Neurocomputing 398 (2020) 531–538

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Single image super-resolution with enhanced Laplacian pyramid network via conditional generative adversarial learning

Xiaolu Zhang, Huihui Song, Kaihua Zhang*, Jiaojiao Qiao, Qingshan Liu

Jiangsu Key Laboratory of Big Data Analysis Technology, Nanjing University of Information Science and Technology, Nanjing, China

ARTICLE INFO

Article history:
Received 1 December 2018
Revised 8 April 2019
Accepted 11 April 2019
Available online 25 July 2019

Keywords:
Single image super-resolution
GAN
Conditional GAN
Laplacian pyramid,

ABSTRACT

Despite much progress has been made by applying generative adversarial network (GAN) to single image super-resolution (SISR), obvious difference remains between the details of reconstructed high-frequency and ground-truth because GAN is unstable that has a very high degree of freedom. To address this issue, we exploit conditional GAN (CGAN) for SISR, which leverages the ground-truth high-resolution (HR) image as its conditional variable to guide to learn a more stable model. To better reconstruct image with a large-scale factor, we further design an enhanced Laplacian pyramid network (ELapN) as the generator model of CGAN, which progressively reconstructs HR images at multiple pyramid levels. The proposed ELapN fuses low-and high-level features for the residual image learning achieves better generalization than those only based on high-level information. Finally, we train the proposed network via deep supervision using a combination of multi-level CGAN, VGG and robust Charbonnier loss functions to obtain high-quality SR results. Extensive evaluations on three benchmark datasets including Set5, Set14, B100 demonstrate superiority of the proposed method over state-of-the-art methods in terms of PSNR, SSIM and visual effect.

1. Introduction

The studies of SISR have attracted increasing attention in the deep learning, aiming to reconstruct an HR image based on a single input low-resolution (LR) image [1–6]. With demonstrated breakthrough progress in a variety of vision tasks [7–9], the deep convolutional neural networks (CNNs) have been successfully applied to SISR that also achieve significant performance improvement compared to the traditional ones [3]. Among them, Dong et al. [10] first introduce a three-layer CNN for SISR. Afterwards, much deeper CNNs are applied to SISR with promising performance. Kim et al. [11] present a very deep CNN for SISR that increases the network depth to 20 layers. Lim et al. [12] propose an enhanced deep CNN that removes unnecessary modules in conventional residual networks, leading to significant performance boosting. However, the CNN based approaches largely concentrate on minimizing the mean squared reconstruction error, leading to perceptually unpleasing results with less high-frequency details [13].

To address this issue, recently, GAN [14] has become more and more popular in the field of SISR [13,15–18]. Different from the CNN based SR that focuses on minimizing the mean squared reconstruction error, GAN trains a generative model by minimizing a loss that tries to distinguish whether the output HR image is real or fake. Ledig et al. [13] propose SRGAN that reconstructs the HR images under GAN framework and trains the GAN by combining pixel-wise mean squared error (MSE) loss, perceptual loss and adversarial loss to generate realistic image. Despite demonstrated success of GAN based SISR, unsatisfying HR outputs may be generated due to the intrinsic limitations of GAN: First, the adversarial loss function cannot reflect the degree of network training, and hence when training the network, the generator always tends to generate some samples in the real data with high scores. Second, the degree of freedom of GAN in the unconditional generation model is very large, making it difficult to control the mode to generate desirable data, resulting in the reconstructed image having some strange textures as shown by Fig. 1.

To resolve these problems of GAN [14], Mirza and Osindero [19] improve GAN with conditional constraint, which takes conditional variables as the additional information of the model to restrain its generation process. After adding the instructive conditional variables, the model has a large number of overlapping parts between the generator's output distribution and the real data distribution, the training of the network becomes more directional and easier. These conditional variables can not only be class labels [19], but also the data in different modalities [20,21]. Motivated by this, in this paper, we leverage CGAN with ground-truth HR image as conditional variable to learn a generative model for

* Corresponding author.
E-mail address: zhkhua@gmail.com (K. Zhang).

https://doi.org/10.1016/j.neucom.2019.04.097
0925-2312/

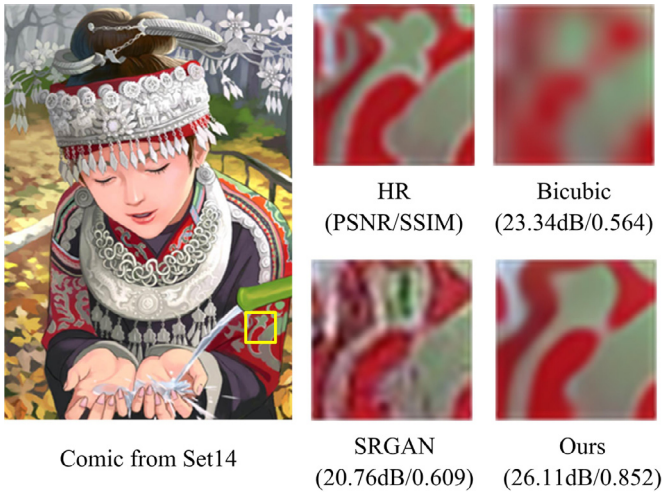

Fig. 1. 4× SR results of our method compared with Bicubic and SRGAN [13] methods.

SISR. Moreover, to better reconstruct image with a large-scale factor, motivated by the deep Laplacian pyramid SR network (LapSRN) [22], we propose an enhanced Laplacian pyramid network (ELapN) as the generator network in CGAN that progressively reconstructs HR images at multiple pyramid levels. Besides, existing residual learning based SR methods [11,12] only leverage the top-layer features to learn the residual image, which is ineffective to learn high-frequency details for any images because the top-layer features contain more high-level semantic information that has less generalization capability than the low-level features. To solve this problem, we concatenate low-and high-level features along the channel axis to learn the residual image. Extensive evaluations on several benchmark datasets demonstrate that the proposed method termed as ELapCGAN achieves better PSNR, SSIM and visual effect than state-of-the-art methods.

