	Set14	BD100
Texture images	A+	<b>29.130</b>	<b>28.180</b>	0.89330	0.77970
	TLA	29.064	28.178	<b>0.89362</b>	<b>0.78046</b>

The PSNR and SSIM results for texture images are shown Table 2. From Table 2, we find that transfer learning based on texture images cannot improve the performance on PSNR. We also find the performance has been improved on SSIM metric. The reason is that structural information in texture images could be useful to improve the performance of SSIM metric. However, although TLA based on texture images has a little improvement on SSIM metric, the increase is smaller than the case of transfer learning based on face images in Table 1. This indicates that if our incremental information is far different from the original data, the performance cannot be improved effectively. These results also suggest that the PSNR and SSIM of TLA with face images are better than that of TLA with texture images, and these two models are significantly different in a paired t-test ( $p < 0.001$ ).

In common sense, we believe face images and texture images are different. To explore why the performance of transfer learning based on face images is better than texture images, we analyze several variables such as Discrete Cosine Transform (DCT) Coefficient Ratio (DCT-CR), Intensity Mean (IM) and Standard Deviation (SD). These

variables are utilized to represent the essential nature of images in Experiment 1, Experiment 2 and the basic training and test images.

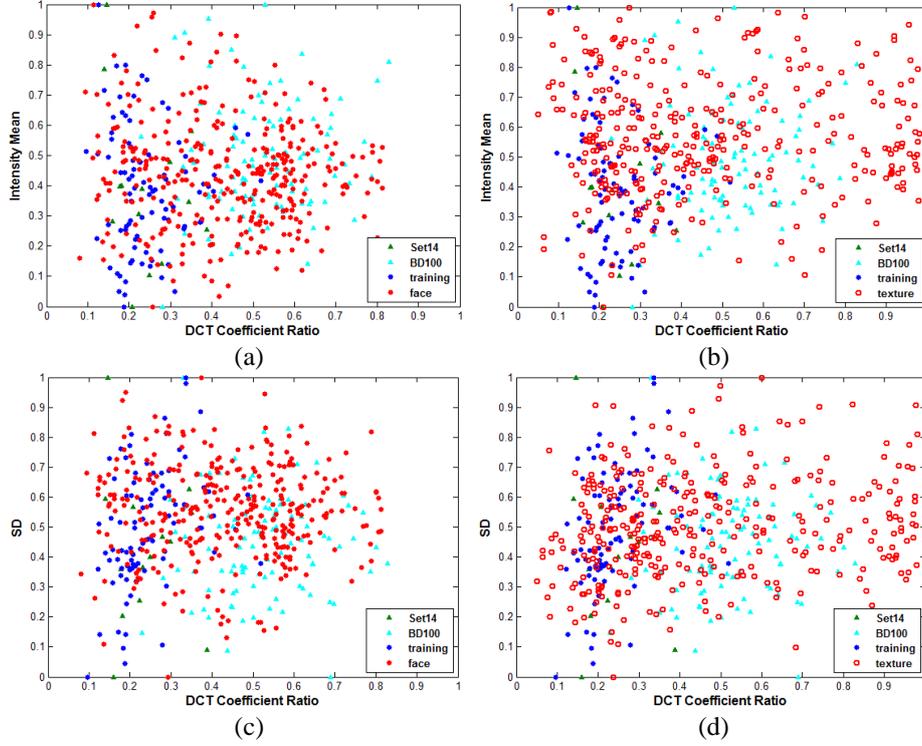
The DCT Coefficient Ratio for an image is  $c = n_1/n$  where  $n_1$  is the number of non-zero value in DCT coefficient and  $n$  is the number of all the values. DCT Coefficient Ratio is used to present the complexity of an image while Intensity Mean represents image's luminance. SD is a measure that is used to quantify the amount of variation or dispersion of each image. In our experiment, we normalize the values of three variables between 0 and 1.



