Calibrating Uncertainty for Semi-Supervised Crowd Counting
—Supplementary Material—

Chen Li [*] Xiaoling Hu Shahira Abousamra Chao Chen
Stony Brook University

A. Appendix

Append. A.1 shows quality results about which high uncertain patches are filtered out.

Append. A.2 shows that our method is reliable even with only 5% labeled samples.

Append. A.3 provides more ablation study results.

Append. A.4 illustrates the implementation details of choosing pseudo-labels.

Method	Ratio	Part A	
		MAE	RMSE
sup.only	5%	88.48	162.42
w/o filtering	5%	111.39	174.87
softmax	5%	94.64	155.65
Ours	5%	**74.48**	**127.51**

Table 9: The ablation study results of 5% labeled data.

A.1. Illustrate which pseudo-labels were selected and which inaccurate ones were filtered out.

Fig. 6 shows results on unlabeled samples. Green dots are predictions and red rectangles are high-uncertainty patches. On the left, we found a high-uncertainty patch within the sparse region, containing only one false positive (on the traffic light). In the middle and right samples, the high-uncertainty patches contain many false negatives due to occlusion or dark shades.

A.2. Is 5% labeled data enough for training reliable uncertainty estimator?

Empirical results show that 5% labeled data is sufficient to achieve superior performance on ShanghaiTech A (main paper Tab. 5 & Tab. 9) and B (Tab. 10) datasets.

A.3. Extra ablation study results.

In this section, we show the effectiveness of our method under 5% and 40% labeled images on the ShanghaiTech part-B dataset with extra ablation study experiments. As shown in Tab. 10, our method achieves better performance

Figure 6: Qualitative results on uncertainty filtering.

under both 5% and 40% labeled image scenarios. This indicates our method can obtain superior performance for semi-supervised crowd counting under various labeled ratios on different datasets.

In Fig. 7, we show additional ablation study results.

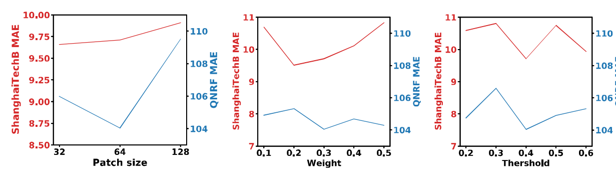

Figure 7: The hyperparameter ablation study results on ShanghaiTech B and UCF-QNRF.

A.4. Details of pseudo-labeling

Here we show the details of linearly increasing uncertainty threshold u_t for choosing pseudo-labels:

$$u_t = startunc + \frac{endunc - startunc}{endep - startep}(t - startep)$$

where u_t is the uncertainty threshold for choosing image patches, i.e., the image patches with uncertainty estimation higher than u_t are blanked out, and t is the current epoch number. The increase of u_t begins at epoch $startep$ and ends at epoch $endep$. The uncertainty threshold increases from $startunc$ to $endunc$. By using this strategy, we can utilize high-quality model predictions at different training stages properly.

Since it takes several training iterations for multitask model to capture valid crowd and uncertainty information,

[*]Email: Chen Li (li.chen.8@stonybrook.edu).

we start leveraging unlabeled information from $10th$ epoch i.e. $startep = 10$. The model predictions on unlabeled images are error-prone, thus uncertainty threshold at the beginning $startunc$ is 0.1. Besides, we have $endep = 130$ and $endunc = 0.6$.

Method	Type	Ratio	Part B	
			MAE	RMSE
MT [4]	SSL	5%	19.3	33.2
L2R [2]	SSL	5%	20.3	27.6
GP [3]	SSL	5%	15.7	27.9
PA [5]	PAL	5%	16.50	25.28
DAcount [1]	SSL	5%	12.6	22.8
ours	SSL	5%	**11.03**	**20.93**
MT [4]	SSL	40%	15.9	25.7
L2R [2]	SSL	40%	16.8	25.1
DAcount [1]	SSL	40%	9.6	14.6
Ours	SSL	40%	**7.79**	**12.70**

Table 10: The ablation study results of labeled ratio on the ShanghaiTech part-B dataset.

A.5. Implementation details

In practice, for the convenience of implementation, we use (1 - batch normalized ASM) as a surrogate to train the uncertainty branch, and the model confidence output will be used to filter out unreliable patches.

References

[1] Hui Lin, Zhiheng Ma, Xiaopeng Hong, Yaowei Wang, and Zhou Su. Semi-supervised crowd counting via density agency. In *ACM MM*, 2022.

[2] Xialei Liu, Joost Van De Weijer, and Andrew D Bagdanov. Leveraging unlabeled data for crowd counting by learning to rank. In *CVPR*, 2018.

[3] Vishwanath A Sindagi, Rajeev Yasarla, Deepak Sam Babu, R Venkatesh Babu, and Vishal M Patel. Learning to count in the crowd from limited labeled data. In *ECCV*, 2020.

[4] Antti Tarvainen and Harri Valpola. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In *NeurIPS*, 2017.

[5] Yanyu Xu, Ziming Zhong, Dongze Lian, Jing Li, Zhengxin Li, Xinxing Xu, and Shenghua Gao. Crowd counting with partial annotations in an image. In *ICCV*, 2021.