NTIRE 2020 Perceptual Extreme Super-Resolution Challenge Factsheet: Investigating Loss Functions for Extreme Super-Resolution

Younghyun Jo, Sejong Yang, and Seon Joo Kim

1. Team details

- Team name: CIPLAB
- Team leader: Younghyun Jo
- Address: Room 707, Engineering hall 4, Yonsei university, 50 Yonsei-ro, Seodaemun-gu, Seoul 03722, Korea
- Phone number: +82-1096097097
- Email: yh.jo@yonsei.ac.kr
- Team members: Sejong Yang, Seon Joo Kim
- Affiliation: Yonsei university
- No one is involved with NTIRE2020 sponsors
- User name on Codalab: heyday097
- (Best) Entries on Codalab: Please check our last submissions for each development and test phase.
- Codes: https://github.com/kingsj0405/ciplab-NTIRE-2020
- Full results: https://drive.google.com/file/d/1kmiBM_jfhfWcxXTJB17MvHW_9XM0sbZe/view?usp=sharing
- Fix results: We found that 1000x1000 cropped images for 1602, 1612, 1648, 1677 are included in the above link. Please overwrite them by downloading from this link: https://drive.google.com/file/d/1rbOj_HfNndxuFrXJ-x3gGD2Yqm7o60ve/view?usp=sharing

2. Details for the final report paper

CIPLAB

Title: Investigating Loss Functions for Extreme Super-Resolution

Members: *Younghyun Jo*[1] (*yh.jo@yonsei.ac.kr*), Sejong Yang[1], Seon Joo Kim[1]

Affiliation:

[1] Yonsei University

Results

(*Please complete Table 1.*)

CIPLAB

We propose to use GAN [1] with LPIPS [5] loss for extreme super-resolution (SR) instead of using GAN with VGG perceptual loss [2]. LPIPS is trained with a dataset of human perceptual similarity judgments, and we expect it is more proper choice for perceptual SR. In addition, we use U-net structure discriminator [3] to consider both local and global context of an image. Our loss combination shows better results in Fig. 3.

3. Method Description

The performance of image super-resolution (SR) has been greatly improved by using convolutional neural networks. However, most of the methods have been studied up to x4 upsampling, and few studies were studied for x16 upsampling. There are three aspects to consider for the new x16 upsampling method: the first is datasets, the second is network designs, and the last is loss functions. Here, we focus on investigating new loss functions for the perceptual x16 SR.

Loss functions: Adversarial loss + Feature matching loss + LPIPS loss + MSE loss General choice of the loss functions for perceptual SR is the adversarial loss [1] with the VGG perceptual loss [2]. This loss combination has worked well for x4 SR, however, we empirically found that it is not work well for x16 SR due to highly hallucinated noise and less precise details (Fig. 3). Because VGG network is trained for image classification, it may not the best choice for the SR task. To this end, we use the learned perceptual similarity (LPIPS) proposed in [5] instead of the VGG perceptual loss. LPIPS is trained with a dataset of human perceptual similarity judgments, and we expect it is more proper choice for perceptual SR. In addition, the discriminator's feature matching loss helps to increase the quality of the results, and mean square error (MSE) on the pixel space prevents color permutation.

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	Ensamble None/model/self	GPU Xp/2080Ti	Extra training datasets (e.g., DIV2K)
CIPLAB	heyday097	33M	3.00	PyTorch	None	TITAN XP	N/A

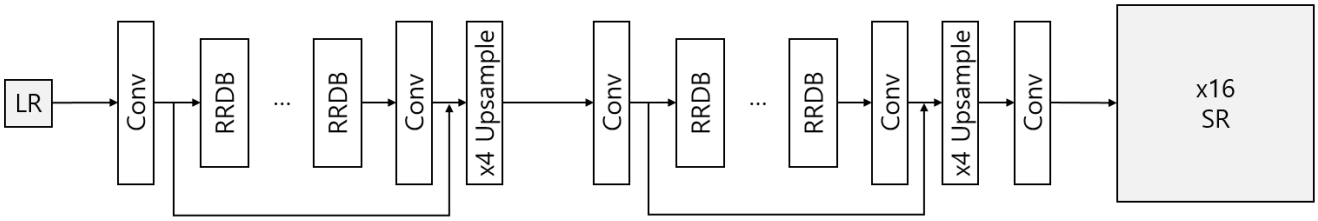

Figure 1: Our generator structure for x16 SR. There are total 46(23+23) RRDBs, and please refer [4] for RRDB details.

