EESR:Edge Enhanced Super-Resolution

Ying-Yue Li[1], Yun-Dong Zhang[2], Xue-Wu Zhou[2], Wei Xu [1]*

[1]State Key Laboratory of ASIC and System, Fudan University, No.825 ZhangHeng Rd. Shanghai 201203, China

[2] Vimicro AI Chip Technology Corporation, No. 35, XueYuan Road, Beijing 100191, China

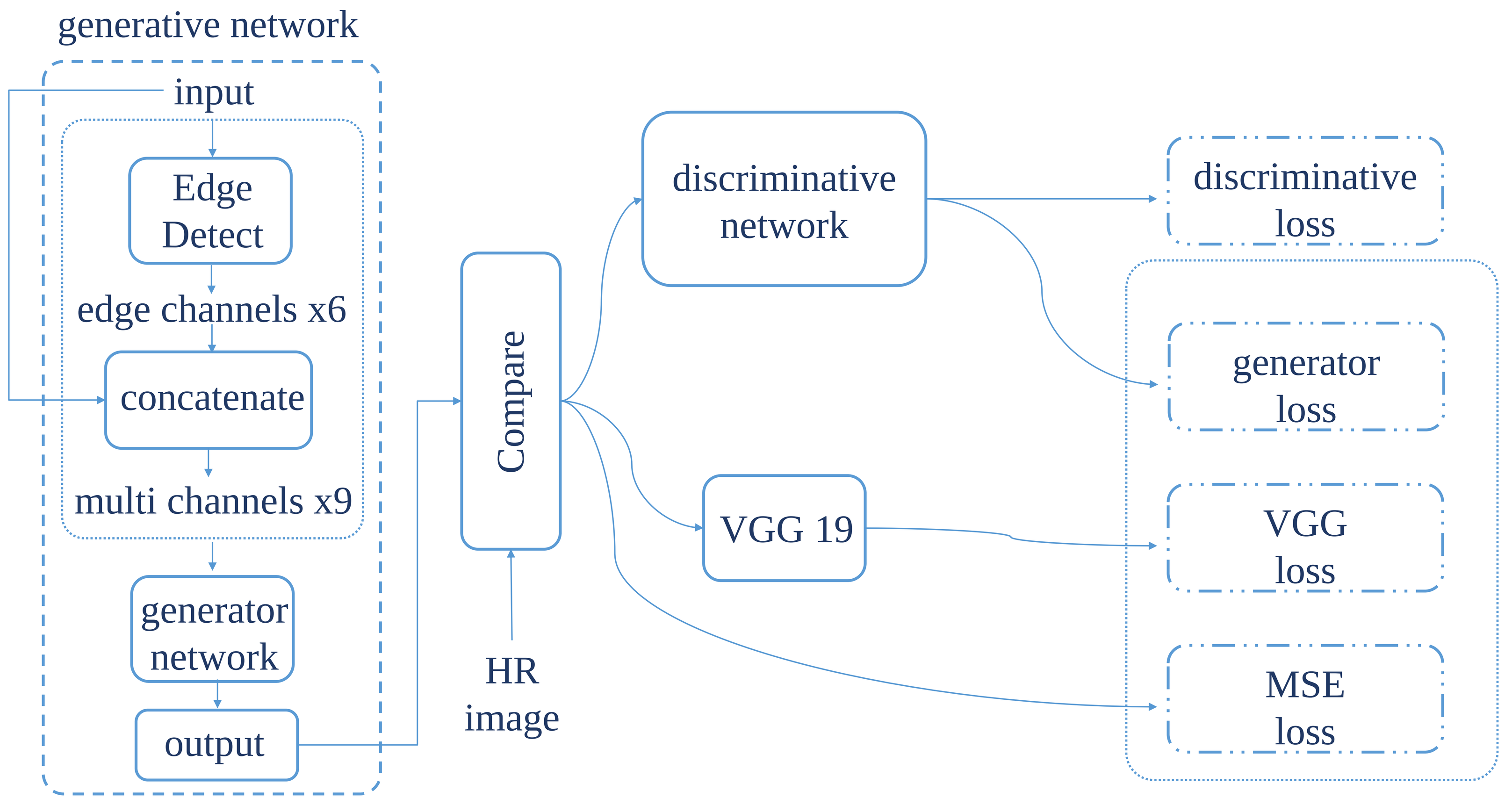
* Email: wei_xu@fudan.edu.cn

INTRODUCTION

- **Motivation – To generate high-frequency structures in blind super resolution**
  - Super Resolution (SR): Enhancing the resolution of and imaging system.
  - Traditional methods: Over-smoothly as a result of pursuing higher PSNR (peak signal to noise) and SSIM (structural similarity).
