Recurrent Embedded Hourglass Network for Single Image Super-Resolution

NANYAN LIU, GUANGPU GAO, AND XINYU XU
Department of Computer Science and Technology, Xi’an University of Science and Technology, Xi’an 710054, China
Corresponding author: Guangpu Gao (1065574172@qq.com)

ABSTRACT The single image super-resolution has become an important topic of discussion due to the demand for high-quality digital images in the field of visual artificial intelligence. The deep learning-based approach has achieved great success because of its excellent ability to control complex features. Using simply extending or deepening the network, performance can be improved slightly, but not significantly. There are still blurry edges and rough texture details. In this paper, we propose a Recurrent Embedded Hourglass Network (SRRHN) for super-resolution reconstruction. We use the hourglass network to combine deep and shallow features and embed Gated Recurrent Unit (GRU) in each layer of the hourglass network to improve the long-range correlations. Finally, sub-pixel convolution is adopted to avoid image distortion during up-sampling. Extensive experiments on several standard benchmarks show that our proposed method achieves better performance compared with state-of-the-art methods.

INDEX TERMS Super resolution, convolutional neural network, gated recurrent unit, hourglass network.

I. INTRODUCTION
The single image super-resolution (SISR) is a typical ill-posed inverse problem [1]. It refers to the recovery of the high-resolution image (HR) by a single frame low-resolution image (LR). Due to the interference of various factors such as poor equipment and environment, the final image quality is poor. However, the low-quality image obviously cannot meet the requirements in video surveillance, satellite remote sensing, medical image, and other areas with strict requirements on image quality [2]. Therefore, a SISR algorithm with a clear reconstruction effect is very necessary.

SISR has made great progress in recent years. The mainstream method of SISR is based on deep learning. Dong [3], [4] et al. proposed the SRCNN [3] method for the first time. They designed three layers Convolutional Neural Network (CNN) to learn the mapping between LR and HR, which achieved good performance and proved the applicability of deep learning in the SISR problem.

Then Dong et al. proposed a fast version based on the existing architecture to learn the LR to RH mapping through the deconvolution layer, which is called a Fast Super-Resolution Convolutional Neural Network (FSRCNN) [5]. It indicates that the network can learn the amplification filter directly and improve the accuracy and speed of training. On this basis, Shi [6] et al. designed an efficient sub-pixel convolutional layer which is only used at the end of the network. The method can learn an array of upscaling filters to up-scale the final LR feature maps into the HR output and avoid resource distortion in bottom up-sampling effectively.

Inspired by VGG-NET [7], Kim et al. designed a highly deep convolutional network named VDSR [8], and they used 20 layers of CNN and residual learning. Their experiment showed that the model without the BN layer could save about 40% of memory usage and maintain performance in the training process [10].

However, deepening networks aggravate the vanishing-gradient problem, and the extracted features will become more abstract. How to integrate with the local information captured in the shallow layer becomes a big challenge to improve network performance. Also, the traditional neural network model is fully connected between the input layer, hidden layer, and output layer, and the network nodes between each layer are disconnected. After receiving the input information, there is no memory information. Therefore, the information is incomplete during training, and the
reconstruction of the image edge effect and texture detail is unsatisfactory. To solve the problems mentioned above, we put forward a novel Recurrent Embedded Hourglass Network (SRRHN) for Super-Resolution architecture, which solves the aforementioned problems by concatenating shallow and deep feature maps via hourglass network. At the same time, embedded Gated Recurrent Unit [11], [12] to solve the problem of a convolution neural network is difficult to improve the long-range correlations. At the end of the architecture, we used sub-pixel operation [6] to execute a mapping module to generate HR images. Extensive experiments on several standard benchmarks show that our proposed method achieves better performance compared with state-of-the-art methods Peak-signal-to-noise-ratio(PSNR) value and Structural similarity (SSIM) value of this algorithm are improved by 0.5dB and 0.03 respectively. The edge effect and texture detail of the image are improved either.

The contributions of this paper are summarized as shown below:

We propose a novel SRRHN model to combine shallow and deep features to collect a more adequate feature atlas and to improve the long-range correlations for SR performance enhancement.

