Active Crowd Counting with Limited Supervision – Supplementary Material –

Zhen Zhao[1⋆], Miaojing Shi[2⋆], Xiaoxiao Zhao[1], and Li Li[1,3]

[1] College of Electronic and Information Engineering, Tongji University
[2] King's College London
[3] Institute of Intelligent Science and Technology, Tongji University
zhenzhao0917@gmail.com; miaojing.shi@kcl.ac.uk; lili@tongji.edu.cn

This supplementary material provides more results and examples of AL-AC on standard counting benchmarks, i.e. ShanghaiTech [11], UCF_CC_50 [2], TRANCOS [1] and DCC [6], to demonstrate the generalization ability of our method.

1 More results

ShanghaiTech First, instead of using limited labeled data as in the paper, we keep increasing M till 280 and report the MAE on SHA in Fig. 1. It can be seen that, with about 80-100 labeled data (nearly 30%) labeled data, AL-AC already reaches the performance close to the fully-supervised method, as in [4] (Table 3 in the paper). The performance will saturate after some point and converge to that of baseline. This is also observed in other works [5, 8].

Second, we offer a variant of AL-AC inspired by [10] and compare to our original AL-AC. [10] is a state of the art active learning method where a variational autoencoder (VAE) and an adversarial network are trained to play a min-max game discriminating between unlabeled and labeled data with domain label 0 and 1, respectively. Samples from those predicted as "unlabeled" with the lowest probabilities (near 0) are selected for active annotations. We find this min-max idea to be similar to the gradient reversal layer (GRL) in our proposed distribution alignment between labeled and unlabeled data. The GRL assumes a domain label 0 for the unlabeled data and 1 for the labeled data. It multiplies the gradient by a negative constant (-1) during the network back propagation which enforces the feature distributions over the labeled and unlabeled data as indistinguishable as possible for the distribution classifier. We therefore select unlabeled samples with the lowest probabilities from our domain classifier. In this sense, the distribution alignment with latent MixUp is included in every learning cycle. We denote it by AL-AC-v as a variant of AL-AC and compare it to the full version of our AL-AC in the default setting (M = 40, m =10) on SHA and SHB in Table 1: Left. Our AL-AC still works clearly better than this variant.

Last, to test the generalization ability of our method, we offer the results under the default setting M = 40, m =10 by training on SHA and testing on

⋆ Authors contributed equally.

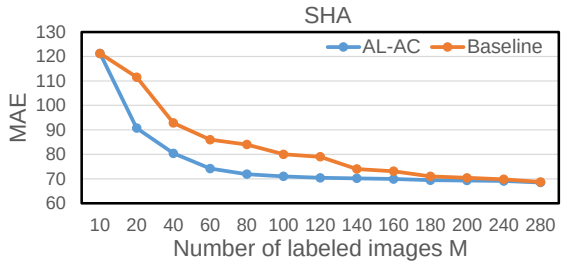

Fig. 1: Comparison of AL-AC against baseline. MAE is reported on SHA. The labeling budget M increases from 10 to 280.

M=40, m=10	SHA		SHB	
Method	MAE	MSE	MAE	MSE
Baseline	93.8	150.9	17.9	27.3
AL-AC-v	85.6	143.7	14.8	23.7
AL-AC	80.4	138.8	12.7	20.4

M=40, m=10	SHA → SHB		SHB → SHA	
Method	MAE	MSE	MAE	MSE
Baseline	39.2	53.8	167.4	283.2
AL-AC	30.5	45.0	144.0	238.5

Table 1: Left: Comparison of AL-AC to its variant AL-AC-v inspired by [10]. Right: Cross dataset performance of AL-AC. Experiments are on ShanghaiTech dataset with default setting $M = 40$ and $m = 10$.

SHB (SHA → SHB), and vice versa (SHB → SHA). The MAE and MSE for our proposed AL-AC and baseline are reported in Table 1: Right. It can be seen that our AL-AL improves the baseline substantially in this transfer setting.

UCF_CC_50 Our proposed AL-AC is mainly composed of two parts: 1) partition-based sample selection with weights (PSSW); and 2) distribution alignment with latent MixUp. We present detailed results of both components on the UCF_CC_50 dataset.

