The Visual Computer
https://doi.org/10.1007/s00371-020-01925-2

ORIGINAL ARTICLE

Learning wavelet coefficients for face super-resolution

Liu Ying[1] · Sun Dinghua[1] · Wang Fuping[2] · Lim Keng Pang[3] · Chiew Tuan Kiang[4] · Lai Yi[5]

Abstract
Face image super-resolution imaging is an important technology which can be utilized in crime scene investigations and public security. Modern CNN-based super-resolution produces excellent results in terms of peak signal-to-noise ratio and the structural similarity index (SSIM). However, perceptual quality is generally poor, and the details of the facial features are lost. To overcome this problem, we propose a novel deep neural network to predict the super-resolution wavelet coefficients in order to obtain clearer facial images. Firstly, this paper uses prior knowledge of face images to manually emphases relevant facial features with more attention. Then, a linear low-rank convolution in the network is used. Finally, image edge features from canny detector are applied to enhance super-resolution images during training. The experimental results show that the proposed method can achieve competitive PSNR and SSIM and produces images with much higher perceptual quality.

Keywords Deep learning · CNN · Wavelet · Super-resolution

1 Introduction

Face Super-Resolution (SR) is an important subset of image super-resolution technology for public security. Face SR computes Low-Resolution (LR) face images, which are often acquired by low quality surveillance camera, to estimate High-Resolution (HR) face images.

Due to the constraints of the environment, the faces acquired by surveillance cameras are unclear in many cases. One way is to upgrade the imaging system to a more expensive and higher resolution system [1]. However, it is cumbersome and expensive to realize. In addition, it cannot resolve the issue of small face images that were captured far away from the camera. Therefore, researchers have proposed SR algorithm to enhance the image quality, and SR is now widely used in most situations [1–6].

The intent of SR is to infer from LR images a priori information to obtain the HR images with clearer details. In single face image super-resolution, only one LR face image can be utlized to reconstruct the desired HR face image. Since desired HR face image has more pixels than the LR face image, it is a morbid inverse problem. Traditional solution is applying constraints based on the features of the face to the HR estimation process. These techniques can be broadly classified into three categories: interpolation, reconstruction, and learning-based methods [3]. Interpolation-based method [7] samples a given LR image and imposes smoothing constraints on the interpolation of missing information in the HR image. It is simple to implement, but the reconstructed image is blurry. Reconstruction-based method adds a priori knowledge that forces a constraint on the process of down sampling to generate the original LR image to reconstruct the HR image [7]. Learning-based method maps LR to HR images by learning the relationship between LR and its corresponding HR images. With the development of deep learning, the performance of learning-based methods has gradually surpassed all other SR methods.

In recent years, SR algorithms based on deep learning have attracted tremendous attention in super-resolution research community. Dong et al. [8] combined image super-resolution

✉ Sun Dinghua
d_t_sdh@163.com

1 Xi'an University of Posts and Telecommunications, Xi'an, China

2 Key Laboratory of Electronic Information Processing for Crime Scene Investigation, Ministry of Public Security, Xi'an, China

3 Xsecpro Pte Ltd, 449 Tagore Industrial Avenue, Great land Industrial Building, Singapore, Singapore

4 Rekindle Pte Ltd, 70 Gardenia Road, Singapore, Singapore

5 International Joint-Research Center for Wireless Communication and Information Processing, Xi'an, China

Published online: 21 July 2020

techniques with deep learning to design a Convolutional Neural Network (CNN) with only three layers of convolutional layers. Compared with the traditional super-resolution method based on sparse coding [9], the method in [8] has greatly improved the performance; however, SR image losses significant information due to the shallow network structure. Kim et al. [10] observed that the low-resolution image and its corresponding high-resolution image are similar, i.e., the low-frequency information of the LR image is similar to the low-frequency information of the high-resolution image. Therefore, if the residual of the high-frequency information between the high-resolution image and the low-resolution image can be accurately predicted [11], it is possible to obtain a high-quality SR image while reducing the computational burden. However, it is difficult to find a satisfactory threshold to achieve the best SR effect due to the gradient threshold strategy used during the training process. Most SR method based on deep learning initially used interpolation of the low-resolution image for a high-resolution image first before it is computed in the neural network, which incurred higher computational cost. Lai et al. [12] designed a multi-resolution CNN, which performs 2 times up sampling at each stage, predicting HR image step-by-step, thus reducing computation time. Christian et al. [13] think that although using MSE as loss function during training can obtain a high peak signal-to-noise ratio, the predicted images usually lose high-frequency details. Reference [13] uses perceptual loss and adversarial loss to improve the realism of the predicted image. Zhang Y et al. [14] proposed channel attention mechanism based on deep residual networks, adaptively learning channel characteristics by considering the interdependence between channels, thereby improving the perceptual quality of predicted images.Reference [15] designed a two-step deep network structure: coarse network (corresponding to coarse loss) and refinement network(corresponding to refinement loss and GAN loss). In addition, an attention mechanism is introduced to give higher weight to features similar to the missing parts in the impainting process. In reference [16], non-local operation is introduced into the end-to-end neural network to capture the correlation between features and their adjacent features, and it is proved that limiting the range of adjacent features is very important when calculating feature similarity. At the same time, the use of RNN improves the utilization rate of parameters and improves the robustness of the model. It is worth noting that super-resolution method based on CNN can achieve good performance in terms of peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) while the output images are often over smooth resulting in poorer perceptual quality.