We expand the network by using an hourglass network embedded in GRU to concatenate shallow and deep feature maps and improve the long-range correlations.

Extensive experiments present that our approach achieves state-of-the-art performance on multiple benchmarks, demonstrating the effectiveness of our network.

II. RELATED WORK

Hourglass network is a kind of CNN architecture, which has been widely used in semantic segmentation. It is originally proposed by Olaf Ronneberger [13], and its architecture is mainly divided into two parts, called encoder and decoder. The encoder adopts several sample convolutional layers continuously to obtain different levels of image features. The decoder performs multi-layer deconvolution in the final resource map of the encoder to combine features of different levels in the sampling process of each level. It connects deep features to shallow features and recovers the size of the original input image. The hourglass network - also called U-net because of the u-shaped - is generally used to accomplish the task of end-to-end semantic segmentation and.

Gated Recurrent Unit network is another network architecture of CNN. Unlike other traditional neural networks, the nodes between the hidden layers of recurrent neural networks are interconnected. The input for each moment is not only the input for the current time but also contains the output from the hidden layer of the previous moment. It enables recurrent neural networks to process richer historical information and apply it to current learning. This network architecture has a memory function, which can well improve the long-range correlations and learn more image features [11], [14].

III. PROPOSED METHOD

Based on the network architectures mentioned in chapter 2 which achieve good results in their respective fields, we propose a Recurrent Embedded Hourglass Network (SRRHN) for SR. We build a network architecture with hourglass shape embedded GRU to extract features, as shown in Fig. 1. The SRRHN consists of four key components: 1) reduce convolution block (ReConv), 2) deconvolution block (DeConv), 3) convolutional GRU (ConvGRU), and 4) up-sampling layer. ReConv is on the left side of the hourglass network. The result of ReConv is that the perception field is continuously enhanced, the input size of each level is gradually reduced, the number of output feature maps is gradually increased, and features are continuously abstracted. DeConv is on the right side of the hourglass network, which can restore the original size of the image by deconvolution operation. In addition, ConvGRU connects the deep and shallow features of the same level with strong spatial deconvolution operation. To introduce memory information into the network, and obtain complete feature information. Finally, the up-sampling layer adopts sub-pixel convolution to avoid image distortion in image reconstruction.

A. RECONV BLOCK

As shown in Fig. 1, the yellow block is a reduced convolution block, following a typical convolutional network architecture. However, different from the traditional hourglass network, which adopts the pooling operation to shrink the features and size, this paper adopts the convolution with a stride size of 2, which can reduce the loss of image features caused by pooling operation. The specific architecture is shown in Fig. 2(a). It mainly consists of the convolution of a $3 \times 3$ filter with a step size of 2 and a $1 \times 1$ filter with a convolution step size of 1. Then, each convolution is composed of a Rectified Linear Unit (ReLU) [15] for skip connection operation. Contract in each step, we will be the channel characteristics, doubling the number of image size in half to achieve different levels of image characteristics, the characteristics of double channel number to obtain a variety of image features, The convolution result of each contraction is taken as the input of the next contraction and the same layer ConvGRU. Because EDSR has achieved good results after deleting the BN layer [9], this method does not adopt BN operation.

B. DECONV BLOCK

The blue blocks are shown in Fig. 1 represents the deconvolution block and its specific structure is shown in Fig 2(b). In order to merge multiple feature maps and to acquire more detailed information, it utilizes the stack result of the two feature maps computed by ConvGRU convolution and deconvolution as the input. However, it can result in a doubling of the number of channels for a feature. To reduce the network parameters and to accelerate the network training, the paper applies a $1 \times 1$ filter to optimize the convolution characteristics of the channel to minimize the channel quantity.
FIGURE 1. Architecture of the proposed Recurrent Embedded Hourglass Network for super-resolution (SRRHN). X represents the LR image. Y represents the reconstructed HR image. ReConv: $k \times k \times c$ represents s feature maps of size $k \times k$.

FIGURE 2. The detailed structure of the different blocks. (a) reduce convolution block (ReConv), (b) deconvolution block (DeConv), and (c) convolutional GRU block (ConvGRU). ReConv($k \times k \times c$, $s = m$) and DeConv($k \times k \times c$, $s = m$) represents a convolution layer or a deconvolution layer that has $k \times k$ filters, $c$ output maps and $m$ strides. ConvGRU(layers = 3) represents a GRU network with 3 layers.