The main contributions of our ELapCGAN are summarized as follows:

- *Conditional generative adversarial learning.* We extend GAN to CGAN for SISR, which regards the ground-truth HR image as the condition variable to guide the discriminator network learning, making the training of the network more directional and easier.
- *Enhanced Laplacian pyramid network.* We build an enhanced Laplacian pyramid network, which concentrates on creating more feature channels to enhance feature learning ability for SISR.
- *Feature concatenation.* We fuse the low-and high-level features for residual image learning, which enables to improve the network generalization ability for a variety of images.

2. Proposed approach

2.1. Conditional generative adversarial learning

Inspired by the work of Isola et al. [21], we present a CGAN for SISR that is conditioned on the ground-truth HR image. Fig. 2 shows the training process of our CGAN. Specifically, given the input LR image **x**, when optimizing the generator *G*, the output of the generator $G(\mathbf{x})$ tries to confuse the discriminator *D* as much as possible so as to judge it to be true. Meanwhile, when optimizing the discriminator *D*, it receives the input image conditioned on the ground-truth **y**, and distinguishes whether the input

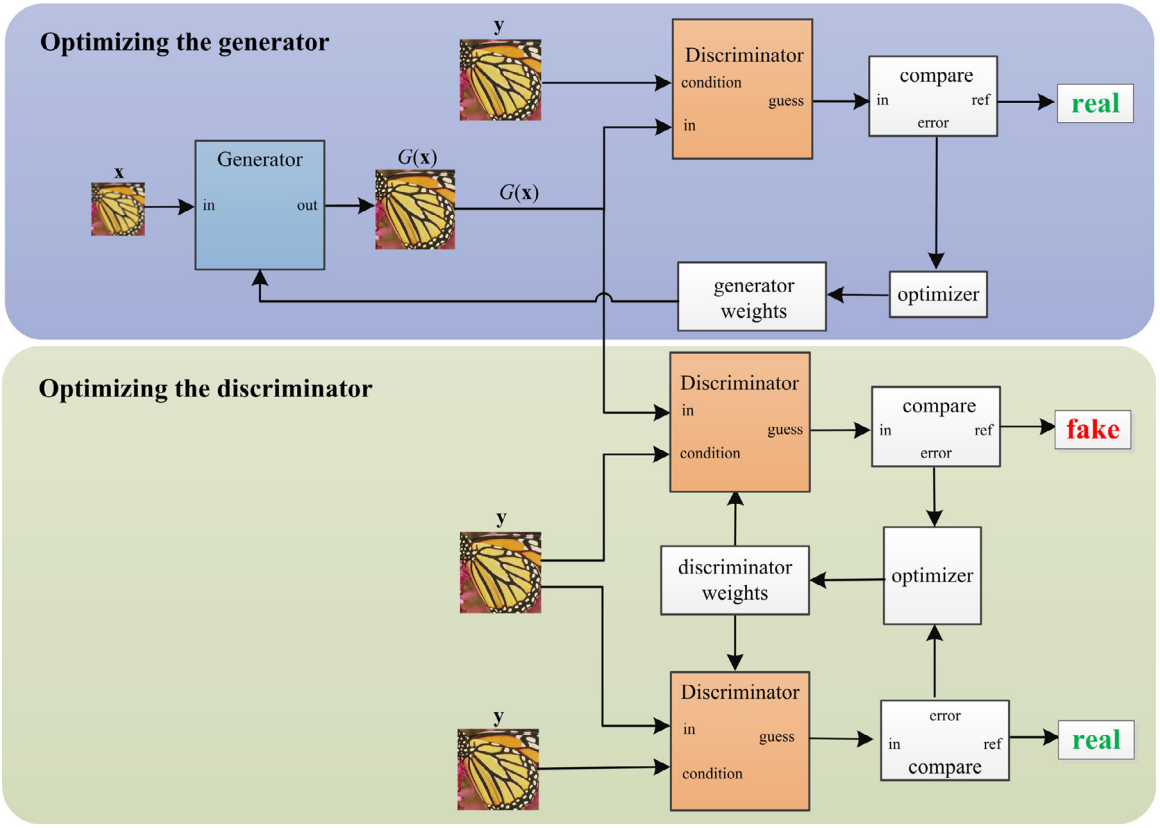

Fig. 2. Optimization process for CGAN. We add ground-truth **y** as the conditional variable to each discriminator. Given the input LR image **x**, the generated image $G(\mathbf{x})$ tries its best to fool the discriminator *D*, when optimizing the generator *G*. Meanwhile, the discriminator tries its best to distinguish the generated image $G(\mathbf{x})$ and the ground-truth **y**, when optimizing the discriminator *D*.

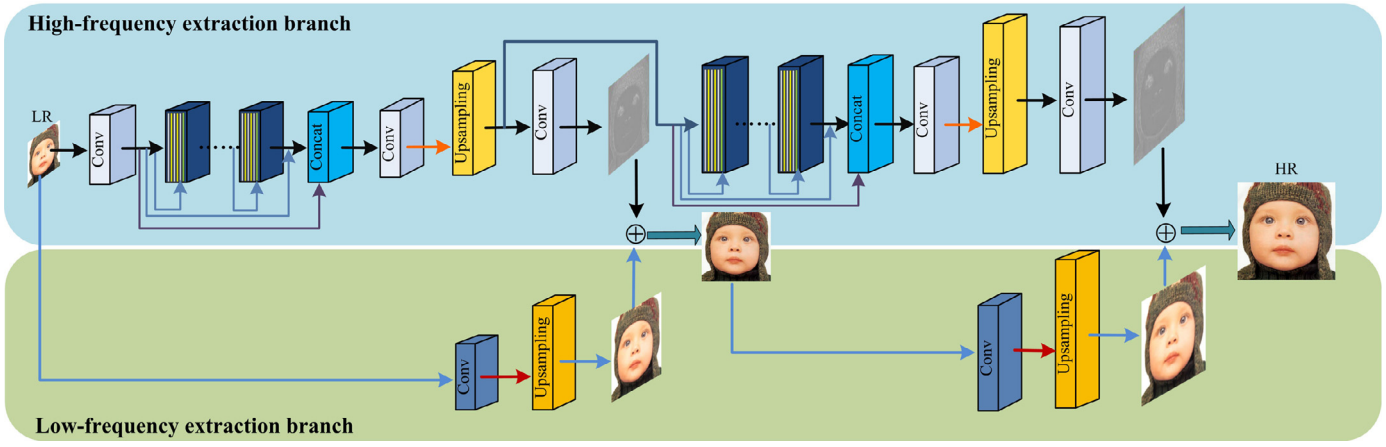

Fig. 3. Sketch architecture of the proposed ELapN for 4× SR.