  - Recent networks using generative adversarial network (GAN) [1]: Having photo-realistic result, but not easy to train.
- **Contribution**
  - An edge detection network is used to extract and preserve high-frequency structures.
  - GAN architecture is used for fine-tuning the generative network, instead of direct training.
  - A new criterion "perceptive texture score" (PTS) is raised for better evaluating an image's high-frequency structures.

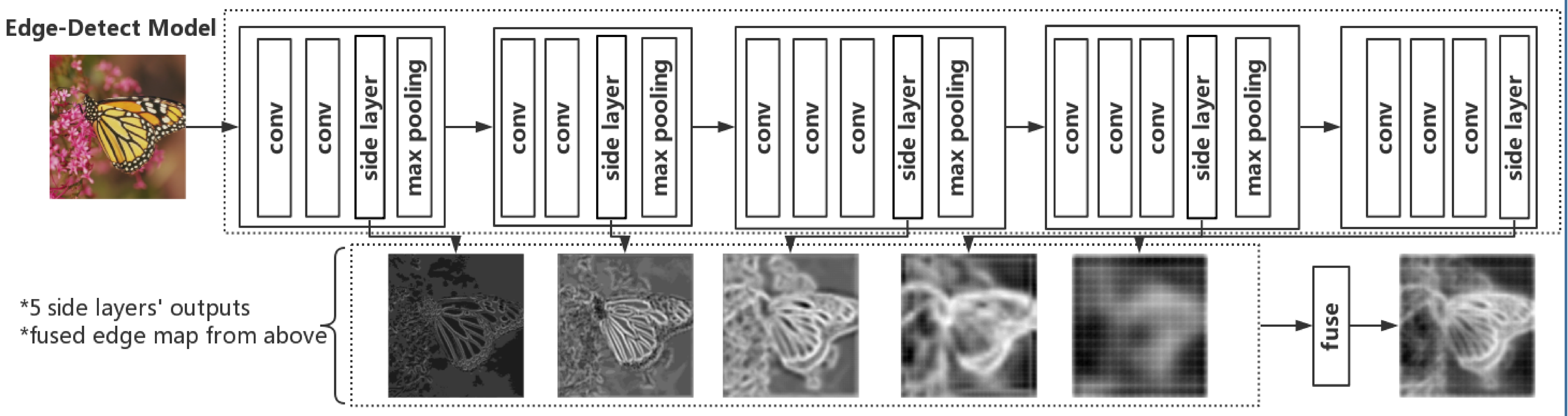
Low resolution input　　Our result　　Ground truth

EESR FRAMEWORK

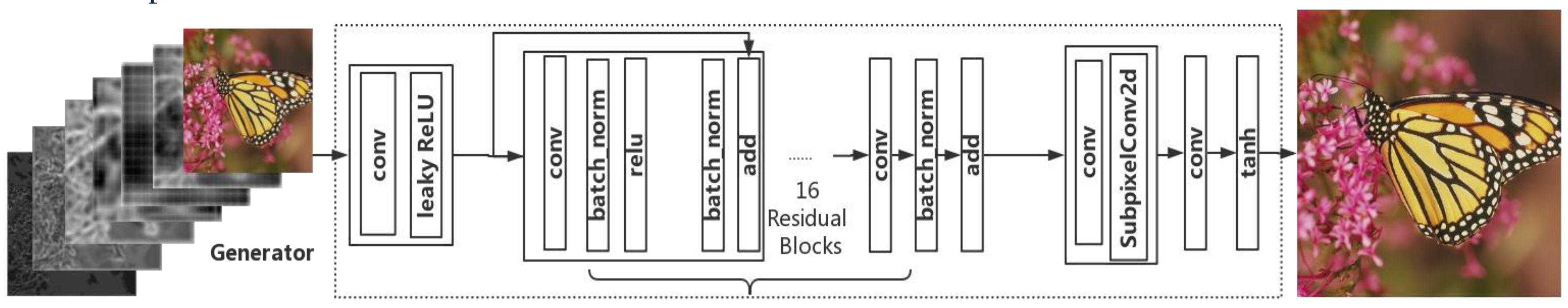


EESR has a generative network for training and a discriminative network for fine-tuning.