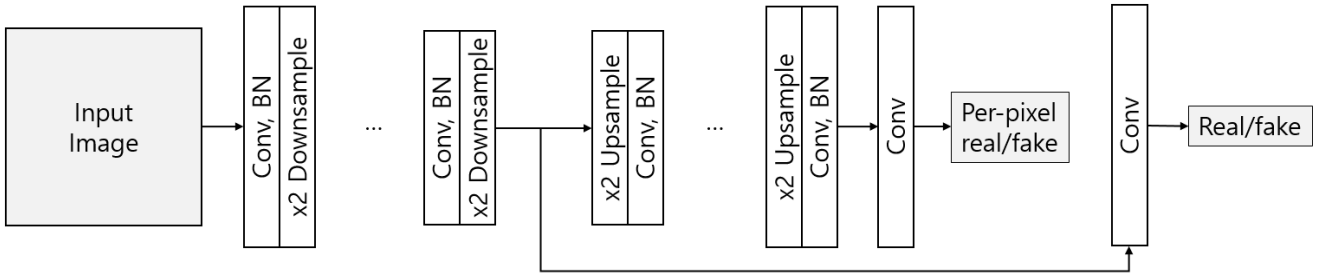

Figure 2: Our discriminator network. There are 6 downsampling and 6 upsampling stages.

Generator: Adopt x4 SOTA one and double the parameters There are few studies for the network structures for x16 SR. We adopt the generator of ESRGAN [4] as our generator network. ESRGAN is the winner of PIRM 2018 challenge on perceptual super-resolution, and it is currently one of state-of-the-art (SOTA) methods for x4 perceptual SR. We double the main network body for x16 SR as shown in Fig. 1.

Discriminator: U-net structure We adopt an U-net discriminator structure [3] for our discriminator (Fig. 2). The discriminator judges real and fake for the compressed space from the encoder head and every pixel from the decoder head. This allows to provide detailed per-pixel feedback to the generator while maintaining the global context. We empirically found that this gives more details for restored images rather than normal encoder structure discriminator.

Implementation details For training, we use provided training data only. We randomly crop 384x384 patches from the training data, and corresponding input patch size is 24x24. For the both networks, we use Adam optimizer and learning rate is set to 0.00001. We first train our generator with mean square error for 50K iterations with mini-batch size 3 (takes 12 hours), and we further train the generator with our discriminator with the proposed loss setting for about 60K iterations with mini-batch size 2 (takes 15 hours).

We use Pytorch 1.2.0 and single NVIDIA TITAN XP GPU (12G). The number of parameters for our generator and discriminator is 33M and 13M respectively. Our generator takes average 3.00 seconds for given test images.

Results Some results are shown in Fig. 3. Our results look better than GAN + VGG results qualitatively. This is the effect of using LPIPS loss as it provides better feature space than VGG for human perception. This is also the effect of U-net discriminator as it considers global context, and the discriminator give effective feedback to the generator.

4. Ensembles and fusion strategies

Our results are from a single model. We expect model ensembles can further improve the results.

5. Other details

Thanks for always opening up many interesting competitions. We hope that interesting competitions will be held in areas where not much research has been done such as this year's x16 SR challenge.

Figure 3: The left images are generated by our method, and the right images are generated from the model trained by the adversarial loss with the VGG perceptual loss. In the right images, it sometimes generates excessive noise or degenerates details. Please zoom for better comparisons. From top to bottom: *1659.png*, *1680.png*, and *1601.png*.

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 *NeurIPS*, pages 2672–2680, 2014. 1

[2] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *ECCV*, 2016. 1

[3] Edgar Schönfeld, Bernt Schiele, and Anna Khoreva. A u-net based discriminator for generative adversarial networks. *arXiv preprint arXiv:2002.12655*, 2020. 1, 2

[4] 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 *ECCVW*, September 2018. 2

[5] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*, pages 586–595, 2018. 1