Neurocomputing xxx (xxxx) xxx

Contents lists available at ScienceDirect

Neurocomputing

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

Single image super-resolution with attention-based densely connected module

Zijian Wang [a,b], Yao Lu [a], Weiqi Li [a], Shunzhou Wang [a,*], Xuebo Wang [a], Xiaozhen Chen [a]

[a] Beijing Laboratory of Intelligent Information Technology, School of Computer Science, Beijing Institute of Technology, Beijing, China
[b] China Central Television, Beijing, China

ARTICLE INFO

Article history:
Received 12 February 2020
Revised 9 August 2020
Accepted 12 August 2020
Available online xxxx
Communicated by Zidong Wang

Keywords:
Super-resolution
Neural attention
Dense connection
Deep neural network

ABSTRACT

Benefited from the abundant features provided by the dense connection block, the densely connected-based super-resolution network has achieved superior performance in the single image super-resolution (SISR) task. However, the abundant features also introduce redundant and conflicting information, resulting in longer training time and unsatisfied image reconstruction results. To solve this problem, we propose an attention-based densely connected module (DAM). DAM consists of two parts: channel attention module (CAM) and dense connection block (DB). CAM is placed at the front of each DB and gives different weights of each channel from received features for suppressing redundant responses. Based on DAM, we propose an Attention-based Densely Connected Network (ADSRNet) for SISR, and explore the effectiveness of DAM on other densely connected-based super-resolution networks. Extensive experiments are performed on commonly-used super-resolution benchmarks. Qualitative and quantitative results demonstrate the effectiveness of our method.

1. Introduction

Single image super-resolution (SISR) is a low-level computer vision task, which aims to reconstruct the high-resolution image (HR) from its counterpart low-resolution image (LR). It has been widely used in multimedia and medical image processing fields and received an amount of research interest recently.

With the renaissance of deep learning, SISR achieves significant progress. Many super-resolution networks [1–10] are designed to reconstruct HR images by learning a nonlinear mapping relationship between the LR-HR pairs. For example, SRCNN [1] applies three convolutional layers to learn the mapping relationship between LR images and HR images; LapSRN [2] uses three networks to estimate the residual information for progressively $\times 8$ super-resolution; DRCN [3] utilizes residual learning to increase the depth of super-resolution network and achieves promising results; SRDenseNet [4] uses dense connection operations proposed by [11] to avoid the vanish gradient problem when model training, and learns a compact model for SISR.

Among them, benefited from the abundant features provided by the dense connection block, SRDenseNet achieves higher reconstruction accuracy compared with the other super-resolution networks [1–3]. However, the abundant features also include irrelevant information, which affects the quality of final image reconstruction. The high computational cost of the dense connection block also brings an obstacle for SRDenseNet to be applied in the real-world application. Therefore, reducing the effect of redundant features provided by the dense connection block is an effective way to improve the performance of SRDenseNet.

Inspired by the neural attention mechanism, many CNN-based attention modules [12–25] are proposed recently, which aim to enhance the task-relevant feature representation and suppress the task-irrelevant feature representation. Therefore, in this paper, we apply the neural attention mechanism to SRDenseNet to decrease feature redundancy, and propose a novel attention-based densely connected module (DAM). DAM includes two parts: channel attention module (CAM) and dense connection block (DB). CAM is placed at the front of each DB and generates channel weight scores to re-weight each channel feature map for suppressing redundant information. Further, we apply the proposed DAM to construct an Attention-based Densely connected network (ADSRNet) for SISR. With the help of DAM, ADSRNet generates more real-realistic SR images than SRDenseNet. Besides, we also explore the effectiveness of our proposed DAM on other densely-connected networks (*e.g.*, RDN [26], and DBPN [27]). We perform extensive experiments on commonly-used super-resolution benchmarks. Experiment results demonstrate the effectiveness of our method.

* Corresponding author.
E-mail address: shunzhouwang@bit.edu.cn (S. Wang).

https://doi.org/10.1016/j.neucom.2020.08.070
0925-2312/© 2020 Elsevier B.V. All rights reserved.

To summarize, the contributions of our work are described as follows:

- We propose the DAM for SISR, which can effectively reduce redundant information from DB and improve image reconstruction accuracy.
- We propose the ADSRNet for SISR, which achieves promising results on commonly-used super-resolution benchmarks.

This paper extends our conference paper [28] in three aspects. Firstly, to evaluate the effectiveness of DAM, we also combine DAM with other densely connected super-resolution networks, such as RDN [26] and DBPN [27], and perform extensive experiments to see the effect of DAM. Secondly, we compare our proposed DAM with other attention modules through quantitative and qualitative analysis. Thirdly, we extend the evaluation benchmark and evaluate ADSRNet on Urban100, and our model also achieves promising results.

The rest of this paper is as follows. We review the related works of our paper in Section 2. In Section 3, we introduce our proposed method in detail. In Section 4, we demonstrate the implementation details and present extensive experiments on public super-resolution benchmarks. We conclude our paper in Section 5.

2. Related work

2.1. Neural network-based single image super-resolution

Benefited from the success of deep learning technology, numerous deep learning-based methods are proposed recently. SRCNN [1] was the first proposed deep convolutional neural network for SISR. They firstly applied the bicubic interpolation to super-resolved the low-resolution image to the target size, and then learned the mapping relation from LR to HR images through a three-layer convolutional neural network. Kim et al. proposed a deeply-recursive convolutional network DRCN [3] and VDSR [29]. DRCN [3] applied the same convolutional layer multiple times in a recursive manner, which remains less computation cost and enlarges the receptive field. In VDSR [29], they applied the residual learning into the super-resolution network for training stability and fast convergence. Lim et al. [30] propose an enhanced deep super-resolution network (EDSR), which is built on ResNet [31] and removed all the batch normalization layers of the network.