Therefore, a convolutional layer of $1 \times 1$ is used to reduce the number of channels, followed by a ReLU and residual operation.

C. CONVGRU BLOCK

The red block in Fig. 1 represents the ConvGRU block. It solves the drawback of a fully connected network in processing time series. Also, a prominent advantage of the recurrent neural network is the parameter sharing strategy. Such a network architecture saves memory footprint and hardware consumption. Since maintaining a long-term memory at the same level of spatial location allows us to find the strong characteristics of the image, a three-layer network unidirectional GRU is adopted to support it to input changes over scaling characteristics of the convolution layer channel. The first layer is 64, the second layer is 128, and the third layer is 256. The result of the convolution is followed by a ReLU and residual operation. The specific architecture is shown in Fig. 2(c).
D. UP-SAMPLING LAYER

Reconstruction refers to the up-sampling layer of the architecture in this paper, as shown in Fig. 1 Up-Sampling block. We use a sub-pixel layer at the end of the architecture for up-sampling. The specific structure is shown in Fig. 3.

FIGURE 3. Sub-pixel convolution reconstruction layer structure.

With different upscale factor $r$, in the last deconvolution output feature graph, a core with a size of $3 \times 3 \times r^2$ is used for convolution operation, and $r^2$ represents the number of feature graphs. Then, the $r^2$ feature graphs generated by convolution are rearranged and combined to reconstruct the HR image with a magnification of $r$. This operator can be described in the following way:

$$I_{HR} = P_t (wF_{last} + b)$$  \hspace{1cm} (1)

where $I_{HR}$ is a tensor in the HR space, $F_{last}$ is the feature map output by the last mapping unit, $w$ and $b$ are learnable network weights and bias. $P_t$ is a periodic transformation operator, which rearranges a tensor with a shape of $h \times w \times c \times r^2$ into a tensor with a shape of $hr \times wr \times c$, to realize the up-sampling of the image.

IV. EXPERIMENTS

A. DATASETS

We used the DIV2K [16], [17] dataset for training and testing the model. DIV2K is a relatively new high-quality image data set, with all data at 2K pixels. It contains 800 training images and 100 validation images. In this case, the validation data set is used for the test. To compare our network with other methods, three benchmark data sets are applied in the evaluation: Set5 [18], Set14 [19], and BSD300 [20].

B. IMPLEMENTATION DETAILS

The DIV2K training set is used for training. More specifically, we used 960 $\times$ 960 RGB input patches from LR images with the corresponding HR patches cropped from 800 images of DIV2K for training. To expand the training data, we augmented our training data by using random horizontal or vertical flips and rotations of 90, 180, and 270 degrees.

We used the Adam [21] optimizer $\beta_1 = 0.9$ and $\beta_2 = 0.999$ to train the model. We set the minimum batch to 64, and the learning rate is initialized to $10^{-3}$ and halved every 50 epochs. After 100 epochs, if the loss value does not continue to change, the training stops. We use L2 loss functions, and we implement the proposed system on the Pytorch framework. The training took 12 hours on Nvidia TITAN XP to get the best model on the training set. We evaluated these methods in Python by Peaking signal to noise ratio (PSNR) and Structure similarity (SSIM) as performance indicators. PSNR reflects the error between the corresponding pixel points of two images. The higher the value, the less the output image distortion, and the better the image reconstruction quality. SSIM is an evaluation index representing the similarity of two images. The closer the value is to 1, the closer the output image is to the original high-resolution image, the better the reconstruction effect.

C. EXPERIMENTS RESULT AND ANALYSIS

We evaluated our proposed network using DIV2K validation data sets and compared it qualitatively and quantitatively with classical methods including SRCNN [3], FSRCNN [5], VDSR [8], EDSR [9], RCAN [22], and HDRN [23] were compared with the method in this paper. Meanwhile, to verify the performance of ConvGRU in the feature extraction stage, we replaced ConvGRU in the network with convolution and called SRUCNN. To verify the hourglass network performance, we replaced the ReConv and DeConv with convolution and called SRGRU.