**Fig. 3.** 3-D representations for DCT Coefficient Ratio, Intensity Mean and SD from Set14, BD100, training images, face images and texture images. (a) Visualization on 3-D space for variables results of original datasets and face images; (b) Visualization on 3-D space for variables results of original datasets and texture images.

Fig.3 shows the 3-D representation of the relationship between DCT Coefficient Ratio, Intensity Mean and SD for Set14, BD100, training images, face images and texture images. In Fig.3, the  $x$ ,  $y$ ,  $z$  axis represents DCT Coefficient Ratio, Intensity Mean and SD, respectively. Green triangles, light blue triangles, blue points, red points and red hollow points represent Set14, BD100, training images, face images and texture images, respectively. From three variable results in Fig.3, we find that face images may be more close to original datasets than texture images. In order to observe the detailed distributions of them clearly, we project these results in two-dimensional spaces in Fig.4.

In Fig.4 (a), (b),  $x$  axis is DCT Coefficient Ratio and  $y$  axis is Intensity Mean. In Fig.4 (c), (d), the  $x$  axis stays still and the  $y$  axis is SD. Green triangles, light blue triangles, blue points, red points and red hollow points represent the same objects with Fig.3. In Fig. 4(a) and Fig. 4(c), we can find the points come from face images could cover the regions of the points in original training images and test images. It means they have similar visual or structural properties. On the contrary, the distribution of the variables in texture images is more disperse than face images. In Fig. 4(b) and Fig. 4(d), we can find some points from original datasets are isolated with the points from texture images. These results could help us understand the differences in their performance for transfer learning.



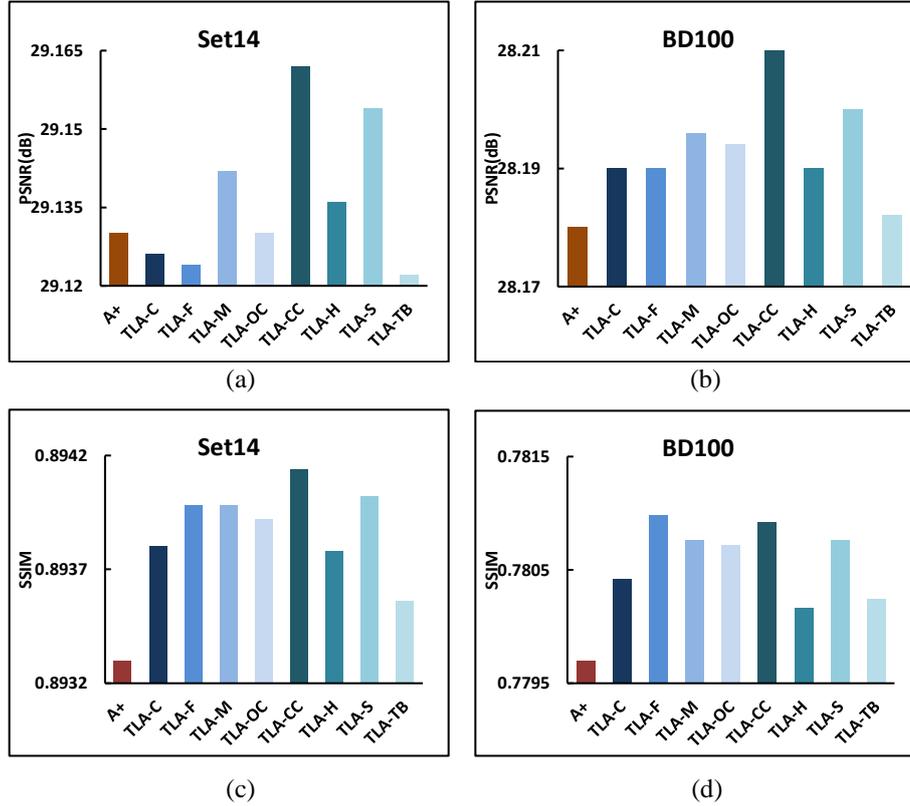
**Fig. 4.** 2-D representations for different variables from Set14, BD100, training images, face images and texture images. (a) Visualization on 2-D space for DCT-CR and IM of original datasets and face images; (b) Visualization on 2-D space for DCT-CR and IM of original datasets and texture images; (c) Visualization on 2-D space for DCT-CR and SD of original datasets and face images; (d) Visualization on 2-D space for DCT-CR and SD of original datasets and texture images.