However, the above methods need large computation cost, which is not suitable for real world application. To reduce the computation resource demand, Shi et al. [32] proposed a sub-pixel convolution layer to improve the computational efficiency of the network. Tai et al. proposed a deep recursive residual network (DRRN) [33] with less computation cost, which benefited from residual image learning.

Besides, most of the current super-resolution methods can get pleasing results only for a limited super-resolved scale. When the super-resolved scale becomes large, the results by one-step upsampling are unsatisfactory. To solve this problem, Lai et al. [2] used a cascaded network to learn high-frequency residual details progressively for $\times$ 8 super-resolution task. Hu et al. [34] proposed a novel method called Meta-SR to firstly solve the super-resolution of arbitrary scale factors with a single model. They designed a MetaUpscale Module [34], which dynamically predicted the weights of the upscale filters by taking the scale factor as input and used these weights to generate the HR image of arbitrary size.

Moreover, with the development of generative adversarial networks (GAN), many GAN-based super-resolution networks are proposed. For example, Ledig et al. [35] proposed the SRGAN, which combined the pixel-wise similarity loss, the perception loss, and the adversarial loss to train the super-resolution network for generating realistic reconstruction images with more texture details. EnhanceNet [36] applied the perception loss and the text matching loss to solve the overly smoothed image reconstruction problem. Wang et al. [37] proposed a deep spatial feature transform to recover textures conditioned on the categorical priors, which generated more realistic and visually pleasing textures in comparison to SRGAN [35] and EnhanceNet [36].

Thanks to powerful feature representative ability provided by dense connection module [11], many networks introduced the dense connection module to super-resolution task and achieved excellent results. DBPN [27] constructed mutually-connected up-sampling and down-sampling to imitate the process of super-resolution and image degradation to learn more about degradation patterns. Zhang et al. [26] proposed the RDN, which applied residual dense block to extract local features and global residual learning to learn global features. Different from the above methods, in this paper, we apply the neural attention mechanism to suppress the redundant information produced by the dense connection block and increase the feature representative abilities of densely connected super-resolution networks.

2.2. Neural attention mechanism

Neural Attention Mechanism is a hot research topic in recent years. It helps the neural network to focus on the important features and suppress the useless representations, which has been widely applied to many computer vision tasks, such as image classification [12,14,16], action recognition [18], human-object interaction detection [13], video object segmentation [15,17,46–48], saliency object detection [23], human parsing [25], video and image restoration [19–22,24], 3D object detection [49] and so on.

Among them, channel attention module (CA) and spatial attention module (SA) are two commonly-used neural attention models. CA focus on modeling the channel relationships of features, and SA aims to extract highlight regions on each feature map's spatial dimension. For CA, Hu et al. [12] proposed the squeeze-and-excitation (SE) block, which squeezed the original features with the global average pooling, and performed an extraction operation by two fully connected layers to learn the weights of different channels. Motivated by SEblock, Woo and Park et al. [14] proposed a convolutional block attention module (CBAM), which expanded channel attention dimension into channel attention and spatial attention dimensions, and processed features with average-pooling and max-pooling operations for increasing the effect of attention module. Based on the CBAM and SE block, the channel-wise and spatial attention residual block (CSAR) [20] and the residual attention module (RAM) [22] were proposed recently for SISR. Both of them achieved promising results on commonly-used super-resolution benchmarks. In this paper, we modify the original SE block and apply it into dense connection module for suppressing the redundant information.

3. Proposed method

In this section, we first introduce the details of our proposed DAM. Then, we demonstrate the detailed components of our proposed ADSRNet.

3.1. Dense attention module

DAM is the key component of our proposed ADSRNet, which aims to reduce the effect of redundant information generated by DB. The detailed structure of the DAM is illustrated in the bottom-left of Fig. 1. DAM includes two parts: DB and CAM. DB

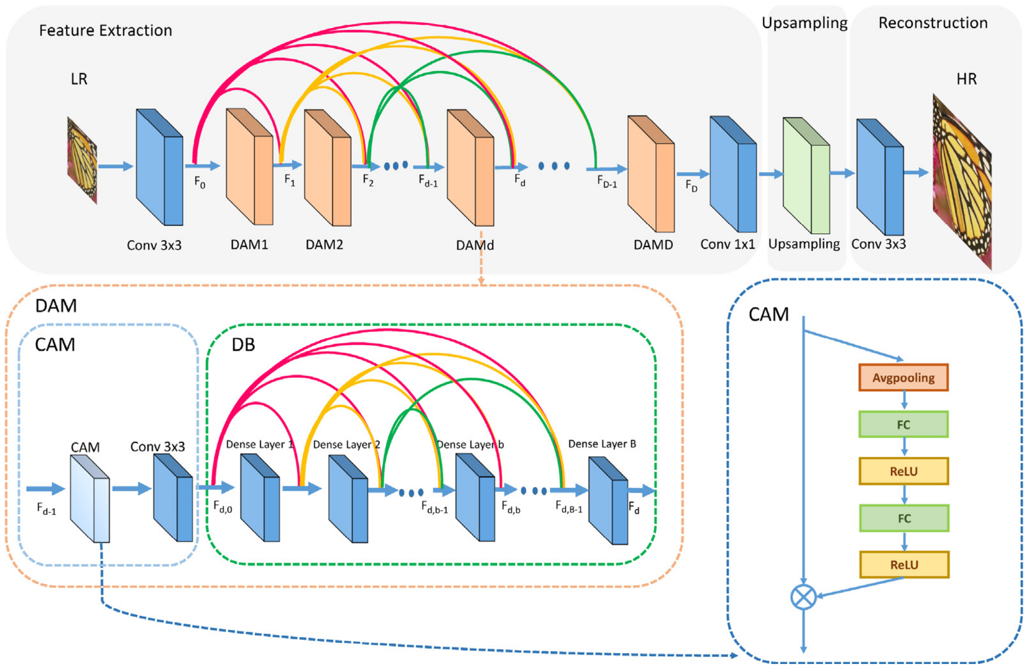


Fig. 1. Overview of our proposed ADSRNet. The sub-figure on the bottom-left presents our proposed DAM, it consists of a CAM and a DB. The sub-figure on the bottom-right shows the structure of CAM. 'Conv' denotes the convolutional layer and 'FC' represents the fully connected layer.

is applied to extract hierarchical features for image reconstruction. CAM process the received features from previous DBs and give different weights to each channel feature map to decrease the effect of the redundant information produced by the dense connection between DBs.

