NTIRE2020 Real World Super-Resolution Challenge Factsheet: Investigating Loss Functions for Extreme Super-Resolution

Younghyun Jo, Sejong Yang, and Seon Joo Kim

March 26, 2020

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
- Rest of the team members: Sejong Yang, Seon Joo Kim
- Team website: https://github.com/kingsj0405/ciplab-NTIRE-2020
- Affiliation: Yonsei university
- No one is involved with NTIRE2020 sponsors
- User name on Codalab: heyday097
- (Best) Entries on Codalab: We have only one submission for each development/test phase.
- Link to the codes: https://github.com/kingsj0405/ciplab-NTIRE-2020
- Link to the restoration results of all frames: https://drive.google.com/file/d/1kmiBM_jfhfWcxXTJB17MvHW_9XM0sbZe/view?usp=sharing

2 Contribution details

- Title: Investigating Loss Functions for Extreme Super-Resolution

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.

- 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.

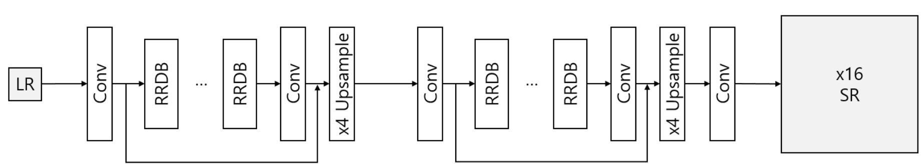

Figure 1: Our generator structure for x16 SR. There are total 46(23+23) RRDBs, and please refer [4] for RRDB details.

- 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.

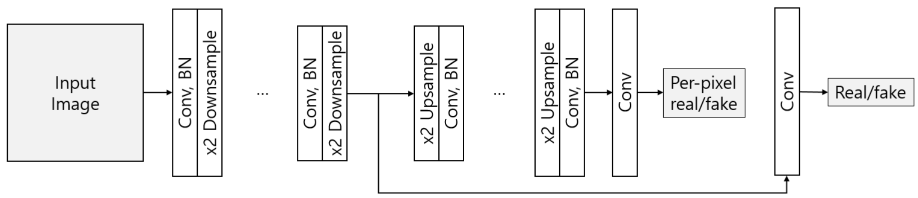

Figure 2: Our discriminator network. There are 6 downsampling and 6 upsampling stages.

3 Global Method Description

- The number of parameters:
  - Generator: 33M
  - Discriminator: 13M
- Additional training data: No
- Training description: 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).
- Results of the comparison to other approaches: Please see Fig. 3.
- Results on other benchmarks: N/A
- Quantitative and qualitative advantages of the proposed solution: It is hard to assess the results quantitatively as we aim for perceptual SR. As can be seen in Fig. 3, our results look better than the opponent's results qualitatively. This is the effect of U-net discriminator as it considers global context, and the discriminator give effective feedback to the generator.

4 Competition particularities

Note that we only apply for Track 1.

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*.

5 Ensembles and fusion strategies

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

6 Technical details

- Language and implementation details:
  - Single NVIDIA TITAN XP GPU (12G)
  - Python 3.6
  - Pytorch 1.2.0
- Testing time per image: Average 3.00 seconds for given test images.
- Generality of the proposed solution: Our proposed loss setting and encoder-decoder structure discriminator can be applied for any generators in GAN-like framework.
- For other sets of downscaling operators: We should retrain our generator for other sets of downscaling operators. However, using the parameters of the existing trained model as the initial value is expected to reduce the training time than random initialization.

7 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.

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.

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

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

[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.

[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.