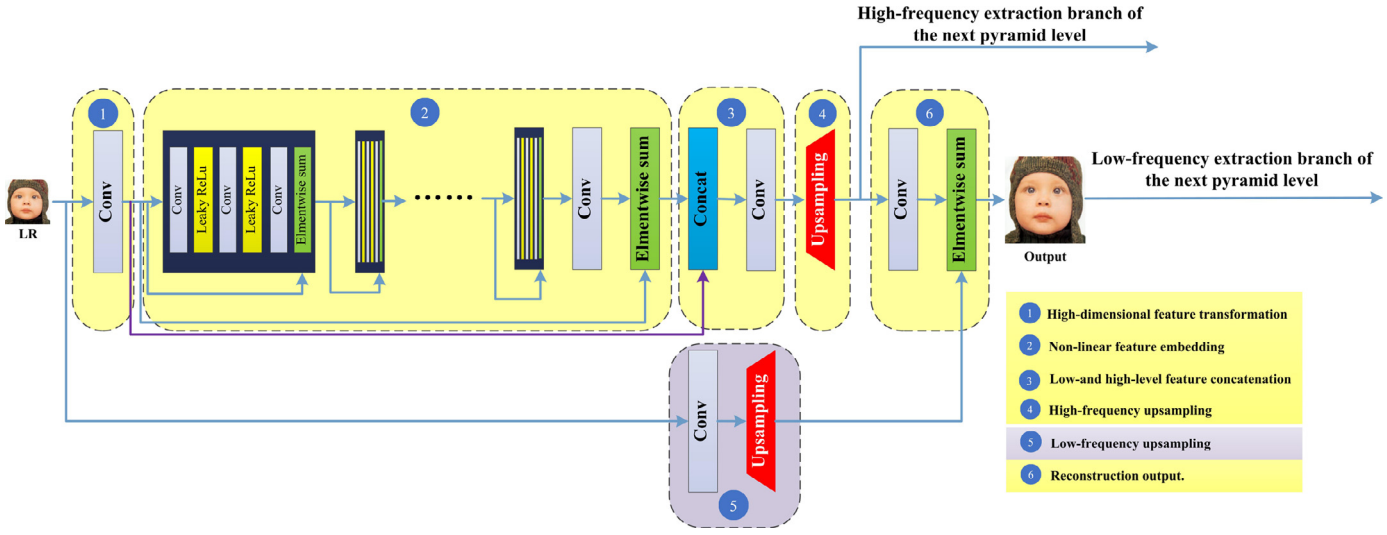

Fig. 4. Detailed architecture of the proposed ELapN at each pyramid level.

image is taken from the real image **y** or "forged" by the generator $G(\mathbf{x})$. When the input image **x** is from the real image **y**, the target of the discriminator D is to judge that it is "real", and when the input image comes from the generator G, the target of the discriminator D is to discriminate it as been "fake". The generator and discriminator in the entire adversarial learning process are similar to two players in the zero-sum game.

It should be noted that for GAN, which is based on unsupervised learning, its generator can well fit an arbitrary distribution, making it become extremely uncontrollable for reconstructing HR image from LR input. Our CGAN leverages the ground-truth HR image as a condition variable which is equivalent to building a bridge between the "fake" image generated by the generator and the ground-truth, and gives them a commonly discriminating direction. The CGAN guides the output distribution of the generator to be consistent with the target distribution as much as possible, making the reconstructed image resemblant to the real image as closely as possible.

2.2. Network architecture

Our CGAN consists of two "adversarial" models: generator and discriminator. The generator model is based on the architecture of LapSRN [22] and the residual block networks therein are replaced by our proposed modules. The discriminator model is based on the VGG19 network [23]. The detailed network structure is introduced as follows.

2.2.1. Enhanced Laplacian pyramid network

As shown in Fig. 3, we propose an enhanced Laplacian pyramid network (ELapN) for the generator model, which mainly consists of two branches: a high-frequency extraction branch (related to residual images or image details), and a low-frequency extraction branch. The two branches progressively reconstruct the HR images in a coarse-to-fine manner.

The proposed ELapN is a cascade of CNNs with the same network structure at each pyramid level. If the input image is an LR image downsampled by $1/S$ times (S is an upsampling scale factor), then the entire network needs $\log_2 S$-level pyramids to predict the HR image. As shown in Fig. 4, the detailed network architecture for each cascade is structured as follows:

1. *High-dimensional feature transformation.* We use a convolutional layer to transform the input LR image into high-dimensional feature maps.
2. *Non-linear feature embedding.* It is the core of the generator network. Compared to the LapSRN and SRGAN methods [13,22], we make some effective extensions: First, different from SRGAN [13] that extracts non-linear features directly from the LR image, our network only extracts non-linear features from the residual image, thus the training is easier to converge,

making the network enable to learn more effective features. Second, different from LapSRN [22] that only uses short skip connections, we stack long and short skip connections, allowing more rich information to bypass through these identity-based skip connections. Third, inspired by the enhanced deep SR method [12], we remove the redundant batch normalization (BN) layers used in SRGAN [13] in each residual block and stack more convolutional layers in a residual block, further boosting the performance. Although we use more convolutional layers than SRGAN [13] in a residual block, only 7 residual blocks are enough at each pyramid level, which is less than SRGAN [13] that cascades 16 residual blocks.