The detailed structure of CAM is shown in the bottom-right of Fig. 1. We first apply the average pooling to process the input features for acquiring global information. After that, the feature maps are fed to FC-ReLU-FC-ReLU operations, which generates the final re-calibrated weight vectors. Finally, the re-calibrated weight vectors are reshaped to the original input feature size and multiplied with the original input feature to generate the final attentive output. The process of the d-th CAM is formulated as

$$F_{d,0} = f_{att}(F_{d-1}), \tag{1}$$

where F_{d-1} and $F_{d,0}$ ($d \in \{1 \cdots D\}$) are the input and output features of the d-th channel attention module. D represents the number of DAM in the final network. $f_{att}(\cdot)$ is the function of CAM.

DB is firstly proposed in [11], and its structure is shown in the bottom-left of Fig. 1. Each DB accepts the feature maps produced by all previous DBs and processes them to extract hierarchical features. For achieving multi-level information flow, each DAM receives the output of all previous DAMs.

3.2. Attention-based densely connected network

We apply our proposed DAM to SRDenseNet [4] and propose an ADSRNet for SISR. The overall architecture of ADSRNet is illustrated in Fig. 1, which includes three parts: feature extraction part, upsampling part, and image reconstruction part. In the feature extraction part, we first use one 1×1 convolutional layer to extract the low-level features

$$F_0 = f_{low}(I_{LR}), \tag{2}$$

where I_{LR} denotes the LR input image, and F_0 denotes the shallow features extracted from I_{LR},$f_{low}(\cdot)$ denotes the 1×1 convolution operation. Then, we use DAMs to extract the hierarchical high-level features

$$F_d = f_{high}(F_0), \tag{3}$$

where $F_d(d \in \{1 \cdots D\})$ denotes the output of the d-th DAM, $f_{high}(\cdot)$ represents the operations of the DAM. After that, to reduce the dimensions of output feature, we utilize a 1×1 convolutional layer to process the output feature. Finally, the extracted features are fed to upsampling layer and reconstruction layer sequentially for final super-resolution. We choose the deconvolutional layers as the upsamping layer, and 3×3 convolutional layer as the reconstruction layer. The detail process is as follows:

$$I_{SR} = f_{recon}(f_{up}(F_D)), \tag{4}$$

where F_D denotes the output of the D-th DAM, $f_{up}(\cdot)$ denotes the upsampling operation, and $f_{recon}(\cdot)$ denotes the reconstruction operation. We apply $L1$ loss to optimize ADSRNet. The optimization loss L is as follows

$$L = \|I_{SR} - I_{HR}\|_1, \tag{5}$$

where I_{SR} represents the output of ADSRNet, and I_{HR} is the counterpart high-resolution image.

4. Experiment

In this section, we first introduce the evaluation datasets and metrics. Then, we describe the implementation details of the experiments. After that, we introduce the comparison results with other state-of-the-art methods. Finally, we give a comprehensive model analysis of our proposed ADSRNet.

4.1. Datasets and evaluation metrics

ADSRNet is trained on DIV2K [38], which consists of 800 high-quality training images. We use the bicubic interpolation to get the LR images from corresponding HR images. For data augmentation, the patches are cropped from original images and are randomly horizontal flipped, and rotated. ADSRNet is evaluated on four commonly-used benchmarks, such as Set5 [39], Set14 [40], BSD100 [41] and Urban100 [42]. The outputs of ADSRNet are converted to YCbCr space and is evaluated by calculating PSNR and SSIM [43] on only Y channel.

4.2. Implementation details

ADSRNet consists of 7 DAMs. Each DAM includes 8 dense layers. We apply the Adam optimizer to optimize our model. The initial learning rate is set to 0.0001, and the batch size is set to 64. We train our model for 60 epochs in total. All experiments are performed on a single NVIDIA GeForce RTX 2080 Ti with the Pytorch framework.

4.3. Comparisons with state-of-the-art methods

We compare ADSRNet with other state-of-the-art super-resolution networks, such as SRCNN [1], VDSR [29], DRCN [3], DRRN [33], LapSRN [2], SRDenseNet [4], RDN [26], DBPN [27], SRFBN [44], and Soft-Edge [45]. The detailed quantitative results are displayed in Table 1. We see that our proposed ADSRNet achieves a significant improvement over the average PSNR and SSIM compared with the previous network [1,29,3,33,2]. Compared with SRDenseNet [4], ADSRNet is marginally higher by 0.38 dB, 0.23 dB, 0.14 dB and 0.41 dB in PSNR for $\times 4$ SR on Set 5, Set 14, BSD100 and Urban100, respectively. Visualization results of different methods are displayed in Fig. 2. We find that ADSRNet generates more realistic image reconstruction results compared with other methods.

4.4. Model analysis

4.4.1. The analysis of dense connection based super-resolution network with DAM

We employ the DAM to RDN [26] and DBPN [27]. For RDN, we replace all residual dense blocks (RDBs) of RDN into our proposed DAMs with the local residual learning of RDB. For DBPN, we place CAM at the front of each Up/Down projection module and construct a modified 'DAM'. The quantitative results are shown in Table 1.