In general, we could get the conclusion that if the supplemental information has similar characters with the original data, transfer learning can improve the performance. And we can get opposite results when our supplemental information is different with the original ones.

#### 4.4 Experiment 3: What is The Effect of Different Categories

Experiment 1 and Experiment 2 have shown the performance of transfer learning based on face images and texture images, respectively. These experiments have proved that similar information contributes to improve the performance of super-resolution. However, we only use two types of images, the common used face images and an extreme case, texture images. Therefore, in this subsection, we seek to employ more categories of nature images to analyze whether the performance changes with the categories or not.

To statistically analyze this problem, we utilize a standard dataset called Urban and Natural Scene dataset [14]. This dataset includes 2688 authentic images with eight semantically organized categories: Coast, Forest, Mountain, Open Country, Highway, City Center, Street and Tall Building. The dataset is a widely used nature dataset and has been utilized in many other papers [20]. In this experiment, we conduct our experiment on each category in Urban and Natural Scene dataset, respectively.



**Fig. 5.** Average PSNR and SSIM results comparison on test datasets for different categories. (a) PSNR on Set14 for A+ method and TLA based on eight categories. (b) PSNR on BD100 for A+ method and TLA based on eight categories. (c) SSIM on Set14 for A+ method and TLA based on eight categories. (d) SSIM on BD100 for A+ method and TLA based on eight categories.

Fig.5 shows the average PSNR and SSIM results of various categories on test datasets. Fig.5 (a), (b) are PSNR values of A+ and TLA based on the eight categories while Fig.5 (c) and (d) are the SSIM values on the same condition. In Fig.5, the captions TLA-C, TLA-F, TLA-M, TLA-OC, TLA-CC, TLA-H, TLA-S and TLA-TB are used to represent the TLA method when supplementing Coast, Forest, Mountain, Open Country, City Center, Highway, Street and Tall Building images, respectively. The results indicate that when supplemental information is from different categories, the

PSNR and SSIM results are also different. From Fig.5 (a), we can find that the images from three categories including Coast, Forest and Tall Building make the performance worse. For most of categories, the performances of TLA method improve significantly than A+ method, especially for City Center and Street images. We infer the reason is that the images from Coast, Forest and Tall Building have different characters with the original data while images from City Center and Street can provide more similar information. Therefore, transfer learning based on different categories will demonstrate different effects on the performance of super-resolution task.

## 5 Conclusion

In this paper, we present a novel transfer learning-based super-resolution algorithm called TLA. We transfer some useful information from other datasets to improve the performance of super-resolution task. Three experiments are conducted in our paper. The first experiment indicates that the transfer learning based on face images is beneficial to the performance of super-resolution task. In the second experiment, we can find that if our incremental information is far different from the original data, the performance cannot improve. The third one proves that transfer learning based on different categories has different effects on super-resolution task. Our experimental results indicate that images containing similar content with original data are helpful. As further research directions, transfer learning can combine with other super-resolution approaches to extract more useful information.

**Acknowledgements:** This work was supported by the National Natural Science Foundation of China (No. 61502311), the Natural Science Foundation of Guangdong Province (No. 2016A030310053), the Science and Technology Innovation Commission of Shenzhen under Grant (No. JCYJ20150324141711640), the Strategic Emerging Industry Development Foundation of Shenzhen (No. JCY20130326105637578), the Shenzhen University research funding (201535), Special Program for Applied Research on Super Computation of the NSFC-Guangdong Joint Fund (the second phase), and the Tencent Rhinoceros Birds Scientific Research Foundation (2015).