TABLE 1. Performance comparison between architectures on DIV2K validation set expressed as PSNR(dB)/SSIM.

| Scale | x2     | x3     | x4     |
|-------|--------|--------|--------|
| Bicubic | 31.5923/0.938 | 28.1312/0.894 | 26.9356/0.848 |
| SRCNN   | 33.0467/0.958 | 29.5540/0.917 | 27.9076/0.867 |
| FSRCNN  | 32.0926/0.932 | 29.0456/0.905 | 28.2534/0.871 |
| VDSR    | 33.3264/0.940 | 30.4308/0.923 | 29.0832/0.901 |
| EDSR    | 34.4537/0.947 | 31.3730/0.936 | 29.6832/0.902 |
| RCAN    | 34.7193/0.948 | 31.4013/0.940 | 29.6981/0.903 |
| HDRN    | 34.7224/0.951 | 31.3043/0.940 | 29.7313/0.908 |
| SRUCNN  | 34.7264/0.950 | 31.3224/0.943 | 29.7213/0.903 |
| SRGRU   | 34.9264/0.963 | 32.0756/0.941 | 29.6832/0.907 |
| SRRHN   | 35.0824/0.971 | 32.5393/0.947 | 29.9439/0.912 |

With an upscale factor of 2 or 3 or 4, via comparing the average values of PSNR and SSIM in the DIV2k test set after being processed by different methods, it can be seen that compared with the traditional Bicubic algorithm, the SRRHN method greatly improved the value of PSNR and SSIM, and it was also superior to other mainstream algorithms based on deep learning.

SRGRU was superior to VDSR [8] and EDSR [9], suggesting that an embedded ConvGRU network could effec-
TABLE 2. Performance comparison between architectures on set5, set14, bsd300 dataset expressed as PSNR(dB)/SSIM.

| Datasets | Scale | Bicubic | SRCNN | FSRCNN | VDSR | EDSR | RCAN | HDRN | Our   |
|----------|-------|---------|-------|--------|------|------|------|------|-------|
| Set5     | ×2    | 33.67/0.929 | 35.46/0.955 | 34.85/0.956 | 36.64/0.962 | 36.66/0.961 | 36.98/0.953 | 37.04/0.956 | 37.10/0.969 |
|          | ×3    | 30.13/0.850 | 32.64/0.906 | 31.31/0.923 | 33.25/0.929 | 33.93/0.929 | 34.69/0.930 | 34.72/0.930 | 34.94/0.934 |
|          | ×4    | 28.69/0.799 | 30.58/0.846 | 30.46/0.889 | 32.58/0.895 | 32.56/0.893 | 32.60/0.897 | 32.63/0.897 | 32.94/0.901 |
| Set14    | ×2    | 30.24/0.883 | 31.28/0.903 | 33.37/0.918 | 34.16/0.916 | 35.03/0.922 | 35.51/0.956 | 35.44/0.943 | 35.57/0.962 |
|          | ×3    | 28.69/0.735 | 29.58/0.826 | 29.46/0.849 | 30.58/0.833 | 29.73/0.850 | 30.79/0.829 | 31.59/0.836 | 31.63/0.841 |
|          | ×4    | 26.69/0.706 | 28.08/0.786 | 28.46/0.779 | 29.58/0.785 | 29.80/0.785 | 29.64/0.783 | 30.06/0.788 | 30.24/0.791 |
| BSD300   | ×2    | 30.47/0.848 | 31.85/0.867 | 30.46/0.930 | 31.84/0.928 | 32.02/0.931 | 31.98/0.931 | 32.14/0.931 | 32.56/0.935 |
|          | ×3    | 28.78/0.803 | 27.04/0.758 | 27.67/0.836 | 28.08/0.876 | 28.99/0.886 | 29.36/0.890 | 29.38/0.883 | 29.66/0.894 |
|          | ×4    | 24.18/0.655 | 25.24/0.729 | 25.91/0.754 | 26.37/0.803 | 26.57/0.797 | 26.63/0.807 | 26.59/0.809 | 26.96/0.812 |

FIGURE 4. Results on the DIV2K validation dataset via different methods with an upscale factor of 2.