Since Wavelet Transformation (WT) can perform multi-frequency analysis and preserve the edges of images well, wavelet-based SR method is often used in image processing [17]. Wavelet-based method performs wavelet transform on the high-resolution image to obtain wavelet sub-band coefficients. Features extracted from LR images are mapped to different sub-band wavelet coefficients. Generally, the features extracted from the LR image are used to predict the wavelet coefficients of unknown HR image, and then reconstruct the desired HR image from predicted coefficients. The task of wavelet-based method is to estimate unknown high frequency coefficients. Traditional solution is learning the scale dependence between low frequency coefficients and high frequency coefficients, and applies the mapping function to estimate detailed coefficients unknown [18,19]. Finding the missing high frequency coefficients accurately in LR images is still a challenging problem. With further development of deep learning, many methods based on estimating the high frequency coefficients have been proposed in [20–25].

In [22], a deep neural network model which combines wavelet transform and CNN is proposed to predict missing details in the wavelet coefficients of low-resolution images. Z. Zhong et al. [24] pointed out that CNN-based method has sharp performance degradation on extremely low-resolution image super-resolution tasks, and the output SR image is over-smoothed. We proposed using wavelet transform to decompose HR image to different sub-band coefficients, and using CNN to predict the coefficients of the HR from LR image to infer SR images. Huang H et al. [25] introduced the method of Generative Adversarial Networks (GAN) [26] in the wavelet-based deep SR network. In this method, wavelet coefficients predicted by CNN and wavelet coefficients decomposed by ground truth are trained. The resultant SR image appears more realistic.

Although wavelet-based deep learning method for super-resolution performs better in detailed texture reconstruction than CNN-based method, but wavelet-based deep learning SR methods lacks translation invariance property because it uses the structural information of the image for super-resolution. In order to overcome this problem, we proposed a new deep SR network, namely, wavelet-based face mask super-resolution network (MWSR). Our work is as follows: (1) A pre-trained segmentation network is used to obtain the facial mask by detecting the facial features of the human face. Data augmentation is then performed. The purpose of the first phase is to focus the attention on the facial features and enable translation invariance to the wavelet-based CNN. (2) We introduce the method of [27] to supplement the image edge with information extracted by the canny edge detection operator. (3) We introduce the linear low-rank convolution operation in the stage of feature embedding, which improves the accuracy of the predicted wavelet coefficients without increasing the computational cost.

2 Discrete wavelet transform

To improve the performance of the wavelet-based SR method, the relationship between high-frequency wavelet coefficients and its LR image was investigated. Reference [25] verified through experiments that the high-frequency wavelet coefficients of the image gradually decrease with increasing blurriness. Whether the high-frequency wavelet coefficients can be restored determines whether the obtained SR image is clear. CNN-based SR method has a sharp degradation in performance on very low-resolution images mainly due to the loss of high-frequency information. In order to reconstruct the high-frequency details of the image, this paper combines the wavelet discrete transform with the deep convolutional neural network to obtain a better SR image.

In this paper, Haar transform [28] is utilized to transform the two-dimensional image signal into four sub-bands using the low-pass and high-pass filters. The high-pass filter is processed horizontally, vertically and diagonally and the flow of the two-dimensional discrete wavelet transform is presented in Fig. 1. Firstly, this paper generates the detailed coefficient of the image by performing two-dimensional wavelet transform of the input image.

An example of the face image and the wavelet coefficients after two-dimensional discrete wavelet transform is presented in Fig. 2. The example face image is from the dataset CelebA [29]. Right part of Fig. 1 is the transformed domain using the two-dimensional discrete wavelet transform to capture the image details in the four sub-bands.

SR task can be considered as a problem of inferring an HR image containing image detail information from a LR image in which image detail information is missing. Wavelet decomposition provides an elegant structure to separate the LR information from the details as shown in Fig. 1 where D, H and V represent the detailed information.

In our proposed method, LR image was fed into an attention-based deep SR network, to predict D, H, V and A of the corresponding HR image using DWT. General DWT decomposition is shown in Fig. 1. After all DWT sub-bands are predicted, computed sub-band blocks are used to recover the SR image by two-dimensional discrete wavelet inverse transform as shown in Fig. 2.

