1. Comparison with Naïve Cross-Scale Non-Local (CS-NL) Attention

In the non-local structure, features are summed and weighted by corresponding spatial attention. Formally, in-scale non-local attention is

\[
Z_{i,j} = \frac{\sum_{g,h} \exp(\phi(X_{i,j}, X_{g,h})) \psi(X_{g,h})}{\sum_{u,v} \exp(\phi(X_{i,j}, X_{u,v}))} \psi(Y_{g,h})
\]

(1)

where red and blue are the same features representation. Naïve cross-scale non-local attention can be straightforwardly evolved as

\[
Z_{i,j} = \frac{\sum_{g,h} \exp(\phi(X_{i,j}, Y_{g,h})) \psi(Y_{g,h})}{\sum_{u,v} \exp(\phi(X_{i,j}, Y_{u,v}))} \psi(Y_{g,h})
\]

(2)

where red and blue are still the same but changed to \(Y = X \downarrow_s\), that are the down-scaled features by by scaling factor \(s\). The naïve cross-scale attention is based on the correlation between features in different scales but summarises down-scaled features. The down-scaling operation will eliminate high-frequency details and lead performance regression in super-resolution tasks.

The proposed cross-scale non-local attention summaries corresponding features in target scale without down-scaling operation, and can be formalized as

\[
Z_{s \times s}^{i \times j} = \frac{\sum_{g,h} \exp(\phi(X_{i,j}, Y_{g,h})) \psi(X_{s \times s \downarrow g,h})}{\sum_{u,v} \exp(\phi(X_{i,j}, Y_{u,v})) \psi(X_{s \times s \downarrow u,v})}
\]

(3)

where red and blue are in different scales but one-to-one corresponded spatially. In this way, the proposed cross-scale attention can keep high-resolution information in feature maps, utilize the original self-exemplar hints and benefits super-resolution performance.

Experiments in Table 1 shows that the naïve cross-scale attention is negligible better than in-scale one, and the proposed cross-scale attention significantly outperforms other approaches.

|       | Proposed Cross-scale | Naïve Cross-scale | In-scale |
|-------|----------------------|-------------------|----------|
| PSNR  | 33.74                | 33.65             | 33.62    |

Table 1. Comparison with Naïve Cross-Scale Non-Local (CS-NL) Attention on Set14 [9] (×2).

2. More Qualitative Comparison

In Fig. 1-2, we provide more visual results to compare with other state-of-the-art methods. One can see that our approach reconstructed better image details, demonstrating the superiority of the proposed CSNLN.

References

[1] Tao Dai, Jianrui Cai, Yongbing Zhang, Shu-Tao Xia, and Lei Zhang. Second-order attention network for single image super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 11065–11074, 2019.

[2] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. In European conference on computer vision, pages 184–199. Springer, 2014.

[3] Muhammad Haris, Gregory Shakhnarovich, and Norimichi Ukita. Deep back-projection networks for super-resolution. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 1664–1673, 2018.

[4] Xiangyu He, Zitao Mo, Peisong Wang, Yang Liu, Mingyuan Yang, and Jian Cheng. Ode-inspired network design for single image super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 1732–1741, 2019.

[5] 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.

[6] Wei-Sheng Lai, Jia-Bin Huang, Narendra Ahuja, and Ming-Hsuan Yang. Deep laplacian pyramid networks for fast and
accurate super-resolution. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 624–632, 2017. 2, 3

[7] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyung Mu Lee. Enhanced deep residual networks for single image super-resolution. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 624–632, 2017. 2, 3
Figure 2. Visual comparison for $4 \times$ SR on Manga109 dataset.

ence on computer vision and pattern recognition workshops, pages 136–144, 2017. 2, 3

[8] Ying Tai, Jian Yang, and Xiaoming Liu. Image super-resolution via deep recursive residual network. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 3147–3155, 2017. 3
[9] Roman Zeyde, Michael Elad, and Matan Protter. On single image scale-up using sparse-representations. In International conference on curves and surfaces, pages 711–730. Springer, 2010. 1

[10] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In Proceedings of the European Conference on Computer Vision (ECCV), pages 286–301, 2018. 2

[11] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual dense network for image super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 2472–2481, 2018. 2, 3