FIGURE 5. Results on the DIV2K training dataset via different methods with an upscale factor of 3.
tively improve SR capability. One reasonable explanation was that ConvGRU combined the information of the input layer and the hidden layer to learn more comprehensive information than the convolution. The SRUCNN results...
were also superior to the VDSR [8] and EDSR [9], indicating that the use of the hourglass network enhanced the ability to combine the shallow information with the deep information, enabling the network to learn different levels of image features. Therefore, the SRRHN method, which combines the hourglass network with the ConvGRU network, achieves good experimental results. In the DIV2K test set, the PSNR value was improved by about 2.3dB on average compared with the SRCCN [3], and the SSIM was improved by about 0.03. Compared with VDSR [8], EDSR [9], RCAN [22], HDRN [23], PSNR was improved by 1.5dB, 0.70dB, 0.58dB, 0.60dB and SSIM by 0.020, 0.015, 0.012, 0.010.

To show the effect of this algorithm compared with other algorithms more intuitively, Fig. 4 and Fig. 5 shows different reconstruction results.

To show the effect of SRRHN compared with other algorithms more intuitively, Fig. 4 and Fig. 5 respectively showed the processing results of two images in DIV2k data set by different algorithms. To highlight the detail texture, we enlarged and compared the local area of the image. SRRHN produces clearer and clearer details, the letters in Fig. 4 are clearer, the ghosting is significantly reduced, and the reconstruction is closer to the original. The proposed algorithm in Fig 5 also clearly restores the spire part, which is more effective than other algorithms.

Subjective judgments are shown in Fig. 6 and Fig. 7. With an upscale factor of 2, Fig. 6 shows the super-resolution reconstruction result of the butterfly image in the Set5 test set and the effect can be identified from the figure. The method of Bicubic has the fuzziest processing effect, SRCNN [3] and FSRCNN [5] algorithm have some shortcomings in the detail area of butterfly wings. The SRGRU and SRUCNN algorithms proposed in this paper have similar effects and the processing results are better than VDSR [8], and while the SRRHN method has better effects in restoring ghosting and edge.

With an upscale factor of 3, Fig. 7 shows the super-resolution reconstruction result of building an image in the BSD300 test set. It can be seen from the results in the texture of the grille wooden bridge, SRRHN processing results significantly better than the traditional interpolation algorithm, compared with other algorithms based on deep learning at the same time, this algorithm processing images of the stripe edges appear sharper, details more clear.

V. CONCLUSION
We propose a super-resolution method using Recurrent Embedded Hourglass Network, which can extract the complementary features and improve the long-range correlations via the hourglass network embedded in the GRU, and reconstruct the SR image by using the sub-pixel layer without distorting the information. The quantitative and qualitative experimental results obtained using benchmark datasets demonstrate the superior performance of our method.

