Supplementary Material
NAS-Count: Counting-by-Density with Neural Architecture Search

Yutao Hu[1] *, Xiaolong Jiang[4] *, Xuhui Liu[1], Baochang Zhang[5], Jungong Han[6], Xianbin Cao[1,2,3] †, and David Doermann[7]

[1] School of Electronic and Information Engineering, Beihang University, Beijing, China
[2] Key Laboratory of Advanced Technologies for Near Space Information Systems, Ministry of Industry and Information Technology of China
[3] Beijing Advanced Innovation Center for Big Data-Based Precision Medicine,China
[4] YouKu Cognitive and Intelligent Lab, Alibaba Group
[5] Beihang University, Beijing, China
[6] Computer Science Department, Aberystwyth University, SY23 3FL, UK
[7] Department of Computer Science and Engineering, University at Buffalo, New York, USA
{huyutao, bczhang, xbcao}@buaa.edu.cn, xainglu.jxl@alibaba-inc.com, xuhui_cc@126.com, jungonghan77@gmail.com, doermann@buffalo.edu

Comparison with various decoder

In the Sec.4.3 of the main paper, we compare the searched AMSNet decoder with single-path decoder. In fact, in our two-level search, we leverage the macro-level search to explore an optimal multi-path encoder-decoder architecture, which helps us achieve the sufficient feature aggregation. To fully elaborate the efficacy of our macro-level search, we conduct experiments on the ShanghaiTech Part_A dataset to compare our AMSNet decoder with various baseline decoder. Specifically, based on the decoding block and hierarchical fusion strategy in [1], we fuse multi-scale features from the last three (two) cells. The experimental results are shown in Table 1. The first row shows the counting accuracy of single-path decoder. The second and third row show the performance of trellis multi-path decoder that fuses features from the last two cells and last three cells respectively. The fourth row reports the performance of our AMSNet decoder that fusing features from last three cells. Additionally, the four networks are all equipped with AMSNet encoder and optimized with SPPLoss. We can find multi-path decoder surpasses the single-path decoder, which coincides with the idea that fusing multi-scale features from different stages can improve the performance. Furthermore, the decoder in our AMSNet also achieves a better performance than trellis multi-path decoder, which not only demonstrates its superior ability in multi-scale feature aggregation but also shows the great effectiveness of the macro-level search in obtaining an efficient multi-path encoder-decoder network.

*Contribute equally
†Corresponding author

Table 1: Performance comparison among methods with different decoder configurations on the ShanghaiTech Part_A.

Network	MAE↓	MSE↓
Single-Path Decoder	59.5	98.3
Trellis Multi-Path Decoder(two cells)	58.5	96.2
Trellis Multi-Path Decoder(three cells)	57.9	95.8
AMSNet Decoder	**56.7**	**93.4**

References

1. Jiang, X., Xiao, Z., Zhang, B., Zhen, X., Cao, X., Doermann, D., Shao, L.: Crowd counting and density estimation by trellis encoder-decoder networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2019) 6133–6142