First, we compare our PSSW with random selection (RS) in Table 2: Left. We choose by default $M = 10$ and $m = 3$ (initial m is 4). The mean MAE at the starting point ($M = 4, m = 4$) is 645.8 for both PSSW and RS. For PSSW, it reduces to 479.2 with $M = 7$, and 387.3 with $M = 10$; in contrast, the MAE for RS is 505.8 and 444.7 for $M = 7$ and $M = 10$, respectively. PSSW produces clearly lower MAE than RS. We also present the result of PSSW with $M = 20$ and $m = 10$: the MAE is also much lower than that of RS.

Next, we study the effect of the proposed distribution alignment with latent MixUp in Table 2: Right. Like in the paper, we add GRL (gradient reversal layer) and MX (latent MixUp) to PSSW and report the result. For $M = 10$, $m = 3$, by adding GRL + MX to PSSW, the mean MAE and MSE further reduce 35.9 and 58.8 points, respectively. We also present the result for $M = 20$, $m = 10$, the contribution of GRL and MX is also significant (e.g. 27.2 points decrease on MAE). Notice PSSW + GRL + MX is equivalent to AL-AC in Table 4 in the paper.

Dataset	UCF_CC_50	
Method	PSSW	RS
M=4, m=4	645.8 ± 36.5	645.8 ± 36.5
M=7, m=3	479.2 ± 32.1	505.8 ± 35.3
M=10, m=3	387.3 ± 22.5	444.7 ± 25.9
M=20, m=10	345.9 ± 24.6	417.2± 29.8

Dataset	UCF_CC_50	
M=10, m=3	MAE	MSE
PSSW	387.3	506.9
PSSW + GRL + MX	351.4	448.1
M=20, m=10	MAE	MSE
PSSW	345.9	498.3
PSSW+ GRL + MX	318.7	421.6

Table 2: Ablation study of the proposed partition-based sample selection with weights (PSSW) and distribution alignment with latent MixUp (GRL + MX). Left: comparison of PSSW against random selection (RS), MAE is reported. Right: ablation of GRL + MX, MAE and MSE are reported. Experiments are on UCF_CC_50.

Method	Baseline*	AL-AC*	Lempitsky[3]	Hydra-3s[7]	POCR [6]	CSRNet [4]	CFF [9]
TRANCOS	10.1 ± 1.5	7.5 ± 0.8	13.8	11.0	9.7	3.6	2.0
DCC	7.4 ± 1.2	4.5 ± 0.4	-	-	8.4	-	3.2

Table 3: Comparison of AL-AC with state of the art on TRANCOS and DCC. Notice the labeling budget for TRANCOS and DCC is different. We label 10% of images over each set, such that $M = 80$ and $m = 20$ for TRANCOS, and $M = 10$ and $m = 3$ for DCC. MAE is reported.

TRANCOS and DCC To test the generalization ability of our AL-AC on other counting tasks, we evaluate it on TRANCOS and DCC for vehicle and cell counting, respectively. The global count error MAE is presented in Table 3. We label 10% of the images for each dataset. That is, $M = 80$, $m = 20$ for TRANCOS, and $M = 10$, $m = 3$ for DCC. Our MAE result is 7.5 on TRANCOS with an decrease of 2.6 points from baseline; 4.5 on DCC with an decrease of 2.9 points from baseline. With 10% labeled data, our AL-AC performs close to state of the art, particularly on DCC.

Next, in Table 4, we present the result of labeling 20% data for each; that is, $M = 160, m = 20$ for TRANCOS and $M = 20, m = 5$ for DCC. The mean MAE of AL-AC is 5.9 on TRANCOS with a decrease of 2.9 points from baseline; 3.8 on DCC with a decrease of 2.6 points from baseline. With 20% labeled data, our AL-AC performs quite close to the state of the art [6, 4, 9], which utilize full annotations of the datasets.

2 More examples

Several new examples of AL-AC are illustrated in Fig. 2 over different datsets (e.g. ShanghaiTech, UCF_CC_50, TRANCOS, and DCC).