## References

1. Deng, C., Xu, J., Zhang, K., Tao, D., Gao, X., Li, X.: Similarity constraints-based structured output regression machine: An approach to image super-resolution. *IEEE Transactions on Neural Networks and Learning Systems*, 1-1 (2015)
2. Yu, J., Gao, X., Tao, D., Li, X., Zhang, K.: A unified learning framework for single image super-resolution. *IEEE Transactions on Neural Networks and Learning Systems* **25**(4), 780-792 (2014)
3. Zhu, Y., Li, K., Jiang, J.: Video super-resolution based on automatic key-frame selection and feature-guided variational optical flow. *Signal Processing: Image Communication* **29**(8), 875-886 (2014)

4. Zhang, K., Zhou, X., Zhang, H., Zuo, W.: Revisiting single image super-resolution under internet environment: Blur kernels and reconstruction algorithms. In: *Advances in Multimedia Information Processing-PCM 2015*, pp. 677-687. Springer (2015)
5. Li, K., Zhu, Y., Yang, J., Jiang, J.: Video super-resolution using an adaptive super-pixel guided auto-regressive model. *Pattern Recognition* 51, 59-71 (2016)
6. Li, J., Qiu, M.K., Ming, Z., Quan, G., Qin, X., Gu, Z.: Online optimization for scheduling preemptable tasks on IaaS cloud systems. *Journal of Parallel and Distributed Computing* 72(5), 666-677 (2012)
7. Timofte, R., De Smet, V., Van Gool, L.: A+: Adjusted anchored neighborhood regression for fast super-resolution. In: *Computer Vision-ACCV 2014*, pp. 111-126. Springer (2014)
8. Yang, J., Wright, J., Huang, T., Ma, Y.: Image super-resolution as sparse representation of raw image patches. In: *CVPR*, pp. 1-8. IEEE (2008)
9. Zeyde, R., Elad, M., Protter, M.: On single image scale-up using sparse-representations. In: *Curves and Surfaces*, pp. 711-730. Springer (2010)
10. Pan, S.J., Yang, Q.: A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering* 22(10), 1345-1359 (2010)
11. Zhang, D., Si, L.: Multiple instance transfer learning. In: *IEEE International Conference on Data Mining Workshops*, pp. 406-411. IEEE (2009)
12. Dai, D., Kroeger, T., Timofte, R., Van Gool, L.: Metric imitation by manifold transfer for efficient vision applications. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 3527-3536 (2015)
13. Dong, C., Deng, Y., Change Loy, C., Tang, X.: Compression artifacts reduction by a deep convolutional network. In: *Proceedings of the IEEE International Conference on Computer Vision*. pp. 576-584 (2015)
14. Oliva, A., Torralba, A.: Modeling the shape of the scene: A holistic representation of the spatial envelope. *International journal of computer vision* 42(3), 145-175 (2001)
15. Martin, D., Fowlkes, C., Tal, D., Malik, J.: A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: *IEEE International Conference on Computer Vision*, pp. 416-423. IEEE (2001)
16. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* 13(4), 600-612 (2004)
17. Zhang, Y., Zhang, Y., Zhang, J., Wang, H., Dai, Q.: Single image super-resolution via iterative collaborative representation. In: *Advances in Multimedia Information Processing-PCM 2015*, pp. 63-73. Springer (2015)
18. Deng, J., Dong, W., Socher, R., Li, L., Li, K., Fei-Fei, L.: Bimagnet: A large-scale hierarchical image database. In: *CVPR*. pp. 248-255 (2009)
19. Cimpoi, M., Maji, S., Kokkinos, I., Mohamed, S., Vedaldi, A.: Describing textures in the wild. In: *CVPR*. pp. 3606-3613 (2014)
20. Zhong, S.h., Liu, Y., Chen, Q.c.: Visual orientation inhomogeneity based scale invariant feature transform. *Expert Systems with Applications* 42(13), 5658-5667 (2015)