We see that RDN w/DAM and DBPN w/DAM do not achieve higher reconstruction accuracy than their original model. The reason is that the injection of DAM into the networks affects the original gradient optimization process of RDN and DBPN, which hinders the final reconstruction accuracy. Specifically, for RDN, each RDB of RDN only receives its previous RDB output. There is no dense connection operation between RDBs. Therefore, the feature redundancy problem our proposed DAM aims to solve does not exist in RDN. On the contrary, compared with the original RDB, our proposed DAM has the newly added module, CAM, which increases the depth and complexity of original RDN and brings an obstacle to the gradient optimization process. Therefore, the performance of RDN w/DAM does not outperform the original RDN.

For DBPN, DAM process the input of each Up/Down-projection module from DBPN. The Up/Down projection module has corrected the projection error of each hierarchical feature, and all modified hierarchal features are fed to the following projection modules with dense connection operation. However, the CAM is introduced to the front part of each projection module, which changes the representatives of the modified hierarchal features via enhancing some features and weakening the other feature. The noise information is introduced to the modified features with the operation of CAM, which are not appropriate for final image reconstruction. Therefore, the performance of DBPN w/DAM also drops.

Based on the above analysis, our proposed DAM is inappropriate for densely connected super-resolution networks with complicated information flows. However, ADSRNet (SRDenseNet w/ DAM) performs better than SRDenseNet on $\times 2$ and $\times 4$ SR images, and achieves comparable results with RDN w/DAM and DBPN w/DAM on $\times 4$ SR images, even outperforms them on Set14 and BSD100, which demonstrates the effectiveness of our proposed DAM to SRDenseNet. From Fig. 3, we can see that the reconstruction results

Table 1
Comparison results of different methods for $\times 2$ and $\times 4$ image super-resolution. Red and blue indicate the best results and the second best results, respectively. '–' denotes the results we can not provide due to the limited GPU memory.

Scale	Method	Set5		Set14		BSD100		Urban100	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
2×	Bicubic	33.68	0.9304	30.24	0.8691	29.56	0.8435	27.39	0.8410
	SRCNN [1]	36.66	0.9542	32.45	0.9067	31.36	0.8879	29.52	0.8950
	VDSR [29]	37.53	0.9597	33.05	0.9127	31.90	0.8960	30.77	0.9140
	LapSRN [2]	37.52	0.9591	32.99	0.9124	31.80	0.8949	31.05	0.9100
	DRCN [3]	37.63	0.9588	33.04	0.9118	31.85	0.8942	30.76	0.9130
	DRRN [33]	37.74	0.9591	33.23	0.9136	32.05	0.8973	31.23	0.9190
	SRFBN [44]	38.18	0.9611	33.90	0.9203	32.34	0.9015	32.80	0.9341
	SeaNet [45]	38.15	0.9611	33.86	0.9198	32.31	0.9013	32.68	0.9332
	DBPN [27]	38.09	0.9600	33.85	0.9190	32.27	0.9000	33.02	0.9310
	RDN [26]	38.24	0.9614	34.01	0.9212	32.34	0.9017	33.09	0.9368
	SRDenseNet	37.83	0.9605	33.25	0.9154	32.06	0.8994	–	–
	DBPN w/DAM	37.96	0.9610	33.37	0.9162	32.14	0.9000	–	–
	RDN w/DAM	38.11	0.9616	33.47	0.9168	32.20	0.9008	–	–
	ADSRNet(**ours**)	37.88	0.9606	33.30	0.9153	32.09	0.8994	31.59	0.9233
4×	Bicubic	28.43	0.8109	26.00	0.7023	25.96	0.6678	23.14	0.6577
	SRCNN [1]	30.48	0.8628	27.50	0.7513	26.90	0.7103	24.52	0.7221
	VDSR [29]	31.35	0.8838	28.02	0.7678	27.29	0.7252	25.18	0.7530
	LapSRN [2]	31.54	0.8866	28.09	0.7694	27.32	0.7264	25.21	0.7553
	DRCN [3]	31.53	0.8854	28.02	0.7670	27.23	0.7233	25.14	0.7520
	DRRN [33]	31.68	0.8888	28.21	0.7720	27.38	0.7284	25.44	0.7638
	SRFBN [44]	32.56	0.8992	28.87	0.7881	27.77	0.7419	26.73	0.8043
	SeaNet [45]	32.33	0.8970	28.72	0.7855	27.65	0.7388	26.32	0.7942
	DBPN [27]	32.47	0.8980	28.82	0.7860	27.72	0.7400	27.08	0.7950
	RDN [26]	32.47	0.8990	28.81	0.7871	27.72	0.7419	26.61	0.8028
	SRDenseNet	31.54	0.8834	28.12	0.7712	27.32	0.7296	25.36	0.7640
	DBPN w/DAM	31.92	0.8890	28.33	0.7749	27.43	0.7322	–	–
	RDN w/DAM	31.95	0.8897	28.35	0.7767	27.44	0.7333	25.78	0.7782
	ADSRNet(**ours**)	31.92	0.8894	28.35	0.7756	27.46	0.7325	25.77	0.7757