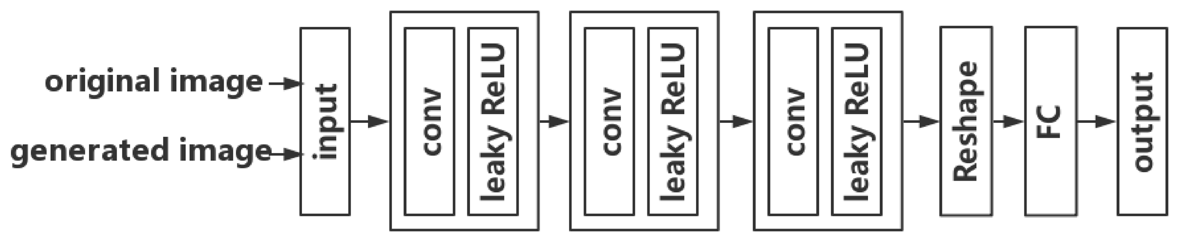
- Inference Steps
  - Compute the edge channels using Edge Detect network.
  - Concatenate edge channels with the input LR image to form a 9-channel input.
  - Feed the generator network with the 9-channel input.
- Edge Detect Network
  - HED[2] model is used for edge detection.
  - 5 edge maps, together with 1 fused map, are concatenated with 3-channel RGB (from original LR image) to form a 9-channel input map of the generator network.



- Generator Network
  - Responsible for mapping LR image I^{LR} to SR image $G(I^{LR})$.
  - Able to have a good preservation of high-frequency features of the original image ,as the feature maps are multi-scale and multi-level.



- Discriminator Network
  - Discriminate SR samples from original HR images.
  - Induce the generative network to generate indistinguishable SR images.
  - Only trained in fine-tuning iterations.



- Loss Function
  - EESR network has two loss functions: generative and discriminative.
  - Discriminative loss is optimized during fine-tuning:

$$L_{GAN-D} = \mathbb{E}_{\delta}(D(I^{HR}), \mathbb{I}) + \mathbb{E}_{\delta}(D(G(I^{LR})), \mathbb{O})$$

  - Generative loss:
    - The loss from generator in GAN: $L_{GAN-G} = \mathbb{E}_{\delta}(D(G(I^{LR})), \mathbb{I})$
    - Mean Square Error (MSE): $L_{MSE} = \frac{1}{r^2WHC} \sum_{x=1}^{rW} \sum_{y=1}^{rH} \sum_{z=1}^{C} (I^{HR}_{x,y,z} - G(I^{LR})_{x,y,z})^2$
    - VGG[3] loss: $L_{VGG} = \frac{1}{MNK} \sum_{i=1}^{M} \sum_{j=1}^{N} \sum_{k=1}^{K} (V(I^{HR})_{i,j,k} - V(G(I^{LR}))_{i,j}, k)^2$