Method	Baseline*	AL-AC*	Lempitsky[3]	Hydra-3s[7]	POCR [6]	CSRNet [4]	CFF [9]
TRANCOS	8.8 ± 1.4	5.9 ± 0.9	13.8	11.0	9.7	3.6	2.0
DCC	6.4 ± 1.1	3.8 ± 0.5	-	-	8.4	-	3.2

Table 4: Comparison of AL-AC with state of the art on TRANCOS and DCC datasets. *Notice that the labeling budget of AL-AC on TRANCOS and DCC is different. We label 20% of images over each set, such that $M = 160, m = 20$ for TRANCOS, and $M = 20, m = 5$ for DCC. MAE is reported.

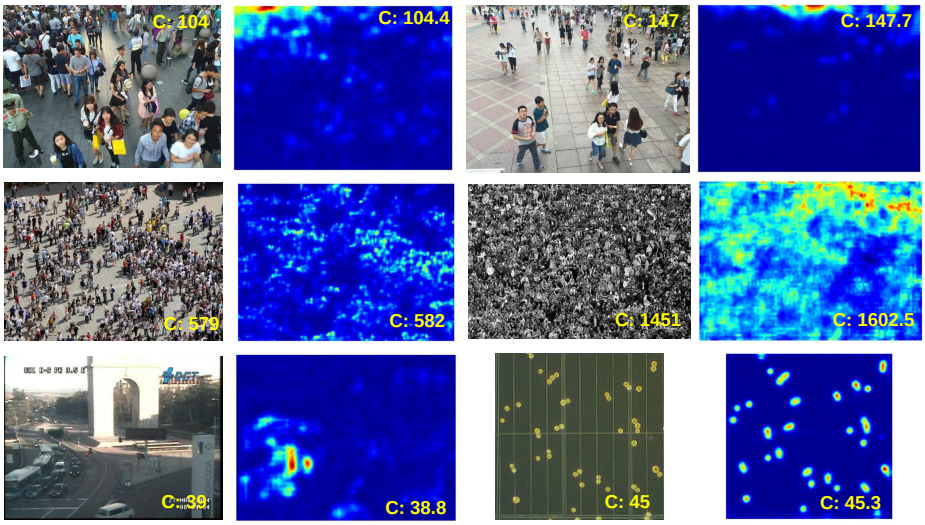

Fig. 2: Examples of AL-AC. Ground truth counts are in the original images while predicted counts in the estimated density maps.

References

1. Guerrero-Gómez-Olmedo, R., Torre-Jiménez, B., López-Sastre, R., Maldonado-Bascón, S., Onoro-Rubio, D.: Extremely overlapping vehicle counting. In: Iberian Conference on Pattern Recognition and Image Analysis (2015)
2. Idrees, H., Saleemi, I., Seibert, C., Shah, M.: Multi-source multi-scale counting in extremely dense crowd images. In: CVPR (2013)
3. Lempitsky, V., Zisserman, A.: Learning to count objects in images. In: NIPS (2010)
4. Li, Y., Zhang, X., Chen, D.: Csrnet: Dilated convolutional neural networks for understanding the highly congested scenes. In: CVPR (2018)
5. Liu, X., Van De Weijer, J., Bagdanov, A.D.: Exploiting unlabeled data in cnns by self-supervised learning to rank. IEEE transactions on pattern analysis and machine intelligence (2019)
6. Marsden, M., McGuinness, K., Little, S., Keogh, C.E., O'Connor, N.E.: People, penguins and petri dishes: adapting object counting models to new visual domains and object types without forgetting. In: CVPR (2018)
7. Onoro-Rubio, D., López-Sastre, R.J.: Towards perspective-free object counting with deep learning. In: ECCV (2016)

8. Sam, D.B., Sajjan, N.N., Maurya, H., Babu, R.V.: Almost unsupervised learning for dense crowd counting. In: AAAI (2019)
9. Shi, Z., Mettes, P., Snoek, C.G.: Counting with focus for free. In: ICCV (2019)
10. Sinha, S., Ebrahimi, S., Darrell, T.: Variational adversarial active learning. In: ICCV (2019)
11. Zhang, Y., Zhou, D., Chen, S., Gao, S., Ma, Y.: Single-image crowd counting via multi-column convolutional neural network. In: CVPR (2016)