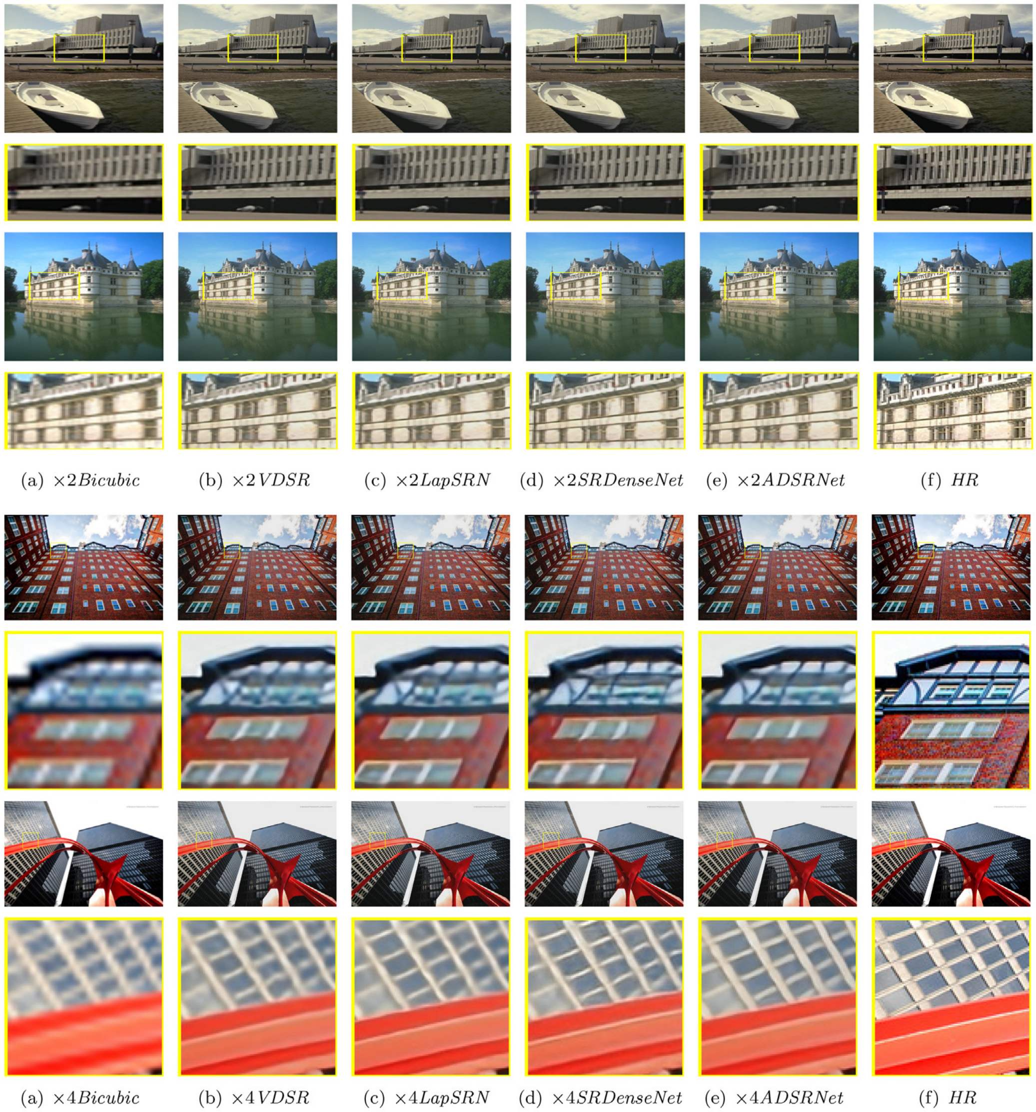


Fig. 2. Qualitative comparison for $\times 2$ and $\times 4$ SR images. Our proposed ADSRNet can generate more realistic reconstruction images.

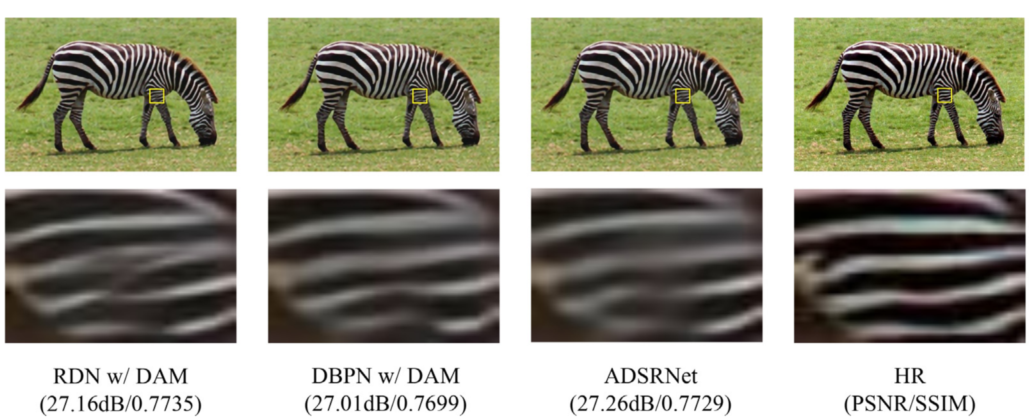


Fig. 3. Visual comparisons of different dense connection based super resolution networks with DAM for $\times 4$ SR images on Set14.

of ADSRNet are more realistic than the other methods, which further proves the effectiveness of our DAM. To summarize, our proposed DAM is suitable for densely connected based super-resolution network with plain information flows, like SRDenseNet.