3. *Low-and high-level feature concatenation.* Existing related works [11,12] only leverage high-level features to extract the high-frequency information, which are ineffective to learn high-frequency details for any images because the high-level features contain more semantic information that cannot generalize better than the low-level features. On the contrary, we exploit the interdependencies among feature channels and concatenate low-level and high-level features along the channel axis to learn better residual image with more high-frequency details.
4. *High-frequency upsampling.* We use a sub-pixel convolutional layer for upsampling. The LapSRN [22] uses a transposed convolutional layer for upsampling, leading to losing the reconstructed high-frequency details of the HR image. Furthermore, the transposed convolution operation in LapSRN [22] is less efficient than the sub-pixel convolutional operation used by our method [24].
5. *Low-frequency upsampling.* To efficiently transfer more low-frequency information of the input LR image to the SR image, we only resort to a convolutional layer and a sub-pixel upsampling layer at each pyramid level.
6. *Reconstruction output.* SRGAN [13] directly combines both low- and high-frequency information and then upsamples the combined features by a sup-pixel upsampling layer, which may result in interference between low-frequency and high-frequency. To avoid this, we separately upsample the features of low-and high-frequency information via two different sub-pixel upsampling layers, and then combine them together to output the HR image at each pyramid level.

2.2.2. Discriminator network

In CGAN, the discriminator network is to distinguish the real data source from the generator "forgery" source. Our discriminator network has the same architecture as the VGG19 network, where we replace the ReLU activation function by the Leaky ReLU function, and the number of channels of the convolutional kernels is increased by a factor of 2 from 64 to 512. Then a fully connected layer and a sigmoid activation function are used to obtain the probability of the sample class.

2.3. Loss function

Our loss function is a combination of the CGAN, Charbonnier, and VGG losses as

$$\mathcal{L} = \alpha \mathcal{L}_{\text{CGAN}} + \beta \mathcal{L}_C + \gamma \mathcal{L}_{VGG}. \tag{1}$$

where $\mathcal{L}_{\text{CGAN}}$, $\mathcal{L}_C$ and $\mathcal{L}_{VGG}$ are defined by (2)–(4) respectively, α, β and γ are trade-off parameters.

2.3.1. CGAN loss

By optimizing the discriminator D and the generator G alternately, the loss function for CGAN is defined as

$$\mathcal{L}_{CGAN}(G, D) = \mathbb{E}_{\mathbf{y}}[\log D(\mathbf{y}, \mathbf{y})] + \mathbb{E}_{\mathbf{x},\mathbf{y}}[\log(1 - D(\mathbf{y}, G(\mathbf{x}))], \tag{2}$$

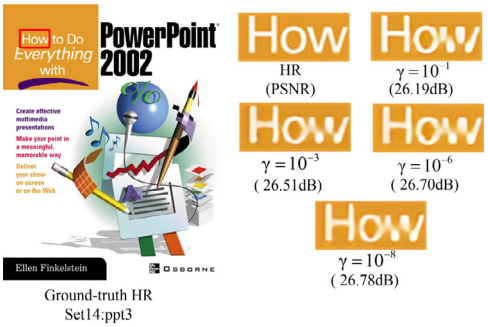

Fig. 5. The comparison of 4 × results using different VGG loss parameters γ with 100 epochs.

where $\mathbf{x}$ is the input image and $\mathbf{y}$ is the ground-truth. The generator G is optimized alternatively along with the discriminator D, i.e., $G^{\star} = \min_G \max_D \mathcal{L}_{CGAN}(G, D)$. Both the generator G and the discriminator D therein are nonlinear mapping functions. The D network is used to estimate the probability that samples from the ground-truth datasets rather than the G network. The G network is used to learn the distribution close to the ground-truth datasets.

2.3.2. Charbonnier loss

The MSE loss is usually used to optimize the network in SR. However, an LR patch is usually affected by multiple HR patches in the process of downsampling, so that the MSE loss cannot capture the underlying multi-modal distributions of the HR patches, making the reconstructed HR image over-smoothed. To address this issue, we leverage the Chabonnier loss function [25] that is robust to outliers.

Given the input LR image $\mathbf{x}$, our aim is to learn the mapping function G to generate the HR image $\hat{\mathbf{y}} = G(\mathbf{x})$, and make the generated $\hat{\mathbf{y}}$ approach to the ground-truth $\mathbf{y}$. We use bicubic downsampling to adjust the ground-truth $\mathbf{y}$ to the image $\mathbf{y}_l$ with size appropriate for the level l. Since our ELapN is a cascaded network, where $\mathbf{y}_l$ and $\hat{\mathbf{y}}_l$ are the ground-truth and the reconstructed image of the level l respectively, the overall loss function is defined as

$$\mathcal{L}_C(\mathbf{y}, \hat{\mathbf{y}}) = \sum_{l=1}^{\log_2 S} \sqrt{\|\mathbf{y}_l - \hat{\mathbf{y}}_l\|_2^2 + \varepsilon^2}, \tag{3}$$

where ε=1e−3 and S is the upsampling scale factor.

2.3.3. VGG loss

As SRGAN [13], we define the VGG loss function using a Leaky ReLU activation layer based on the pre-trained VGG19 network as

$$\mathcal{L}_{VGG} = \sum_i \frac{1}{w_i h_i c_i} \|\phi_i(\mathbf{y}) - \phi_i(\hat{\mathbf{y}})\|_2^2, \tag{4}$$

where $\hat{\mathbf{y}} = G(\mathbf{x})$ denotes the output of ELapN for input LR image $\mathbf{x}$ and $\mathbf{y}$ denotes the ground truth, the feature map ϕ_i denotes the output of the ith Leaky ReLU activation layer of the VGG19 network, and its corresponding width, height and channel number are w_i, h_i and c_i respectively.