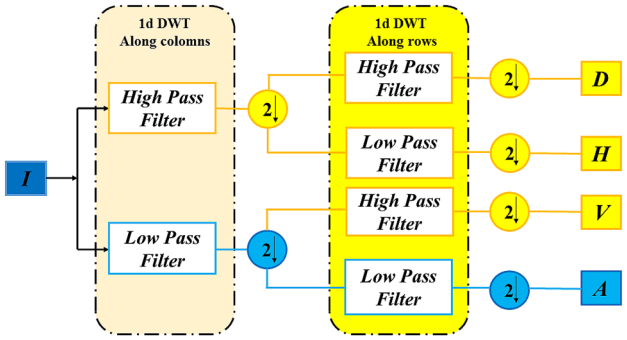

Fig. 1 2d discrete wavelet transform(DWT)

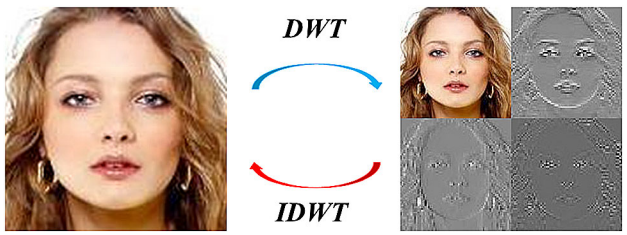

Fig. 2 Image super-resolution based on 2d DWT and 2d inverse discrete wavelet transform (IDWT)

3 Face mask wavelet-based super-resolution Network

3.1 Architecture

As shown in Fig. 2, MWSR consists of three sub-networks, which are mask generator network, attention-based feature embedding network, and wavelet coefficient prediction network.

We use a pre-trained semantic segmentation network to generate facial mask images during training, and we remove it from the MWSR while testing. Due to the problem of spatial information loss when the image is processed in CNN, this paper introduces linear low-rank convolution operation in the feature embedding stage of SR network without incurring additional computational burden to the convolution layer. The skip connection is applied in the wavelet coefficients prediction phase of the network so that it greatly reduces training to learn the low-frequency wavelet coefficients. Finally, two-dimensional inverse discrete wavelet transform is utilized to reconstruct the high-resolution image by using the predicted wavelet coefficients. Because of missing information while the information propagates in CNN, we introduce linear low-rank convolution operation in feature embedding. The wavelet coefficients prediction network adopts the design of residual connection, to greatly reduce training to learn the low-frequency information of the network. Finally, the wavelet coefficients obtained by the prediction network are used to reconstruct the high-resolution image by two-dimensional discrete wavelet inverse transform.

3.2 Facial mask

The visual attention mechanism is a characteristic of our human visual perception system. The human vision acquires the focus of the target area by quickly scanning the global image. The targeted area is generally called the focus of attention. Once targeted area is determined, it starts paying more

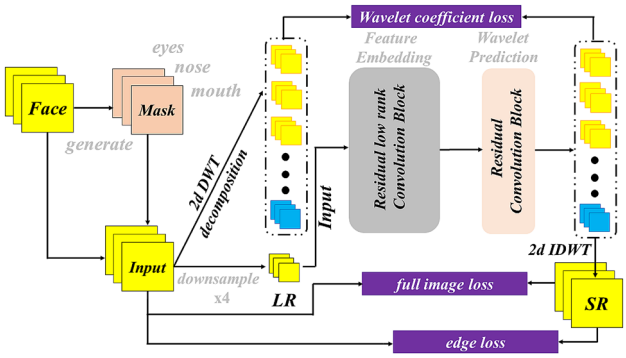

Fig. 3 The architecture of MWSR

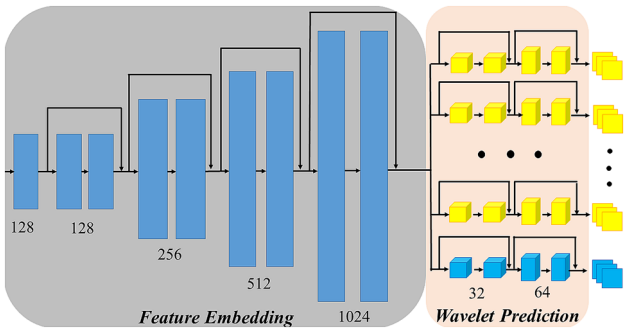

Fig. 4 The backbone of MWSR which is our supplement to implementation details of feature embedding and wavelet prediction mentioned in Fig. 3

attention to the area to obtain more details, ignoring other secondary or useless information.

Inspired by visual attention mechanism, the attention mechanism for deep learning aims to select information from a multitude of information that is more critical to the current mission objectives. Logically, the most recognizable position in a face image is the facial features. Most details of SR image inferred by CNN-based methods are lost [8,12,22,30,31]. Therefore, this paper designs a facial mask method which encourages more attention to the facial features while learning the mapping relationship between the LR face image and the HR face image. This method uses manual selection of a priori information in the face image to give more attention to the facial features.