4.4.2. Compared with other attention modules

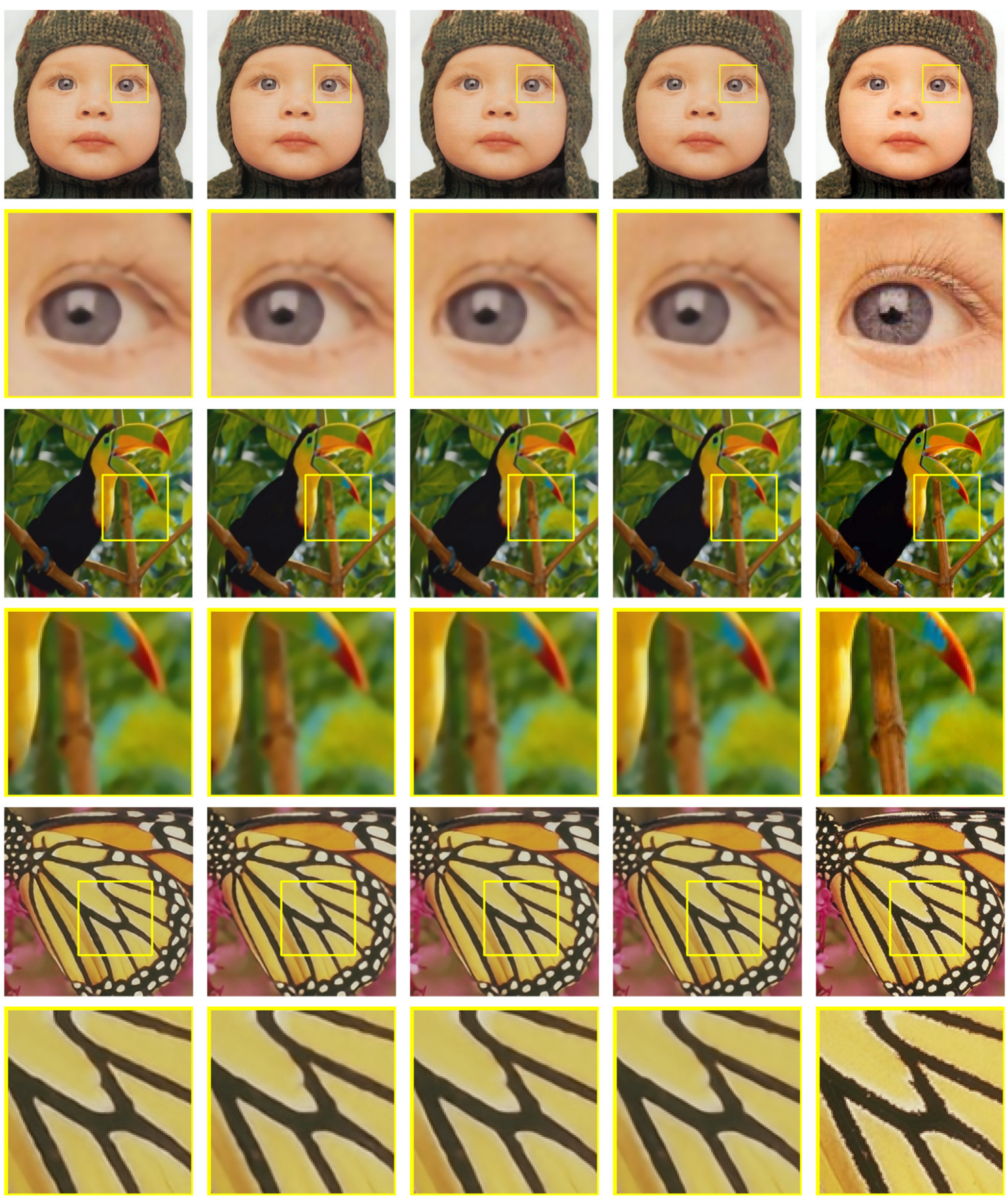
To evaluate the effectiveness of CAM for DB, we also explore the performance of ADSRNet with other attention modules, such as CBAM [14], CSAR [20] and RAM [22]. We replace the CAM in each DAM of ADSRNet with other attention modules. Table 2 displays the comparison results. Although other attention modules have a more complicated structure than a single channel attention module, they don't achieve satisfactory results. We argue that complex attention operations, such as spatial attention (CBAM, CSAR) or second-order attention (RAM), are not conducive to optimizing dense connection modules and degrade the performance of the dense connection-based super-resolution network. However, different attention modules generate similar visual results from the perspective of visual effects, as shown in Fig. 4. To some extent, the attention module filters out invalid information and plays a role in the image reconstruction process.

4.4.3. The comparisons of convergence speed

We further analyze the convergence speed of SRDenseNet and ADSRNet. The convergence curves are presented in Fig. 5. The blue curve denotes the performance of the SRDenseNet, and the yellow curve denotes the performance of our ADSRNet. We can see that the curve of ADSRNet converges more faster than SRDenseNet, and performs high reconstruction accuracy than SRDenseNet, which proves the effectiveness of our proposed ADSRNet.

Table 2
Comparison results of ADSRNet with different attention modules for × 4 SR on Set 5.

Method	ADSRNet w/CBAM	ADSRNet w/RAM	ADSRNet w/CSAR	ADSRNet w/CA (Ours)
PSNR	31.65	31.71	31.80	**31.92**



(a) CBAM (b) RAM (c) CSAR (d) CA(**Ours**) (e) HR

Fig. 4. Visual comparisons of ADSRNet with different attention modules for ×4 SR images on Set5.

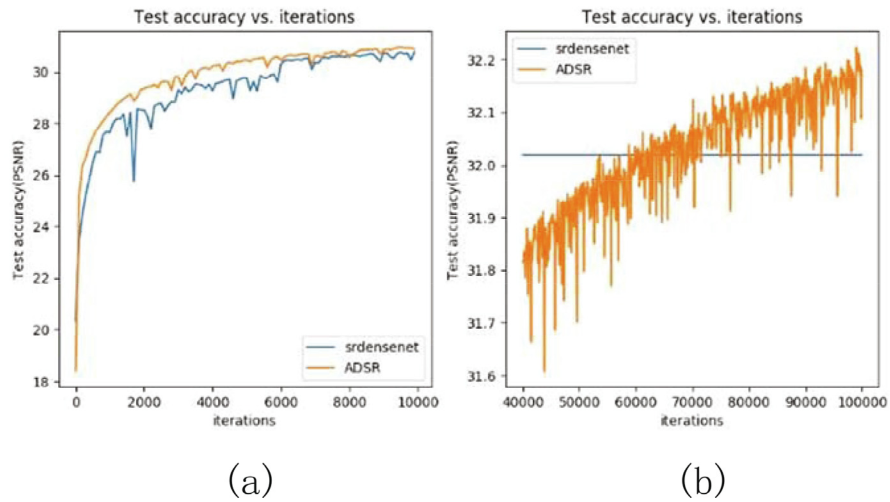


Fig. 5. Convergence curves of SRDenseNet [4] and ADSRNet for ×4 SR images on Set5. (a) presents the convergence curves of the first 10000 iterations, and (b) presents the convergence curves between 40000 and 100000 iterations. (b) is enlarged for better view. X and Y axes represent iterations and PSNR(dB), respectively.