3. Experimental results

3.1. Experimental setup

Each convolutional layer in our network has 256 filters with size 3 × 3. Moreover, to solve the issue that the convolutional feature maps become smaller and smaller, we use zero-padding for each convolutional feature map to keep the size of all the feature maps the same at each pyramid level. Besides, we adopt the Leaky

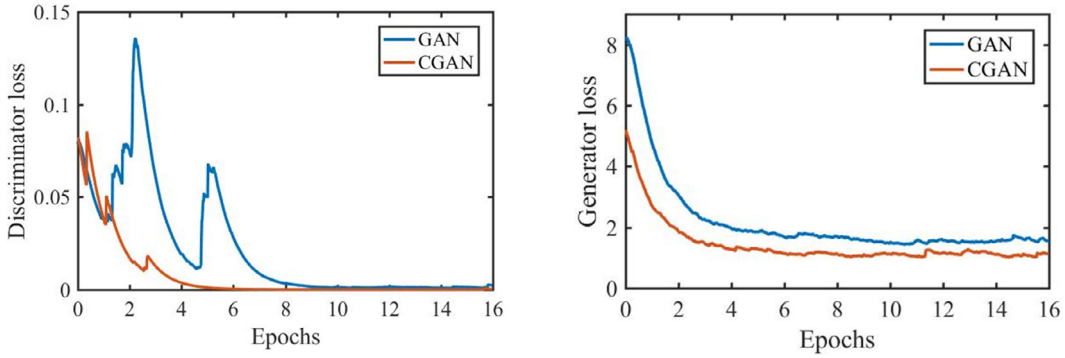

Fig. 6. Plots of discriminator loss and generator loss over epochs for GAN and CGAN on Set5. All training curves were used the same moving average parameters.

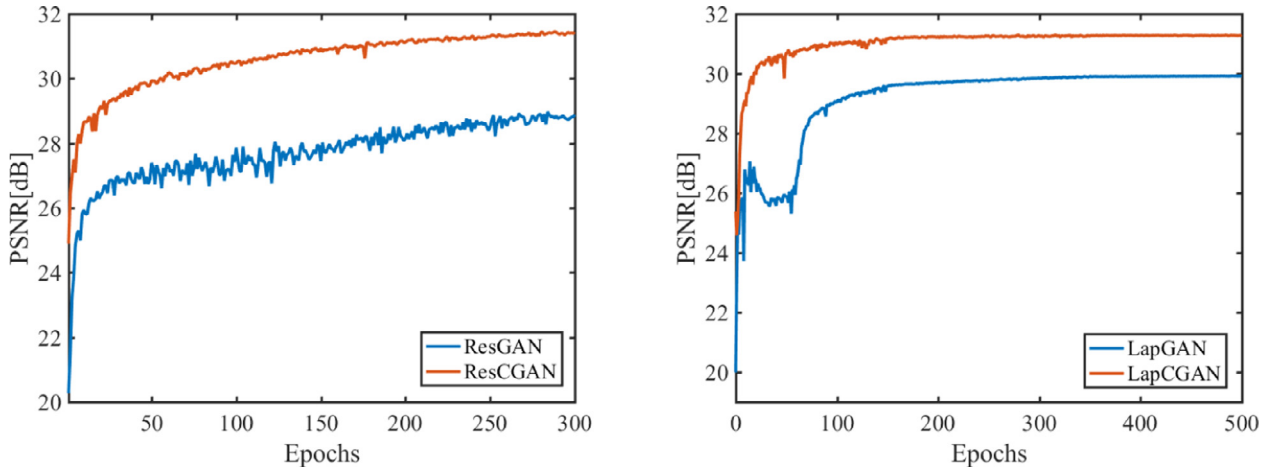

Fig. 7. Left: PSNR convergence curves of CGAN and GAN using single upsamping network. Right: PSNR convergence curves of CGAN and GAN using progressively upsamping network in LapSRN [26] on Set5.

ReLU activation function to give each negative value a non-zero slope.

We use 17,125 images in the VOC2012 dataset to train the network without any data augmentation. Each HR image is cropped into an 88×88 image patch for training. According to the training scheme of existing method [22], we generate an LR image training patch using bicubic downsampling of an HR image by a scale-factor $= 4$. The Adam algorithm with RMSProp is adopted to learn the network parameters with a momentum factor of 0.9 and a weight decay rate of 1e−4. The batch size is set to 16. The learning rates of all layers are initialized to 1e−5. The trade-off parameters in (1) are set to $\alpha = 10^{-3}$, $\beta = 1$, $\gamma = 6 \times 10^{-3}$. Except for the parameter γ, we set the parameters α and β with the same values as SRGAN [13]. As shown in Fig. 5, a larger γ, even with a poor PSNR, can lead to a perceptually clearer image. We balance the objective evaluation result (PSNR) with subjective visual effect (perceptually pleasing) to set $\gamma = 6 \times 10^{-3}$.

3.2. Training curves

Fig. 6 shows the training curves of discriminator loss and generator loss for GAN and CGAN. The discriminator loss of CGAN is fast monotonic-decreasing to approach to zero while that of GAN fluctuates much at the beginning, and then tends to be stable. Meanwhile, the generator losses are also monotonic-increasing to be stable and the stable loss value of the CGAN generator is smaller than that of GAN. Overall, CGAN enables to generate much more stable training than GAN, verifying the effectiveness of using ground-truth as a conditional variable in CGAN that ensures more stabler training than GAN.

