Supplementary Material of 'Crowd Counting With Partial Annotations in an Image'

Yanyu Xu[1], Ziming Zhong[2], Dongze Lian[2], Jing Li[2], Zhengxin Li[2], Xinxing Xu[1]
Shenghua Gao[2,3,4]

[1]IHPC, A*STAR, Singapore. {xu_yanyu,xuxinx}@ihpc.a-star.edu.sg
[2]ShanghaiTech University, China. {zhongzm,liandz,lijing1,lizhx, gaoshh}@shanghaitech.edu.cn
[3]Shanghai Engineering Research Center of Intelligent Vision and Imaging, China.
[4]Shanghai Engineering Research Center of Energy Efficient and Custom AI IC, China.

The organization of our supplementary material is mainly about the the visualization and implementation details of the latent vectors.

1. The effort of Learned VQ Latent Vector

To investigate the learned latent vectors in the memory bank, we show the corresponding relationship between the indices of latent vectors and the crowd counting numbers, as shown in 1. After finishing the training process, given the learned memory bank E, we feed one partial annotated training image into the trained model and get the features $F_0 \in R^{128 \times H/8 \times W/8}$. For each feature vector f_j in annotated regions, we use the equation (2) to find its most similar latent vector e_i and the corresponding vector index i. Thus, we can get an indices map $(H/8 \times W/8)$, where only the annotated regions have values. Since the image and ground truth density map is 8 times larger than feature F_0, each index in indices map is related to the crowd count number of its corresponding 8×8 receptive field region in density map. We repeat the above process on the whole training images and average the counting number of each index. The final indices and counting number table are shown in Fig. 1 (a). Fig. 1 (d) and (g) show the generated density maps based the table (a). We can see that there is a high correlation and similarity between density maps generated by latent vectors and ground truth.
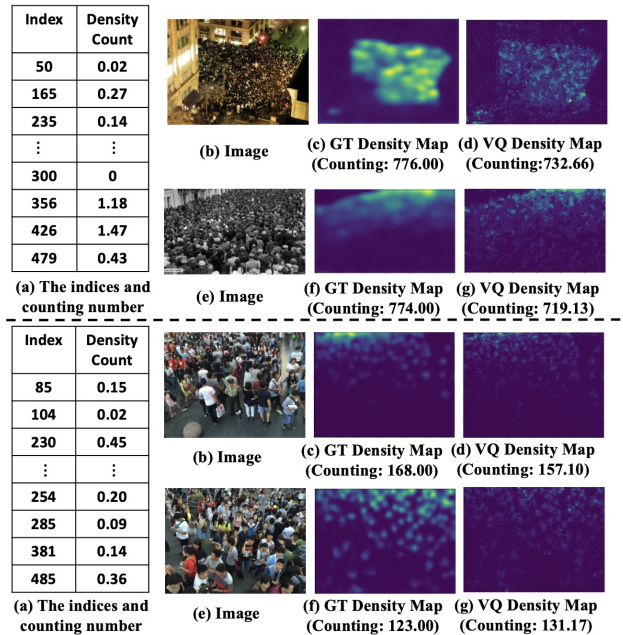
References

Figure 1. The corresponding relationship between the indices of latent vectors and the crowd counting numbers