5. Conclusions

In this paper, we propose a new attention module DAM to decrease redundant information provided by DB, and propose the ADSRNet for SISR. With DAM's help, ADSRNet generates more photo-realistic image reconstruction results compared with other super-resolution networks and achieves high reconstruction accuracy. Experiment results on commonly-used super-resolution benchmarks demonstrate the effectiveness of our method. However, our method is only developed for SISR on synthetic data and does not consider the real-world application situation whose degradation model is unknown. In future work, we will extend our ADSRNet to solve the real-world super-resolution task.

CRediT authorship contribution statement

Zijian Wang: Conceptualization, Methodology, Software, Writing - original draft. **Yao Lu:** Methodology, Writing - review & editing. **Weiqi Li:** Conceptualization, Methodology, Writing - review & editing. **Shunzhou Wang:** Writing - review & editing. **Xuebo Wang:** Writing - review & editing. **Xiaozhen Chen:** Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (No. 61273273), by the National Key Research and Development Plan (No. 2017YFC0112001), and by China Central Television (JG2018-0247).

References

[1] C. Dong, C.C. Loy, K. He, X. Tang, Image super-resolution using deep convolutional networks, IEEE Trans. Pattern Anal. Mach. Intell. 38 (2) (2015) 295–307.
[2] 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, 2017, pp. 624–632.
[3] 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.
[4] T. Tong, G. Li, X. Liu, Q. Gao, Image super-resolution using dense skip connections, in: Proceedings of the IEEE Conference on Computer Vision, 2017, pp. 4799–4807.
[5] Y. Guo, J. Chen, J. Wang, Q. Chen, J. Cao, Z. Deng, Y. Xu, M. Tan, Closed-loop matters: dual regression networks for single image super-resolution, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2020, pp. 5407–5416.
[6] Y. Mei, Y. Fan, Y. Zhou, L. Huang, T.S. Huang, H. Shi, Image super-resolution with cross-scale non-local attention and exhaustive self-exemplars mining, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2020, pp. 5690–5699.
[7] G. Li, Y. Lu, L. Lu, Z. Wu, X. Wang, S. Wang, Semi-blind super-resolution with kernel-guided feature modification, in: IEEE International Conference on Multimedia & Expo Workshops, 2020, pp. 1–6.
[8] J. Liu, W. Zhang, Y. Tang, J. Tang, G. Wu, Residual feature aggregation network for image super-resolution, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2020, pp. 2359–2368.
[9] T. Shang, Q. Dai, S. Zhu, T. Yang, Y. Guo, Perceptual extreme super-resolution network with receptive field block, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2020, pp. 440–441.
[10] Z. Hui, X. Gao, Y. Yang, X. Wang, Lightweight image super-resolution with information multi-distillation network, in: Proceedings of the 27th ACM International Conference on Multimedia, 2019, pp. 2024–2032.
[11] G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, Densely connected convolutional networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 4700–4708.
[12] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 7132–7141.
[13] T. Zhou, W. Wang, S. Qi, H. Ling, J. Shen, Cascaded human-object interaction recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2020, pp. 4263–4272.
[14] S. Woo, J. Park, J.-Y. Lee, I. So Kweon, Cbam: convolutional block attention module, in: Proceedings of the European Conference on Computer Vision, 2018, pp. 3–19.
[15] W. Wang, X. Lu, J. Shen, D. J. Crandall, L. Shao, Zero-shot video object segmentation via attentive graph neural networks, in: Proceedings of the IEEE Conference on Computer Vision, 2019, pp. 9236–9245.
[16] F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang, X. Tang, Residual attention network for image classification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 3156–3164.
[17] W. Wang, H. Song, S. Zhao, J. Shen, S. Zhao, S.C. Hoi, H. Ling, Learning unsupervised video object segmentation through visual attention, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 3064–3074.
[18] X. Wang, R. Girshick, A. Gupta, K. He, Non-local neural networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 7794–7803.