It is realized by detecting the facial features from the pre-trained segmentation network [31] to generate a corresponding mask image to be trained together with the original face image. CNN can be seen as an approximator used to fit the mapping relationship between input and target, when the distribution of training data is less complex, the accuracy of CNN prediction is higher. It is worth noting that the small-angle rotation and translation operation of the generated face mask image can overcome the adverse effects of the wavelet-based method which inherently is not translation invariance. Some examples of facial mask are shown in Fig. 5. The first

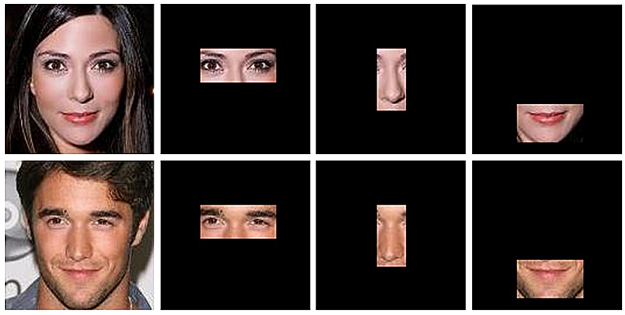

Fig. 5 Examples of face mask image

column on the left are the original face images, and the three columns on the right are the mask image generated by mask generator network.

3.3 Canny edge detector

The edge of images inferred by CNN-based SR methods is blurry because the loss of information during forward propagation. In order to supplement the edge of the image, we use the canny operator to extract the edge features of the face image and use it as a loss function during the training process. Experiments are verified that it can restore more details of our facial image. The image edge loss function is defined as:

$$l_{edge} = \left\| C\left(\tilde{I}_i\right) - C\left(I_i\right) \right\|_2^2 \tag{1}$$

where in formula (1), I_i refers to the i-th input image, $\tilde{I}_i$ refers to the prediction of I_i, and $C(\cdot)$ represents the canny edge detector.

3.4 Linear low-rank convolution

Earlier CNN-based SR networks [8,10,11] are inferior than deep residual CNN-based SR network which has more convolution layers [12,14,30]. In order to enhance the performance of SR Network, the depth of the SR Network can be increased by stacking the convolution layer.

However, when the depth of the SR network is increased to a certain extent, information propagation between the convolution layers is hindered. Researchers often use residual learning to overcome this problem, and residual connection can be combined with image texture and semantic features to generate better quality representations [12,13,13,14,32,33].

Due to the truncation effect of the rectified linear unit (ReLU) on CNN activation, some information is lost when the information flow is transmitted in the CNN [34]. To overcome this problem, the number of filters chan-

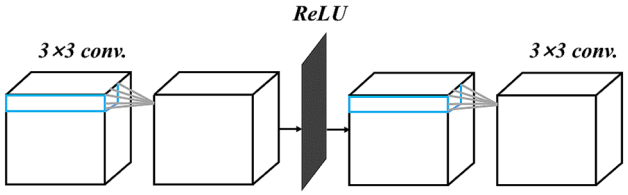

Fig. 6 Original convolution block

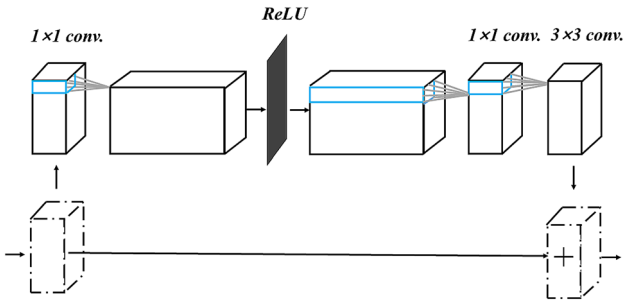

Fig. 7 Low-rank convolution block

nel is increased at the expense of higher computational cost.

Linear low-rank convolution operation was proposed to reduce the complexity while preserving the information flow through the networks. The architecture is shown in Figs. 6 and 7. Linear low-rank convolution operation greatly alleviates the phenomenon of information loss caused by the truncation effect of ReLU.

3.5 Loss function

In this paper, three types of loss functions are used: image edge loss, wavelet-based loss and image pixel-based loss. We use Mean Square Error (MSE) to compute the image pixel-based loss of the SR Network [17,35–38]. In [22,25], wavelet coefficient-based loss function is defined as the weighted sum of the sub-band coefficients and image texture-based loss function.

MSE is also commonly used in the wavelet loss function to compute the errors of the low- and high-frequency sub-band coefficients. However, it is observed in [39–41] that direct MSE loss function cannot achieve good perceptual SR images due to the characteristic distribution of the high-frequency sub-band coefficients of the face image. Thus, we redefine the loss function of the high-frequency sub-band coefficients and formulated as follows:

$$l_{HF}\left(\bar{o}_i, o_i\right) = \sum_{i}^{N} p\left(\bar{o}_i, o_i\right) \log\left(\frac{p\left(o_i\right)}{q\left(\bar{o}_i\right)}\right) \tag{2}$$

where o_i and $\bar{o}_i$ refers to ground truth high-frequency sub-band coefficients by wavelet decomposition and predicted high-frequency sub-band coefficient of the SR image, respectively. $p\left(o_i\right)$ denotes the characteristic distribution of o_i , and $q\left(\bar{o}_i\right)$ is the characteristic distribution of $\bar{o}_i$. Low-frequency sub-band coefficients uses MSE as the loss function, so wavelet coefficient-based loss function is defined in this paper as follows:

$$\begin{aligned} l_{\text{wavelet}}\left(\bar{o}_i, o_i\right) &= \alpha l_{HF}\left(\bar{o}_i, o_i\right) + \beta l_{LF}\left(\bar{o}_i, o_i\right) \\ &= \alpha \sum_{i}^{N} p\left(\bar{o}_i, o_i\right) \log\left(\frac{p\left(o_i\right)}{q\left(\bar{o}_i\right)}\right) + \beta \left\| \bar{o_i} - o_i \right\|_F^2 \end{aligned} \tag{3}$$

In the above formulation, α and β are hyperparameters and they are the weight of high- and low-frequency sub-band coefficients loss function, respectively.