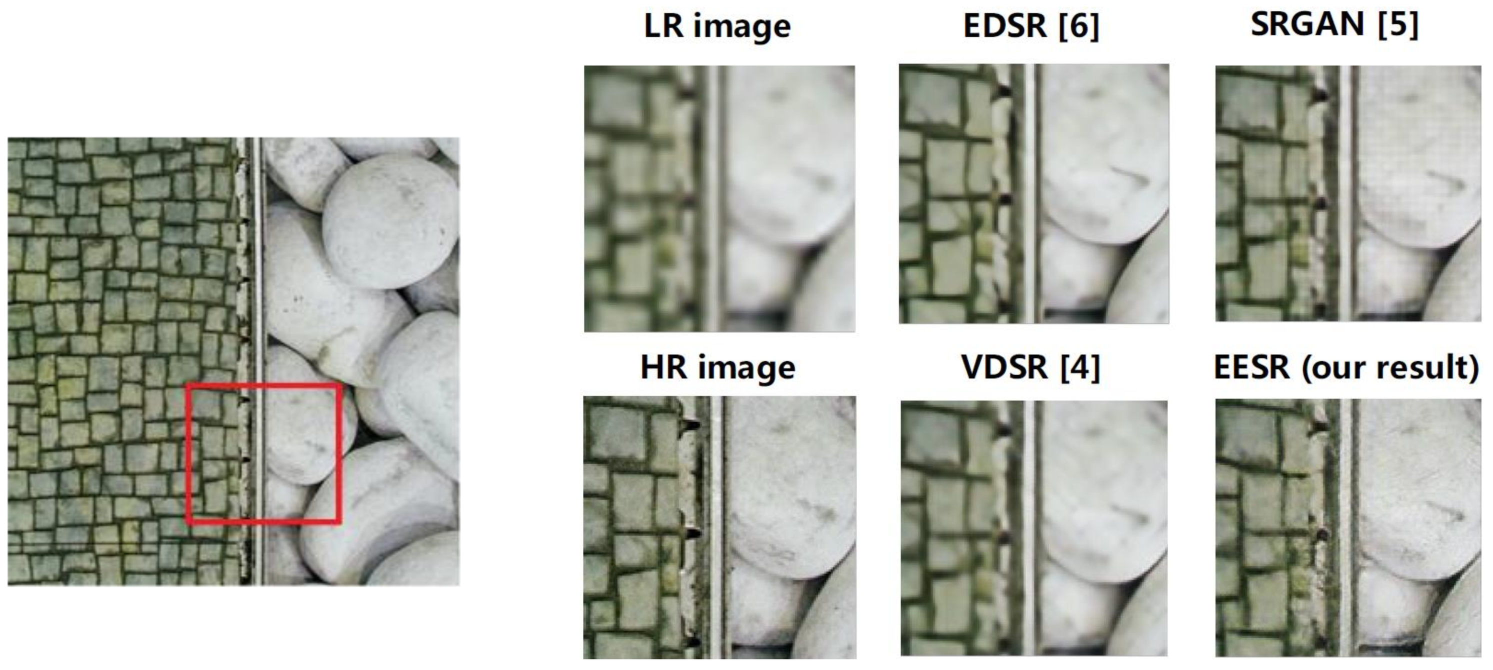
    The final loss function of generative network is a weighted mean of these three losses.

EVALUATION AND EXPERIMENTS

- Evaluation - perceptive texture score (PTS)
  - To quantify the quality of generated super resolved images.
  - Definition: $\text{PTS} = \frac{\sum_i^W \sum_j^H abs(G(I^{LR}) \otimes \mathbb{K}_{PTS})_{i,j}}{\sum_i^W \sum_j^H abs(I^{HR} \otimes \mathbb{K}_{PTS})_{i,j}}$, where $\mathbb{K}_{PTS}[i] = \begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}, (i = 0,1,2)$
- Experiments
  - Developing platform : Tensorflow 1.4.1, NVIDIA TITAN XP GPU
  - Training and testing dataset : DIV2K
  - Training details:
    - Initial learning rate 1e-4, decay rate of learning rate 1e-5;
    - Training iterations 200 epochs, fine-tuning iterations 2000 epochs
    - Batch size 16

Comparison with previous works on DIV2K

	Bicubic	VDSR[4]	SRGAN[5]	EDSR[6]	EESR
PSNR	22.2467	23.5969	23.4090	23.7904	22.3501
SSIM	0.7414	0.8102	0.8082	0.8288	0.7588
PTS	0.0767	0.1870	0.2383	0.2287	0.6210



4 times super resolution result of our EESR compared with existing algorithms.

SUMMARY

- A new blind SR method, called EESR, is proposed.
- Edge pixels are focused by introducing an edge detection model.
- GAN is used to fine-tune our network
- EESR reaches an outstanding performance for 4 times up-sampling in preserving high-frequency structures and generating high quality SR images.

REFERENCE

[1] Ian J. Goodfellow, Jean Pouget-Abadiey, Mehdi Mirza *et al.* Conference and Workshop on Neural Information Processing Systems (NIPS), pp 2672-2680 (2014).

[2] Saining Xie, Zhuowen Tu. IEEE International Conference on Computer Vision (ICCV), pp 1395-1403 (2015).

[3] K. Simonyan and A. Zisserman. International Conference on Learning Representations (2015).

[4] Jiwon Kim, Jung Kwon Lee and Kyoung Mu Lee. Computer Vision and Pattern Recognition (CVPR), pp 1063-6919 (2016).

[5] Christian Ledig, Lucas Theis, Ferenc Husz´ar *et al.* The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4681-4690 (2017).

[6] Bee Lim, Sanghyun Son, Heewon Kim *et al.* CVPR workshop (2017).