[19] S. Anwar, N. Barnes, Real image denoising with feature attention, in: Proceedings of the IEEE Conference on Computer Vision, 2019, pp. 3155–3164.
[20] Y. Hu, J. Li, Y. Huang, X. Gao, Channel-wise and spatial feature modulation network for single image super-resolution, arXiv:1809.11130 (2018).
[21] Z. Shen, W. Wang, X. Lu, J. Shen, H. Ling, T. Xu, L. Shao, Human aware motion deblurring, in: Proceedings of the IEEE Conference on Computer Vision, 2019, pp. 5572–5581.
[22] J.-H. Kim, J.-H. Choi, M. Cheon, J.-S. Lee, Ram: Residual attention module for single image super-resolution, arXiv:1811.12043 (2018).
[23] W. Wang, S. Zhao, J. Shen, S.C. Hoi, A. Borji, Salient object detection with pyramid attention and salient edges, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 1448–1457.
[24] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, Y. Fu, Image super-resolution using very deep residual channel attention networks, in: Proceedings of the European Conference on Computer Vision, 2018, pp. 286–301.
[25] W. Wang, H. Zhu, J. Dai, Y. Pang, J. Shen, L. Shao, Hierarchical human parsing with typed part-relation reasoning, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2020, pp. 8929–8939.
[26] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, Y. Fu, Residual dense network for image super-resolution, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 2472–2481.
[27] M. Haris, G. Shakhnarovich, N. Ukita, Deep back-projection networks for super-resolution, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 1664–1673.
[28] W. Li, Y. Lu, X. Wang, X. Chen, Z. Wang, Adsrnet: Attention-based densely connected network for image super-resolution, in: Chinese Conference on Pattern Recognition and Computer Vision, 2019, pp. 272–282.
[29] 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.
[30] B. Lim, S. Son, H. Kim, S. Nah, K. Mu Lee, Enhanced deep residual networks for single image super-resolution, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2017, pp. 136–144.
[31] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
[32] 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.
[33] 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, 2017, pp. 3147–3155.
[34] X. Hu, H. Mu, X. Zhang, Z. Wang, T. Tan, J. Sun, Meta-sr: a magnification-arbitrary network for super-resolution, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 1575–1584.
[35] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. 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, 2017, pp. 4681–4690.
[36] M. S. Sajjadi, B. Scholkopf, M. Hirsch, Enhancenet: Single image super-resolution through automated texture synthesis, in: Proceedings of the IEEE Conference on Computer Vision, 2017, pp. 4491–4500.
[37] X. Wang, K. Yu, C. Dong, C. Change Loy, Recovering realistic texture in image super-resolution by deep spatial feature transform, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 606–615.
[38] E. Agustsson, R. Timofte, Ntire 2017 challenge on single image super-resolution: Dataset and study, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2017, pp. 126–135.
[39] M. Bevilacqua, A. Roumy, C. Guillemot, M. L. Alberi-Morel, Low-complexity single-image super-resolution based on nonnegative neighbor embedding, in: British Machine Vision Conference, 2012, pp. 135.1–135.10.
[40] R. Zeyde, M. Elad, M. Protter, On single image scale-up using sparse-representations, in: International conference on curves and surfaces, 2010, pp. 711–730.
[41] D. Martin, C. Fowlkes, D. Tal, J. Malik, A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics, in: Proceedings of the IEEE Conference on Computer Vision, 2001, pp. 416–423.
[42] 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.
[43] 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.
[44] Z. Li, J. Yang, Z. Liu, X. Yang, G. Jeon, W. Wu, Feedback network for image super-resolution, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 3867–3876.
[45] F. Fang, J. Li, T. Zeng, Soft-edge assisted network for single image super-resolution, IEEE Trans. Image Process. 29 (2020) 4656–4668.
[46] T. Zhou, S. Wang, Y. Zhou, Y. Yao, J. Li, L. Shao, Motion-attentive transition for zero-shot video object segmentation, in: Proceedings of the 34th AAAI Conference on Artificial Intelligence, 2020, pp. 13066-13073.
[47] Zhou T., Li J., Wang S., Tao R., Shen J., Matnet: Motion-attentive transition network for zero-shot video object segmentation, IEEE Transactions on Image Processing 29 (2020) 8326–8338.
[48] X. Lu, W. Wang, D. Martin, T. Zhou, J. Shen, V. G. Luc, Video object segmentation with episodic graph memory networks, in: Proceedings of the European Conference on Computer Vision, 2020.
[49] Q. Meng, W. Wang, T. Zhou, J. Shen, L. Van Gool, D. Dai, Weakly supervised 3d object detection from lidar point cloud, in: Proceedings of the European Conference on Computer Vision, 2020.

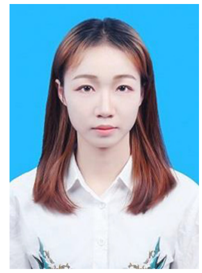
Zijian Wang received the B.S. degree of animation technology from Communication University of China, in 2004, and M.E. degree of software engineering from Beijing University of Posts and Telecommunications, in 2011. He is engaged in computer visual effect development in China Central Television. His research interests include image processing and pattern recognition.

Yao Lu received the B.S. degree in electronics from Northeast University, Shenyang, China, in 1982 and the Ph.D. degree in computer science from Gunma University, Gunma, Japan, in 2003. He was a Lecturer and an Associate Professor with Hebei University, China, from 1986 to 1998, and a foreign researcher with Gunma University in 1999. In 2003, he was an invited professor with the Engineering Faculty, Gunma Universitya Visiting Fellow of University of Sydney, Australia. He is currently a Professor with the Department of Computer Science, Beijing Institute of Technology, Beijing, China. He has published more than 100 papers in international conferences and journals. His research interests include neural network, image processing and video analysis, and pattern recognition.

Weiqi Li received the B.S. degree from Jilin University, in 2017. She is currently pursuing the M.E. degree in Computer Science at Beijing Institute of Technology, Beijing, China. Her main research interests include Image processing and pattern recognition.

Shunzhou Wang is currently pursuing the Ph.D. degree at Beijing Laboratory of Intelligent Information Technology, School of Computer Science, Beijing Institute of Technology, Beijing, China, from 2018. His supervisor is Prof. Yao Lu and his main research interests include computer vision and deep learning.

Xuebo Wang received the B.S. degree from Taiyuan University of Technology, in 2017. He is currently pursuing the M.E. degree in Computer Science at Beijing Institute of Technology, Beijing, China. His main research interests include Image processing and pattern recognition.



Xiaozhen Chen received the B.S. degree from Shenyang Institute of Engineering, in 2015. She is currently pursuing the M.E. degree in Biomedical Engineering at Beijing Institute of Technology, Beijing, China. Her main research interests include Image processing and pattern recognition.