NTIRE 2020 Perceptual Extreme Super-Resolution Challenge Factsheet
A Fast Feedback Network for Large Scale Image Super-Resolution

Zhi Jin, Jiahao Wu, Yifu Chen, Chenming Shang, Huanrong Zhang

1. Team details (required)

- Team name
  sysu-AIR
- Team leader name
  Zhi Jin
- Team leader address, phone number, and email
  University Town, Guangzhou, Guangdong, P.R. China;
  +86 13692244701;
  jinzh26@mail.sysu.edu.cn; jinzhi_126@163.com.
- Rest of the team members
  Jiahao Wu, Yifu Chen, Chenming Shang, Huanrong Zhang.
- Affiliation
  School of Intelligent Systems Engineering, Sun Yat-sen University.
- Affiliation of the team and/or team members with NTIRE2020 sponsors (check the workshop website)
  Non-applicable.
- User names and entries on the NTIRE2020 Codalab competitions (development/validation and testing phases)
  Zhi_Jin_SYSU; 2 entries.
- Link to the codes/executables of the solution
  https://github.com/jzrita/NTIRE2020_sysu-AIR.
- Link to the restoration results of all frames
  https://github.com/jzrita/NTIRE2020_sysu-AIR.

2. Details for the final report paper (required)

Please provide necessary information for the final report paper "**NTIRE 2020 Challenge on Perceptual Extreme Super-Resolution: Methods and Results**".

Your team name

Title: A Fast Feedback Network for Large Scale Image Super-Resolution
Members: *Zhi Jin (jinzh26@mail.sysu.edu.cn)*, Jiahao Wu, Yifu Chen, Chenming Shang, Huanrong Zhang
Affiliation:
School of Intelligent Systems Engineering, Sun Yat-sen University.

Results

(*Please complete Table 1.*)

Your team name

The **sysu-AIR** team proposed **A Fast Feedback Network for Large Scale Image Super-Resolution**. Inspired by SRFBN [4] and IMDN [2], the proposed Fast-SRFBN is still reserved the RNN structure but with a information multi-distillation module (IMDM), which can benefit image SR tasks and accelerate inference speed. As shown in Figure 1, the IMDM recurrently refines the LR image in a "coarse to fine" manner, and it consists of a 1×1 convolutional layer and several stacked information multi-distillation blocks (IMDB) [2]. The IMDB is constructed by progressive refinement module, contrast-aware channel attention (CCA) layer, and a 1×1 convolutional layer. This kind of design enables IMDB to extract features at a granular level, which retains partial information and further treats other features at each step (layer) as illustrated in top part of Figure 1. For aggregating features distilled by all steps, a contrast-aware channel attention layer, specifically related to the low-level vision tasks, is devised to enhance collected various refined information.

As the hidden state in an unfolded RNN, the output of the IMDM is used to achieve the feedback structure. The hidden state at each iteration flows into the next iteration to modulate the input. Hence, one IMDM at the t-th iteration receives the feedback information F_{out}^{t-1} and the low-level representations F_{in}^{t}, which remain the same through all T iterations. Then, it outputs F_{out}^{t} to the next iteration and the following convolutional layer. Cooperating with the such

Table 1. NTIRE 2020 perceptual extreme super-resolution results and final rankings on the DIV8K test set.

Team	Author	Number of parameters	Run-time per testing image	Platforms PyTorch/TensorFlow	Ensemble None/model/self	GPU Xp/2080Ti	Extra training datasets (e.g., DIV2K)
sysu-AIR	Zhi Jin, Jiahao Wu, Yifu Chen, Chenming Shang, Huanrong Zhang	2,099,625	4.35s	PyTorch	None	GeForce RTX 2080Ti	None

IMDM, the RNN structure can extract deep semantic information and still retain low-level representations to refine the feature map iteration by iteration. After T iterations, they finally get totally T SR images, i.e., I^1_{SR}, I^2_{SR}, , I^T_{SR}. Only I^T_{SR} is accepted as the final result, but all T SR images are involved in loss computation.

The network is trained in a generative adversarial network (GAN) way to enhance the visual quality, which introduces the adversarial loss [1]. They also adopt multiple losses in multiple domains to improve the SR performance: pixel loss, perceptual loss [3], novelly proposed Fourier spectrum loss, and total variation loss [5]. The pixel loss computes the distance between SR image and its corresponding ground truth in the spatial domain, while perceptual loss computes that in the feature domain provided by the pre-trained vgg19 [6]. The Fourier spectrum loss is firstly proposed in this work, and it computes the similarity between HR image and SR image in the frequency domain. Since Zhang et.al [9] have pointed out that GAN generated images suffer from visible periodic artifacts in frequency domain, inspired by [9], in this work they propose Fourier spectrum loss to reduce this artifact. The settings of adversarial loss, pixel loss, and perceptual loss follow ESRGAN [8]. Besides, they introduce total variation loss from image restoration tasks.