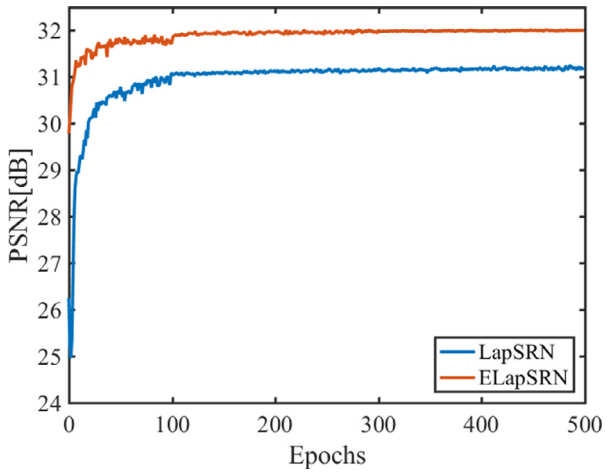

Fig. 8. PSNR convergence curves of LapSRN and ELapSRN on Set5. ELapSRN replaces the LapN in LapSRN by our ELapN.

3.3. Ablative study

The proposed SR model includes two core components: CGAN and ELapN. To verify the effectiveness of each component, we design three different variants. The first two are only for CGAN including a single upsampling network and a progressive upsampling network. The single upsampling network is based on SRResNet [13] while the progressive upsampling network is based on LapSRN [26]. Furthermore, the third one is for ELapN, and based

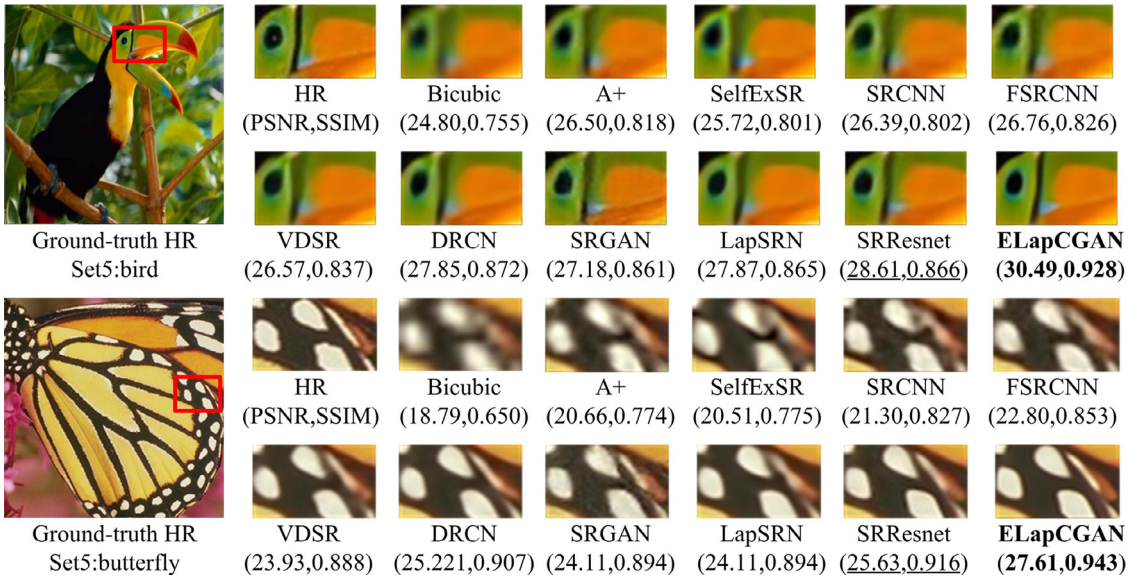

Fig. 9. Visual comparison on the Set5 dataset. **Bold**: best; underline: second best.

Fig. 10. Visual comparison on the Set14 dataset. **Bold**: best; underline: second best.

on LapSRN [26], we compare the proposed ELapN with the LapN in LapSRN.

3.3.1. Comparisons of CGAN with GAN

Single upsampling network SRResnet [13] uses a single upsampling network with promising performance, and we integrate CGAN and GAN into its framework (termed as ResCGAN and ResGAN respectively) for comparisons. The left figure of Fig. 7 shows the PSNR convergence curves of ResCGAN and ResGAN on Set5. We can observe that ResCGAN not only fluctuates smoother than ResGAN, but also achieves much higher PSNR values than ResGAN during the network training. This confirms the effectiveness of our CGAN in the single upsampling network model.

Progressively upsampling network LapSRN [26] proposes a progressively upsampling network for SISR with better results than single upsampling network. For fair comparisons, we utilize the same architecture of LapSRN to train the generative model of CGAN and GAN, and the results are shown by the right figure of Fig. 7. We can observe that the LapSRN with GAN (LapGAN) fluctuates much at the beginning of the training, and the LapSRN with CGAN (LapCGAN) can converge faster and achieve better performance that LapGAN.

3.3.2. Comparisons of proposed ELapN with LapN

To verify the effect of the proposed ELapN, we replace the LapN in LapSRN [26] by the proposed ELapN, which is termed as ELapSRN and then compare LapSRN with ELapSRN in terms of PSNR. Fig. 8 shows the comparison results of LapSRN and ELapSRN on Set5. We can observe that ELapSRN is superior to LapSRN with higher PSNR values during training.

3.4. Comparisons with state-of-the-arts

To further demonstrate advantages of the proposed network, we compare our ELapCGAN with 12 state-of-the-art SR algorithms in terms of PSNR and SSIM [27], including the dictionary-based methods (A+ [28] and RFL [29]), the method based on self-similarity

Fig. 11. Visual comparison on the B100 dataset. **Bold**: best; underline: second best.

Table 1
Combination results of the proposed ELapCGAN and the state-of-the-arts on Set5, Set14 and B100 in terms of PSNR [dB] and SSIM. **Bold**: best; underline: second best.