Selecting MSE as a loss function in deep learning is usually a simple and efficient choice. And we have noticed in our experiments that in image pixel-based loss function, MSE usually gets better results on the evaluation criteria of PSNR. Therefore, it includes in the total objective function defined as:

$$l_{\text{total}} = l_{\text{wavelet}} + \eta l_{\text{image}} + \mu l_{\text{texture}} + \zeta l_{\text{edge}} \tag{4}$$

$$\begin{aligned} l_{\text{total}} = \alpha \sum_{i}^{N} p\left(\bar{o}_i, o_i\right) \log\left(\frac{p\left(o_i\right)}{q\left(\bar{o}_i\right)}\right) + \beta \left\| \bar{o_i}^{-} - o_i \right\|_F^2 \\ + \eta \left\| \tilde{I_i}^2 - I_i \right\|_F^2 + \mu \sum_{i}^{N} \max\left(\lambda \left\| o_i \right\|_F^2 + \varepsilon - \left\| \bar{o_i} \right\|_F^2, 0\right) \\ + \zeta \left\| C\left(\tilde{I}_i\right) - C\left(I_i\right) \right\|_2^2. \end{aligned} \tag{5}$$

η and μ in the formula are the weight of the image-based loss function and texture-based loss function, respectively. The role of λ and ε in the texture-based loss function is to ensure that the value of texture-based loss function is not zero.

MWSR has four sub-loss functions: wavelet loss, image loss, texture loss, and edge loss. Since our intention is to obtain SR image with clearer image edges, the weight of edge loss is appropriately increased. In addition, to predict the high-frequency detail information which is missing in LR images, high frequency coefficients should be given higher weight than that of low frequency coefficients. In order to reduce the negative effects of MSE, a smaller weight will be used for image loss. In our experiments, these weight parameters are all hyper-parameters. We empirically set α,β,η,μ, and ζ to 0.99, 0.01, 0.1, 1, and 1.2, respectively.

Table 1 Quanitative Resultd on CelebA and LFW Test Sets

Dataset	Method	PSNR(dB)	SSIM
CelebA	Bicubic	27.58	0.8453
CelebA	SRCNN	27.94	0.8916
CelebA	RCAN	31.60	0.9495
CelebA	SRGAN	30.02	0.9294
CelebA	U-Net	31.30	0.9447
CelebA	Wavelet-SRNet	30.45	0.9373
CelebA	Ours	31.58	0.9494
LFW	Bicubic	29.51	0.8755
LFW	SRCNN	29.29	0.9143
LFW	RCAN	33.30	0.9616
LFW	SRGAN	31.88	0.9499
LFW	U-Net	32.56	0.9555
LFW	Wavelet-SRNet	32.16	0.9514
LFW	Ours	33.34	0.9627

4 Experiment

The experiments in this paper are implemented on two public face data sets, CelebA [29] and LFW [42]. We selected 10,416 face images from CelebA as training set and 9,230 face images as validation set. At the same time, we selected 1,063 and 653 face images from CelebA and LFW as test sets. All face images are cropped and aligned with a size of 128×128.

We tested SR performance of MWSR by 4 times downsampling the original face image, and compared among some of the classic super-resolution methods. To be more complete, we compared with BICUBIC interpolation method, SRCNN [8], U-Net [31], SRGAN [13], RCAN [14], Wavelet-SRNet [22]. We trained all methods with the same CelebA and LFW training sets. In this paper, PSNR and SSIM are used to evaluate the SR performance of the above method.

Table 1 summarizes the results of the PSNR and SSIM evaluations. The best results are shown in red and the second best results in blue. As can be seen from Table 1, MWSR exceeds the previously mentioned algorithms in the evaluation criteria of PSNR and SSIM. We performed ablation experiments on MWSR as following. Ablation experiments of MWSR investigate the performance of exclusion of mask generation network, edge loss and linear convolution operation separately on the same data set. The result of ablation experiment is shown in Figs. 8 and 9.