3. Method Description

Our solution is mainly based on the recurrent neural network (RNN) framework of SRFBN [4] and integrates many existing technologies for further improvement. Specifically, we first replace the recurrent block (RB) of SRFBN, due to its original RB is very time-consuming and leads a long inferring time during training. The new RB solution is the information multi-distillation module (IMDM), which brings a significant acceleration of inference. Within the IMDM, we adopt several stacked information multi-distillation blocks (IMDBs), which is proposed by [2]. Moreover, we adopt multiple losses in multiple domains to improve the reconstruction performance: pixel loss (spatial domain), perceptual loss (feature domain) [3], adversarial loss (feature domain) [1], Fourier spectrum loss (frequency domain), and total variation loss (spatial domain). Inspired by [7] and [10], we are the first to propose Fourier spectrum loss to penalize the distance between reconstruction image and its corresponding ground truth in the frequency domain, to deal with the fake artifacts caused by GAN [1]. We compute the perceptual loss based on the features from pre-trained vgg19 [6], and the Fourier spectrum loss is computed by 2D discrete Fourier transform.

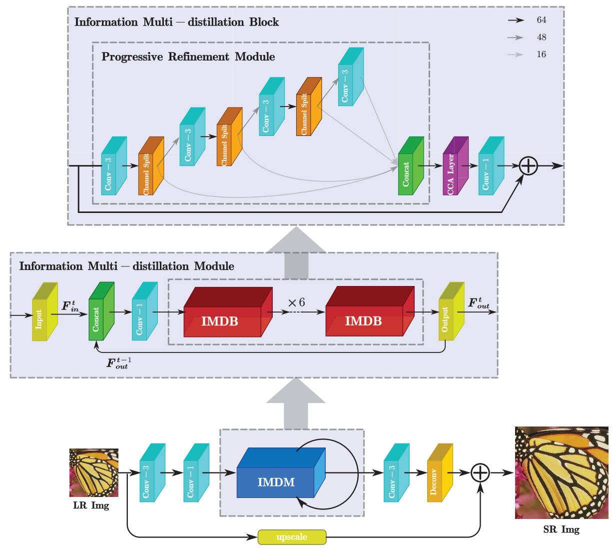

Figure 1. The overall architecture of sysu-AIR team.

4. Ensembles and fusion strategies

- Describe the ensemble strategies (if any).
- What was the benefit over the single method?

Our solution does not utilize any ensembles or fusion strategies.

5. Other details

- Planned submission paper at NTIRE2020 workshop.
- General comments and impressions of the NTIRE2020 challenge.
- What do you expect from a new challenge in image restoration, enhancement and manipulation?

References

[1] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in Neural Information Processing Systems*, pages 2672–2680, 2014. 2

[2] Zheng Hui, Xinbo Gao, Yunchu Yang, and Xiumei Wang. Lightweight image super-resolution with information multi-distillation network. In *Proceedings of the 27th ACM International Conference on Multimedia*, pages 2024–2032, 2019. 1, 2

[3] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *Proceedings of the European Conference on Computer Vision*, pages 694–711, 2016. 2

[4] Zhen Li, Jinglei Yang, Zheng Liu, Xiaomin Yang, Gwanggil Jeon, and Wei Wu. Feedback network for image super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3867–3876, 2019. 1, 2

[5] Leonid I. Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1):259 – 268, 1992. 2

[6] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. 2

[7] Sheng-Yu Wang, Oliver Wang, Richard Zhang, Andrew Owens, and Alexei A Efros. Cnn-generated images are surprisingly easy to spot... for now. *arXiv preprint arXiv:1912.11035*, 2019. 2

[8] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European Conference on Computer Vision Workshops*, pages 0–0, 2018. 2

[9] X. Zhang, S. Karaman, and S. Chang. Detecting and simulating artifacts in gan fake images. In *2019 IEEE International Workshop on Information Forensics and Security (WIFS)*, pages 1–6, Dec 2019. 2

[10] Xu Zhang, Svebor Karaman, and Shih-Fu Chang. Detecting and simulating artifacts in gan fake images. *arXiv preprint arXiv:1907.06515*, 2019. 2