Algorithm	Set5 PSNR/SSIM	Set14 PSNR/SSIM	B100 PSNR/SSIM
Bicubic	28.43/0.811	26.01/0.704	25.97/0.670
A+ [28]	30.32/0.860	27.34/0.751	26.83/0.711
RFL [29]	30.17/0.855	27.24/0.747	26.76/0.708
SelfExSR [30]	30.34/0.862	27.41/0.753	26.84/0.713
SRCNN [10]	30.50/0.863	27.52/0.753	26.91/0.712
FSRCNN [31]	30.72/0.866	27.61/0.755	26.98/0.715
SCN [32]	30.41/0.863	27.39/0.751	26.88/0.711
VDSR [11]	31.35/0.883	28.02/0.768	27.29/0.726
DRCN [33]	31.54/0.884	28.03/0.768	27.24/0.725
DRRN [34]	31.68/0.888	28.21/0.772	27.38/0.728
SRGAN [13]	29.40/0.847	26.02/0.740	25.16/0.669
LapSRN [26]	31.54/0.885	28.19/0.772	27.32/0.727
SRResnet [13]	32.05/0.891	28.53/0.780	27.57/0.735
ELapCGAN	**32.16/0.897**	**28.61/0.796**	**27.69/0.741**

(SelfExSR [30]) and the CNN-based methods (SRGAN [13], SRCNN [10], FSRCNN [31], SCN [32], VDSR [11], DRCN [33], DRRN [34], LapSRN [26], SRResnet [13]). We conduct extensive experiments on three benchmark datasets including Set5, Set14 [35] and B100 [36], which consist of a variety of natural scenes. Table 1 lists the compared results of all evaluated algorithms. Among them, the proposed ELapCGAN achieves the best performance in terms of both PSNR and SSIM on all three datasets, outperforming the baseline SRGAN and LapSRN by a large margin.

To further show the superiority of the proposed ELapCGAN over the state-of-the-arts, Figs. 9–11 show some visual comparisons on Set5, Set14 and B100. For fair comparisons, we use the SR results released by the authors. Compared to the state-of-the-arts, our ELapCGAN presents more visual-pleasing results with sharper edges and richer textures.

4. Conclusion

In this paper, we present a conditional generative adversarial learning framework with an ELapN for accurate SISR. Specifically, we take the ground-truth HR image as the conditional variable to guide to learn the CGAN model, and propose the ELapN that combines low-and high-level features for residual image learning with better generalization. Furthermore, we progressively predict the HR image in a coarse-to-fine manner with deep supervision from multi-level loss functions. Extensive experiments on three benchmark datasets demonstrate that the proposed method achieves better reconstruction results in terms of both quantitative and qualitative metrics with a large upsampling factor $4\times$.

Declarations of interest

None.

Acknowledgments

This work is supported in part by the National Natural Science Foundation of China under Grant nos. 61872189, 61876088, in part by the Natural Science Foundation of Jiangsu Province under Grant no. BK20170040, in part by Six talent peaks project in Jiangsu Province under Grant nos. XYDXX-015, XYDXX-045.

References

[1] Q. Yang, R. Yang, J. Davis, D. Nistér, Spatial-depth super resolution for range images, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2007, pp. 1–8.
[2] W.W.W. Zou, P.C. Yuen, Very low resolution face recognition problem, IEEE Trans. Image Process. 21 (1) (2012) 327.
[3] K. Nasrollahi, T.B. Moeslund, Super-resolution: a comprehensive survey, Mach. Vis. Appl. 25 (6) (2014) 1423–1468.
[4] H. Song, Q. Liu, G. Wang, R. Hang, B. Huang, Spatiotemporal satellite image fusion using deep convolutional neural networks, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 11 (3) (2018) 821–829.
[5] N. Kien, F. Clinton, S. Sridhan, T. Massimo, N. Mark, Super-resolution for biometrics: a comprehensive survey, Pattern Recognit. 78 (2018) 23–42.
[6] Z. Yan, C. Xiang, X. Yi, Z. Xianyi, Y. Jin, Joint residual pyramid for joint image super-resolution, J. Vis. Commun. Image Represent. 58 (2019) 53–62.
[7] Y. Jun, H. Chaoqun, R. Yong, T. Dacheng, Multitask autoencoder model for recovering human poses, IEEE Trans. Ind. Electron. 65 (6) (2018) 5060–5068.
[8] Y. Zhou, Y. Jun, X. Chenchao, F. Jianping, T. Dacheng, Beyond bilinear: generalized multimodal factorized high-order pooling for visual question answering, IEEE Trans. Neural Netw. Learn. Syst. 27 (2018) 1–13.
[9] Z. Jian, Y. Jun, T. Dacheng, Local deep-feature alignment for unsupervised dimension reduction, IEEE Trans. Image Process. 27 (5) (2018) 2420–2432.
[10] C. Dong, C.C. Loy, K. He, X. Tang, Learning a deep convolutional network for image super-resolution, in: Proceedings of the European Conference on Computer Vision, Springer, 2014, pp. 184–199.
[11] J. Kim, J. Kwon Lee, K. Mu Lee, Accurate image super-resolution using very deep convolutional networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1646–1654.
[12] B. Lim, S. Son, H. Kim, S. Nah, K.M. Lee, Enhanced deep residual networks for single image super-resolution, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 1, 2017, p. 4.