Reference [43] points out that both PSNR and SSIM have limitations in evaluating quality of real-world images, hence, we use a new evaluation metric (perceptual similarity) in our experiments. Reference [44] provides a wide-ranging and highly differentiated perceptual similarity dataset, which uses traditional methods (light adjustment, Gaussian kernel blur, noise addition, deformation, color change, etc.) and deep learning methods (denoising, style transfer, encoding, and decoding) to process the ground truth to generate two noise image corresponding to the ground truth. This data set uses the visual perception of different people (number 484k) to determine which noise image is closer to the ground truth, and uses this as an annotation. And the visual perception

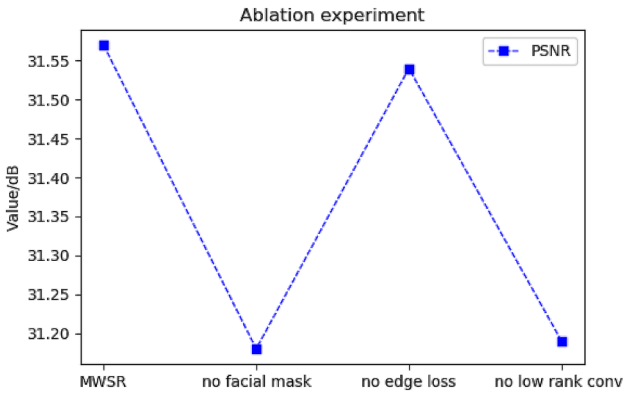

Fig. 8 The PSNR of ablation experiment

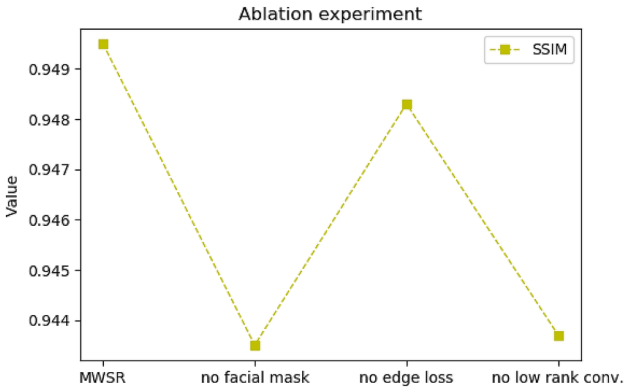

Fig. 9 The SSIM of ablation experiment

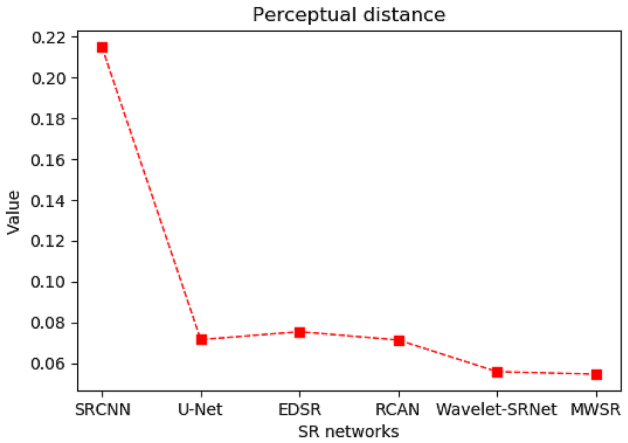

Fig. 10 Perceptual distance of results obtained by classic super-resolution methods

Fig. 11 Comparison with classic super-resolution methods

of different people (484k) is used to determine which noise image is closer to the ground truth as the annotation of the data set. Based on this data set, the authors propose a new perceptual similarity measurement, which performs better than PSNR and SSIM in simulating the underlying perceptual similarity. The results are shown in Fig. 10. The smaller the value of perceptual similarity, the better subjective quality of SR image. Compared with CNN-based networks, the proposed method MWSR shows better subjective quality.

Figure 11 shows the visual quality of SR results from the 4 times down sampled low-resolution input face image using MWSR and SR results of some of the state-of-the-art algorithms.

As can be seen from Table 1 and Fig. 11, although RCAN achieves better performance than Wavelet-SRNet in terms of PSNR and SSIM, the tooth gap is less pronounced than the results of Wavelet-SRNet. The images predicted by SRGAN are affected by the data set with obviously noise. And the SR face image inferred by MWSR recovered the facial features best, especially the details of the teeth can be clearly seen. Compared with classic CNN-based super-resolution methods, MWSR achieved better PSNR results, and recover more facial features.

CNN-based methods generally use MSE as the loss function. As MSE has the function of averaging, SR image derived by using methods such as SRCNN and RCAN could be blurred. Wavelet-SRNet uses deep convolution layers to accurately predict the high-frequency wavelet coefficients describing image details from LR image, then high-quality SR image can be reconstructed using inverse wavelet transform. Inspired by this work, we also use wavelet transform to predict high-frequency wavelet coefficients, and use a pre-trained segmentation network to generate facial mask images, giving higher attention to facial features in face super-resolution. At the same time, under the constraint of the edge loss function, MWSR can obtain better and

clearer image edges in the image reconstruction stage, further improving the subjective quality of SR images.

5 Conclusion

Using CNN-based SR network can perform very well in terms of PSNR and SSIM by simply stacking more residual connection to realize extremely deep networks. However, the reconstructed image is often over-smoothed and for face SR application, facial features are lost.

Wavelet-based neural networks have better subjective quality than direct CNN-based SR algorithm although it has inferior PSNR/SSIM results. By improving wavelet-based neural network in [24], both improvements in subjective and objective metrics can be achieved which shows that wavelet-based approach has more potential for further improvements.