REFERENCES
[1] M. Irani and S. Peleg, “Improving resolution by image registration,” CVGIP, Graph. Models Image Process., vol. 53, no. 3, pp. 231–239, May 1991.
[2] M. W. Thornton, P. M. Atkinson, and D. A. Holland, “Sub-pixel mapping of rural land cover objects from fine spatial resolution satellite sensor imagery using super-resolution pixel-swapping,” Int. J. Remote Sens., vol. 27, no. 3, pp. 473–491, Feb. 2006.
[3] C. Dong, C. C. Loy, K. He, and X. Tang, “Image super-resolution using deep convolutional networks,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 38, no. 2, pp. 295–307, Feb. 2016.
[4] C. Dong, C. C. Loy, K. He, and X. Tang, “Learning a deep convolutional network for image super-resolution,” in Proc. Eur. Conf. Comput. Vis., Sep. 2014, pp. 184–199.
[5] C. Dong, C. C. Loy, and X. Tang, “Accelerating the super-resolution convolutional neural network,” in Proc. Eur. Conf. Comput. Vis. (ECCV), Amsterdam, The Netherlands, Oct. 2016, pp. 391–407.
[6] W. Shi, J. Caballero, F. Huszar, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, “Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network,” in Proc. IEEE Comput. Vis. Pattern Recognit. (CVPR), Las Vegas, NV, USA, Jun. 2016, pp. 1874–1883.
[7] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large scale image recognition,” in Proc. Int. Conf. Learn. Represent., San Diego, CA, USA, May 2015, pp. 1–14.
[8] J. Kim, J. K. Lee, and K. M. Lee, “Accurate image super-resolution using very deep convolutional networks,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Las Vegas, NV, USA, Jun. 2016, pp. 1646–1654.
[9] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, “Enhanced deep residual networks for single image super-resolution,” in Proc. IEEE Comput. Vis. Pattern Recognit. Workshops (CVPRW), Honolulu, HI, USA, Jul. 2017, pp. 1132–1140.
[10] S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” in Proc. Int. Conf. Mach. Learn., 2015, pp. 448–456.
[11] N. Ballas, L. Yao, C. Pal, and A. Courville, “Delving deeper into convolutional networks for learning video representations,” in Proc. Int. Conf. Learn. Representations, 2015, p. 1.
[12] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, “Empirical evaluation of gated recurrent neural networks on sequence modeling,” 2014, arXiv:1412.3555. [Online]. Available: http://arxiv.org/abs/1412.3555.
[13] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional networks for biomedical image segmentation,” in Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent., Munich, Germany, Nov. 2015, pp. 234–241.
[14] P. Pinheiro and R. Collobert, “Recurrent convolutional neural networks for scene labeling,” J. Mach. Learn. Res., vol. 1, no. 32, pp. 82–90, 2014.
[15] X. Glorot, A. Bordes, and Y. Bengio, “Deep sparse rectifier neural networks,” in Proc. 14th Int. Conf. Artif. Intell. Statist. (AISTATS), Fort Lauderdale, FL, USA, Apr. 2011, pp. 315–323.
[16] E. Agustsson and R. Timofte, “NTIRE 2017 challenge on single image super-resolution: Dataset and study,” in Proc. IEEE Comput. Vis. Pattern Recognit. Workshops (CVPRW), vol. 3, Jul. 2017, pp. 1122–1131.
[17] R. Timofte, E. Agustsson, L. Van Gool, M. H. Yang, L. Zhang, B. Lim, S. Son, H. Kim, S. Nah, K. M. Lee, “NTIRE 2017 challenge on single image super-resolution: Methods and results,” in Proc. IEEE Comput. Vis. Pattern Recognit. Workshops (CVPRW), Jul. 2017, pp. 1110–1121.
[18] M. Bevilacqua, A. Roumy, C. Guillemot, and M.-L.-A. Morel, “Low-complexity single-image super-resolution based on nonnegative neighbor embedding,” in Proc. Brit. Mach. Vis. Conf., Surrey, U.K., 2012, pp. 1–10.
[19] R. Zeyde, M. Elad, and M.Protter, “On single image scale-up using sparse-representations,” in Proc. Int. Conf. Curves Surf. Avignon, France, Jun. 2010, pp. 711–730.
[20] D. Martin, C. Fowlkes, D. Tal, and J. Malik, “A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics,” in Proc. 8th IEEE Int. Conf. Comput. Vis. (ICCV), Jul. 2001, pp. 416–425.
[21] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” in Proc. Int. Conf. Learn. Represent. (ICLR), San Diego, CA, USA, May 2015, pp. 1–15.
[22] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, “Image super-resolution using very deep residual channel attention networks,” in Proc. Eur. Conf. Comput. Vis., 2018, pp. 294–310.

[23] K. Jiang, Z. Wang, P. Yi, and J. Jiang, “Hierarchical dense recursive network for image super-resolution,” Pattern Recognit., vol. 107, Nov. 2020, Art. no. 107475.

NANYAN LIU is currently an Associate Professor of computer science and technology. She has been engaged in teaching and research of computer science and application for 25 years with the Xi’an University of Science and Technology. Her research interests include computer graphics and image processing, database technology, and computer vision.

GUANGPU GAO received the bachelor’s degree in software engineering from Northwestern University, in 2017. He is currently pursuing the master’s degree in computer science and technology with the Xi’an University of Science and Technology. His research interests include deep learning, cyclic neural networks, image super resolution from reconstruction, and image segmentation.

XINYU XU is currently pursuing the master’s degree in computer science and technology with the Xi’an University of Science and Technology. His research interests include deep learning and image super resolution.

***