[13] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A.P. Aitken, A. Tejani, J. Totz, Z. Wang, 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, 2, 2017, p. 4.
[14] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, in: Proceedings of the Advances in neural information processing systems, 2014, pp. 2672–2680.
[15] M.S. Sajjadi, B. Schölkopf, M. Hirsch, Enhancenet: Single image super-resolution through automated texture synthesis, in: Proceedings of the IEEE International Conference on Computer Vision, IEEE, 2017, pp. 4501–4510.
[16] E.L. Denton, S. Chintala, R. Fergus, et al., Deep generative image models using a Laplacian pyramid of adversarial networks, in: Proceedings of the Advances in Neural Information Processing Systems, 2015, pp. 1486–1494.
[17] J. Johnson, A. Alahi, L. Fei-Fei, Perceptual losses for real-time style transfer and super-resolution, in: Proceedings of the European Conference on Computer Vision, Springer, 2016, pp. 694–711.
[18] Z. Masoumeh, C. M Emre, Y. Jie, Diverse adversarial network for image super-resolution, Signal Process. Image Commun. (2019).
[19] M. Mirza, S. Osindero, Conditional generative adversarial nets, arXiv:1411.1784. (2014).
[20] S. Reed, Z. Akata, X. Yan, L. Logeswaran, B. Schiele, H. Lee, Generative adversarial text to image synthesis, in: Proceedings of the International Conference on Machine Learning, 2016.
[21] P. Isola, J.Y. Zhu, T. Zhou, A.A. Efros, Image-to-image translation with conditional adversarial networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 5967–5976.
[22] W.-S. Lai, J.-B. Huang, N. Ahuja, M.-H. Yang, Fast and accurate image super-resolution with deep Laplacian pyramid networks, IEEE Trans. Pattern Anal. Mach. Intell. (2018), doi:10.1109/TPAMI.2018.2865304.
[23] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, 2014.
[24] W. Shi, J. Caballero, F. Huszár, J. Totz, A.P. Aitken, R. Bishop, D. Rueckert, Z. Wang, Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1874–1883.
[25] A. Bruhn, J. Weickert, C. Schnörr, Lucas/kanade meets horn/schunck: combining local and global optic flow methods, Int. J. Comput. Vis. 61 (3) (2005) 211–231.
[26] W.-S. Lai, J.-B. Huang, N. Ahuja, M.-H. Yang, Deep Laplacian pyramid networks for fast and accurate super-resolution, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2, 2017, p. 5.
[27] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Process. 13 (4) (2004) 600–612.
[28] R. Timofte, V. De Smet, L. Van Gool, A+: adjusted anchored neighborhood regression for fast super-resolution, in: Proceedings of the Asian Conference on Computer Vision, Springer, 2014, pp. 111–126.
[29] S. Schulter, C. Leistner, H. Bischof, Fast and accurate image upscaling with super-resolution forests, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3791–3799.
[30] J.B. Huang, A. Singh, N. Ahuja, Single image super-resolution from transformed self-exemplars, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 5197–5206.
[31] C. Dong, C.C. Loy, X. Tang, Accelerating the super-resolution convolutional neural network, in: Proceedings of the European Conference on Computer Vision, Springer, 2016, pp. 391–407.
[32] Z. Wang, D. Liu, J. Yang, W. Han, T. Huang, Deep networks for image super-resolution with sparse prior, in: Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 370–378.
[33] J. Kim, J. Kwon Lee, K. Mu Lee, Deeply-recursive convolutional network for image super-resolution, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1637–1645.
[34] Y. Tai, J. Yang, X. Liu, Image super-resolution via deep recursive residual network, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1, 2017, p. 5.
[35] R. Zeyde, M. Elad, M. Protter, On single image scale-up using sparse-representations, in: Proceedings of the International Conference on Curves and Surfaces, Springer, 2010, pp. 711–730.
[36] P. Arbelaez, M. Maire, C. Fowlkes, J. Malik, Contour detection and hierarchical image segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 33 (5) (2011) 898–916.

Xiaolu Zhang has been a master student in the School of Automation in Nanjing University of Information Science and Technology since Sep. 2017. She received her bachelor degree in Measurement and Control Technology and Instrumentation from Nanjing University of Information Science and Technology in Jun. 2017. Her research interest includes Image super-resolution reconstruction algorithms.

Huihui Song is a Professor with the Jiangsu Key Laboratory of Big Data Analysis Technology, Nanjing University of Information Science and Technology, Nanjing, China. She received her B.S. degree in technology and science of electronic information from Ocean University of China in 2008, master degree in communication and information system from University of Science and Technology of China (USTC) in 2011, and Ph.D. degree in geography and resource management from the Chinese University of Hong Kong in 2014. Her research interests include remote sensing image processing and image fusion. He received the B.S. degree in Technology and Science of Electronic Information from Ocean University of China (OUC) in 2006, the M.S. degree in Signal and Information Processing from the University of Science and Technology of China (USTC) in 2009 and Ph.D. degree from the Department of Computing in the Hong Kong Polytechnic University in 2013. From Aug. 2009 to Aug. 2010, he worked as a Research Assistant in the Department of Computing, The Hong Kong Polytechnic University. His research interests include image segmentation, level sets, and visual tracking.

Kaihua Zhang is a professor with the Jiangsu Key Laboratory of Big Data Analysis Technology, Nanjing University of Information Science & Technology, Nanjing, China. He received the B.S. degree in Technology and Science of Electronic Information from Ocean University of China (OUC) in 2006, the M.S. degree in Signal and Information Processing from the University of Science and Technology of China (USTC) in 2009 and Ph.D. degree from the Department of Computing in the Hong Kong Polytechnic University in 2013. From Aug. 2009 to Aug. 2010, he worked as a Research Assistant in the Department of Computing, The Hong Kong Polytechnic University. His research interests include image segmentation, level sets, and visual tracking.

Jiaojiao Qiao has been a master student in the School of Automation in Nanjing University of Information Science and Technology since Sep. 2017. She received her bachelor degree in Electrical Engineering and Automation from Nanjing Institute of Technology in Jun. 2017. Her research interest includes digital image processing and Image super resolution

Qingshan Liu is a Professor with the School of Information and Control, Nanjing University of Information Science and Technology, Nanjing, China. He received the Ph.D. degree from the National Laboratory of Pattern Recognition, Chinese Academic of Science, Beijing, China, in 2003, and the M.S. degree from the Department of Auto Control, Southeast University, Nanjing, in 2000. He was an Assistant Research Professor with the Department of Computer Science, Computational Biomedicine Imaging and Modeling Center, Rutgers, The State University of New Jersey, New Brunswick, NJ, USA, from 2010 to 2011. Before he joined Rutgers University, he was an Associate Professor with the National Laboratory of Pattern Recognition, Chinese Academic of Science, and an Associate Researcher with the Multimedia Laboratory, Chinese University of Hong Kong, Hong Kong, from 2004 and 2005. He was a recipient of the President Scholarship of the Chinese Academy of Sciences in 2003. His current research interests are image and vision analysis, including face image analysis, graph and hypergraph-based image and video understanding, medical image analysis, and event-based video analysis.