In this paper, a wavelet-based face image SR algorithm is proposed by using a facial mask to help trained the attention-based neural network.

The neural network learns the relationship between the wavelet coefficients of the LR face image and the HR face image by paying more attention on facial features. Wavelet structure inherently separates the low-frequency information from the details by storing this information in different sub-bands. This helps MWSR to predict SR wavelet coefficients in the different sub-bands which have the same size as LR face, thus simplifying the mapping relationship to be learned.

The masking operation allows the network to focus on the facial features, further reducing the computational burden and enhancing the accuracy of the network. Therefore, compared with most existing methods, MWSR has achieved competitive results in terms of PSNR and SSIM, as well as the best visual perceptual quality.

Compliance with ethical standards

Conflict of interest This work was supported in part by National Natural Science Foundation of China (Number 61801381). The authors declare that they have no conflict of interest.

References

1. Katsaggelos, A.K.: Digital Image Restoration, pp. 2–3. Prentice Hall, Upper Saddle River (1977)
2. Sun J., Xu Z., Shum H. Y.: Image super-resolution using gradient profile prior. In: 2008 IEEE Conference on Computer Vision and Pattern Recognition. Alaska: IEEE, pp. 1–8 (2008)
3. Park, S.C., Park, M.K., Kang, M.G.: Super-resolution image reconstruction: a technical overview. IEEE Sig. Process. Mag. **20**(3), 21–36 (2003)
4. Zhang, X., Tang, M., Tong, R., et al.: Robust super resolution of compressed video. Vis. Comput. **28**(12), 1167–1180 (2012)
5. Mikaeli, E., Aghagolzadeh, A., Azghani, M., et al.: Single-image super-resolution via patch-based and group-based local smoothness modeling[J]. The Visual Computer 1–17 (2019)
6. Xu, K., Wang, X., Yang, X., et al.: Efficient image super-resolution integration. Vis. Comput. **34**(6), 1065–1076 (2018)
7. Li, X., Orchard, M.T.: New edge-directed interpolation. IEEE Trans. Image Process. **10**(10), 1521–1527 (2001)
8. Dong, C., Loy, C.C., He, K., et al.: Image super-resolution using deep convolutional networks[J]. IEEE Trans on Pattern Analysis and Machine Intelligence **38**(2), 295–307 (2015)
9. Yang, J., Wright, J., Huang, T.S., et al.: Image super-resolution via sparse representation. IEEE Trans. Image Process. **19**(11), 2861–2873 (2010)
10. Kim, J., Kwon Lee, J., Mu Lee, K.,: Accurate image super-resolution using very deep convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, pp. 1646–1654 (2016)
11. He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, pp. 770–778 (2016)
12. Lai, W. S., Huang J. B., Ahuja, N., et al. Deep Laplacian pyramid networks for fast and accurate super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Hawaii: IEEE, pp. 624–632 (2017)
13. Ledig, C., Theis, L., Huszr, F., et al. Photo-realistic single image super-resolution using a generative adversarial network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, pp. 4681–4690 (2017)
14. Zhang Y, Li K, Li K, et al. Image super-resolution using very deep residual channel attention networks. In: Proceedings of the European Conference on Computer Vision. pp. 286–301(2018)
15. Yu, J., Lin, Z., Yang, J., et al. Generative image inpainting with contextual attention. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5505–5514 (2018)
16. Liu, D., Wen, B., Fan, Y., et al. Non-local recurrent network for image restoration. In: Advances in Neural Information Processing Systems, pp. 1673–1682 (2018)
17. Kim S S, Eom I I K, Kim Y S. Image interpolation based on statistical relationship between wavelet sub-bands[C]. 2007 IEEE International Conference on Multimedia and Expo. Beijing: IEEE, 2007: 1723-1726
18. Tian, J., Ma, L., Yu, W.: Ant colony optimization for wavelet-based image interpolation using a three-component exponential mixture model. Expert Syst. Appl. **38**(10), 12514–12520 (2011)
19. Woo, D. H., Eom, I. K., Kim, Y. S.: Image interpolation based on inter-scale dependency in wavelet domain. In: 2004 International Conference on Image Processing. Singapore: IEEE, 2004, 3: pp. 1687–1690 (2004)
20. Kumar, N., Rai, N. K., Sethi, A.: Learning to predict super resolution wavelet coefficients. In: Proceedings of the 21st International Conference on Pattern Recognition. Tsukuba: IEEE, pp. 3468–3471 (2012)

