Data-Free Knowledge Distillation For Image Super-Resolution
Supplementary Material

Yiman Zhang[1], Hanting Chen[1,3], Xinghao Chen[1], Yiping Deng[2], Chunjing Xu[1], Yunhe Wang[1*]
[1] Noah's Ark Lab, Huawei Technologies. [2] Central Software Institution, Huawei Technologies.
[3] Key Lab of Machine Perception (MOE), Dept. of Machine Intelligence, Peking University.
{yiman.zhang, chenhanting, yunhe.wang}@huawei.com

Table 1: The architecture of generator. The **image size** and **channel** in **Initialization** indicate the size and number of channel of input images respectively.

Initialization	init = image size // 4, channel.
Layer 1: Linear	Input: 256
	Weight shape: (256, 128×init×init)
	Output: 128×init×init
Layer 2: Reshape	Input: 128×init×init
	Output: (128, init, init)
Layer 3: Instance Norm	Affine=True
Layer 4: Interpolate	scale=2, mode=nearest
Layer 5: Convolution	Input: (128, init×2, init×2)
	Kernel Size: 128×128×3×3, stride=1, padding=1
	Output:(128, init×2, init×2)
Layer 6: Instance Norm	Eps=0.8, Affine=True
Layer 7: Activation	LeakyReLU, negative slope=0.2
Layer 8: Interpolate	scale=2, mode=nearest
Layer 9: Convolution	Input: (128, image size, image size)
	Kernel Size: 128×64×3×3, stride=1, padding=1
	Output:(64, image size, image size)
Layer 10: Instance Norm	Eps=0.8, Affine=True
Layer 11: Activation	LeakyReLU, negative slope=0.2
Layer 12: Convolution	Input: (64, image size, image size)
	Kernel Size: 64×channel×3×3, stride=1, padding=1
	Output:(channel, image size, image size)
Layer 13: Activation	Tanh

1. The Architecture of Generator

Table 1 shows the details of generator used in our experiments. Considering that the data formats of VDSR [1] and EDSR [2] are different, we set image size and channel as variables. For EDSR, image size is set to 48 and channel is set to 3. And for VDSR, channel is set to 1 and image size is calculated by $48//scale$, while scale is in $2, 3, 4$.

2. More Visualization Results on EDSR

In Figure 1 and 2 we show more visualization results of EDSR. Our distillation method obtains better super resolution results than bicubic and distillation with noise images. It's worth mentioning that Teacher performs worse for the image 202003 in B100, while performing better than Student on whole dataset. Our proposed method even obtains a better visual result than Teacher for this special case.

References

[1] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1646–1654, 2016.

[2] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 136–144, 2017.

Figure 1: $\times 4$ super resolution results of 202003 from B100 and img021 from Urban100 on EDSR.

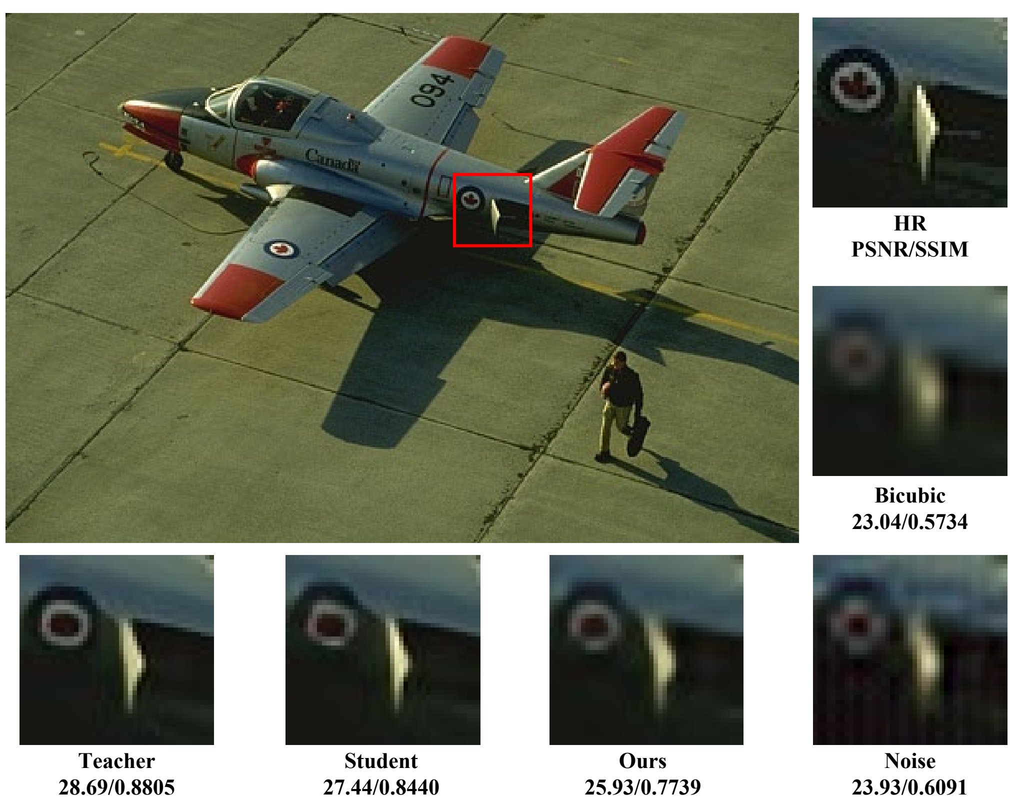

Figure 2: $\times 4$ super resolution results of 37073 from B100 on EDSR.