Discrete-Constrained Regression for Local Counting Models —Supplementary Materials

Haipeng Xiong[0000−0002−8858−3807] and Angela Yao[0000−0001−7418−6141]

National University of Singapore, Singapore
{haipeng,ayao}@comp.nus.edu.sg

1 Architecture of Local Count Networks

For our local count models, we adopted all the convolutional layers in VGG16 [2] to extract feature maps, then we used a regression head consisting of two 3×3 convolutional layers (512 and 1 output channels, respectively) to map local features to local counts, as shown in Fig. 1. The size of the local patch is 32×32.

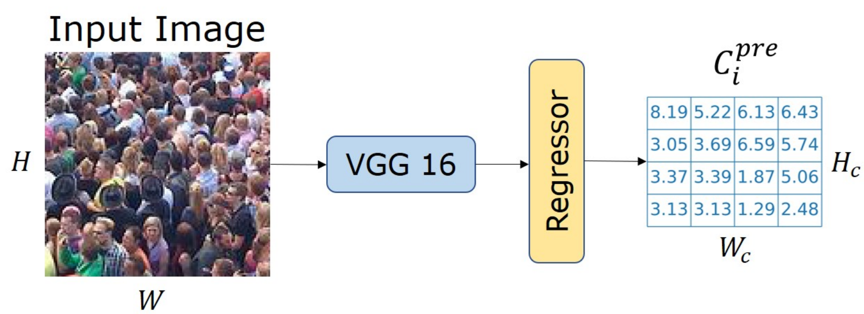

Fig. 1. Architecture of local count models. "VGG16" denotes all of the convolutional layers in VGG16 [2], "regressor" is consisted of two 3×3 convolutional layers with 512 and 1 output channels, respectively. In this example, $H = W = 128$ and $H_c = W_c = 4$.

2 Additional Experiment on Real-World Datasets

2.1 Analysis of Global Count Loss L_{gc}

Fig. 2 presents an example of error matrix E of local count, S^a and S^m. Since the error of the image is 3.8 (> 0), S^a selects the the local image patches with $E(j,k) > 0$. S^m further selects the patches with error 1.50 and 2.3, the sum of which is equal to the global error 3.8.

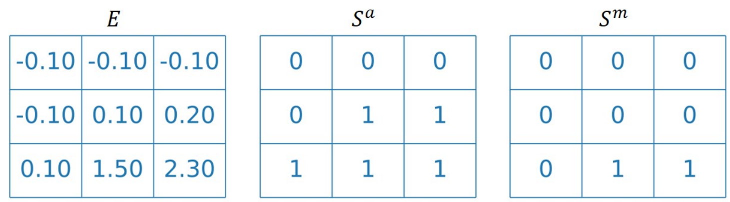

Fig. 2. An example of S_i^a of L_{bias}^0 and S_i^m of L_{bias}^λ. E denotes the error of local counts.

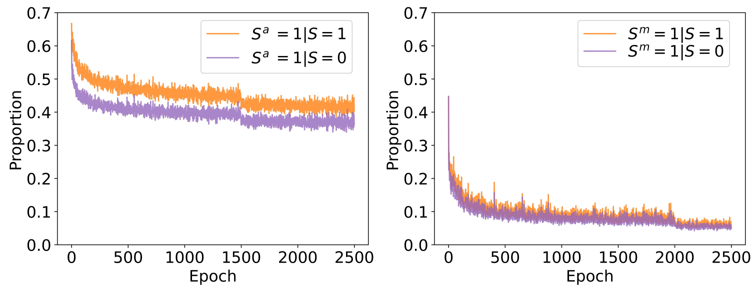

Fig. 3. Visualization of S^a of L_{bias}^0 (left) and S^m of L_{bias}^λ (right) during training. $S = 0/1$ denotes the sample with prediction inside/outside the interval range. $S^m = k|S = j$ denotes the proportion of samples among $S = j$ which satisfies $S^m = k$.

We further compare S^a, S^m during the training phase in Fig. 3. At the beginning of training phase, L_{bias}^0 considers nearly half of the local counts C_i^{pre} within the class intervals, which harms the discrete constraints; while L_{bias}^λ considers a small portion of the samples predicted within the class intervals, which mainly contribute to the error of global counts. In this way, L_{bias}^λ does not harm the discrete regression loss L_{dc} during training, and is helpful to reduce discretization error when most samples are predicted within class intervals at late epochs.

3 DC-regression With Various Backbones

In the paper, we adopt VGG16 [2] as backbone for dc-regression for fair comparison with other methods. Here we evaluate dc-regression with more backbones, including SWIN [1] and efficient network [4].

Table 1. Comparison Different Network Backbone on JHU dataset [3].

Backbone	MAE	MSE
VGG16 [2]	64.8	282.6
SWIN-T [1]	62.2	242.2
SWIN-L [1]	61.5	259.1
effnet-b4 [4]	65.5	251.1

References

1. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B.: Swin transformer: Hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 10012–10022 (2021)
2. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. Computer Science (2014)
3. Sindagi, V.A., Yasarla, R., Patel, V.M.: Jhu-crowd++: Large-scale crowd counting dataset and a benchmark method. Technical Report (2020)
4. Tan, M., Le, Q.: Efficientnet: Rethinking model scaling for convolutional neural networks. In: International conference on machine learning. pp. 6105–6114. PMLR (2019)