21. Gao, X., Xiong, H. A.: hybrid wavelet convolution network with sparse-coding for image super-resolution. In: 2016 IEEE International Conference on Image Processing. Phoenix: IEEE, pp. 1439–1443 (2016)
22. Huang, H., He, R., Sun, Z., et al. Wavelet-SRNet: a wavelet-based CNN for multi-scale face super resolution. In: Proceedings of the IEEE International Conference on Computer Vision. Venice: IEEE, pp. 1689–1697 (2017)
23. Guo, T., Seyed Mousavi, H., Huu, V. T., et al. Deep wavelet prediction for image super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. Honolulu: IEEE, 2017 pp. 104–113 (2017)
24. Zhong, Z., Shen, T., Yang, Y., et al. Joint sub-bands learning with clique structures for wavelet domain super-resolution. In: Advances in Neural Information Processing Systems, pp. 165–175 (2018)
25. Huang, H., He, R., Sun, Z., et al.: Wavelet domain generative adversarial network for multi-scale face hallucination. Int. J. Comput. Vis. **127**(6–7), 763–784 (2019)
26. Goodfellow, I., Pouget-Abadie, J., Mirza, M., et al. Generative adversarial nets. In: Advances in Neural Information Processing Systems. Montreal: NIPS'14, pp. 2672–2680 (2014)
27. Canny, J.: A computational approach to edge detection. Readings in computer vision. Morgan Kaufmann, pp. 184–203 (1987)
28. Mallat, S.G.: A theory for multiresolution signal decomposition: the wavelet representation. IEEE Trans. Pattern Anal. Mach. Intell. **7**, 674–693 (1989)
29. Liu, Z., Luo, P., Wang, X., et al. Deep learning face attributes in the wild. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 3730–3738 (2015)
30. Lim, B., Son, S., Kim, H., et al.: Enhanced deep residual networks for single image super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, pp. 136–144 (2017)
31. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-assisted Intervention. pp. 234–241 (2015)
32. Caballero, J., Ledig, C., Aitken, A., et al.: Real-time video super-resolution with spatio-temporal networks and motion compensation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, pp. 4778–4787 (2017)
33. Huang, G., Liu, Z., Van Der Maaten, L., et al.: Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, pp. 4700–4708 (2017)
34. Tong, T., Li, G., Liu, X., et al.: Image super-resolution using dense skip connections. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 4799–4807 (2017)
35. Szegedy, C., Liu, W., Jia, Y., et al.: Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE, pp. 1–9 (2015)
36. Dong, C., Loy, C. C., Tang, X.: Accelerating the super-resolution convolutional neural network. In: European Conference on Computer Vision. pp. 391–407 (2016)
37. Yu, X., Porikli, F.: Ultra-resolving face images by discriminative generative networks. In: European Conference on Computer Vision. pp. 318–333 (2016)
38. Zhang, Y., Tian, Y., Kong, Y., et al.: Residual dense network for image super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, pp. 2472–2481 (2018)
39. Dahl, R., Norouzi, M., Shlens, J.: Pixel recursive super resolution. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 5439–5448 (2017)
40. Isola, P., Zhu, J. Y., Zhou, T., et al. Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, pp. 1125–1134 (2017)
41. Zhao, H., Gallo, O., Frosio, I., et al.: Loss functions for image restoration with neural networks. IEEE Trans. Comput. Imag. **3**(1), 47–57 (2016)
42. Lu, C., Tang, X.: Surpassing human-level face verification performance on LFW with GaussianFace. In: 20-th AAAI Conference on Artificial Intelligence (2015)
43. Dranoshchuk, A. D., Veselov, A. I.: About perceptual quality estimation for image compression. In: 2019 Wave Electronics and its Application in Information and Telecommunication Systems (WECONF). IEEE (2019)
44. Zhang, R., Isola, P., Efros, A. A., et al. The unreasonable effectiveness of deep features as a perceptual metric. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 586–595 (2018)

Liu Ying was born in 1972. She received the Ph.D. degree from Monash University in Australia in 2007. She is a professor at Xi'an University of Posts and Telecommunications. Her research interests include image retrieval, image enhancement, etc.

Sun Dinghua was born in 1995. He is a M.S. candidate at Xi'an University of Posts and Telecommunications. His research interest is image super-resolution reconstruction.

Wang Fuping received the B.Eng. degree and the Ph.D. degree in signal and information processing from Xidian University, Xi'an, China, in 2011 and 2017, respectively. He is currently a Lecturer with the Xi'an University of Posts and Telecommunications. His research interests include pattern recognition and image processing.

Lim Keng Pang was born in 1969. He received the Ph.D. degree from Nanyang Technological University of Singapore in 2001. He is CEO of Singapore Xsecpro Pte Ltd and Distinguished Professor of Xi'an University of Posts and Telecommunications. His research interests include video coding and image enhancement, etc.

Chiew Tuan Kiang was born in 1967. He Graduated from National University of Singapore with B. Eng (1st Class Honors) and received PhD in Electrical and Electronic Engineering from University of Bristol. He worked in Eti-mad RND (Abu Dhabi), Rekindle Pte Ltd, D'Crypt, STL, I2R-ASTAR. His research interests include Embedded Systems, Energy Management, Media Processing and Data Analysis.

Lai Yi received his PhD degree in pattern recognition and intelligent system from Xi'an Jiaotong University in 2013, and is currently a lecturer at the Institute of Image and Information Processing at Xi'an University of Posts and Telecommunications. His current research interests include image processing and analysis, computer vision and pattern recognition.