Additional Tables and Figures for 'CountGD: Multi-Modal Open-World Counting'

Table 1: **CountGD's performance on 1-shot versus few-shot counting on FSC-147.** Providing more exemplars in the prompt increases the accuracy of CountGD. *Lower errors are better.* MAE := Mean Absolute Error, RMSE := Root Mean Squared Error.

	Val. MAE	Val. RMSE	Test MAE	Test RMSE
1 exemplar + text	8.00	30.29	8.7	83.21
3 exemplars + text	**7.10**	**26.08**	**5.74**	**24.09**

Table 2: **CountGD and GroundingDINO's counting performance on FSC-147.** Using *both* text and exemplars, CountGD achieves significantly better counting accuracy than GroundingDINO, which can only accept text. *Lower errors are better.* MAE := Mean Absolute Error, RMSE := Root Mean Squared Error.

	Val. MAE	Val. RMSE	Test MAE	Test RMSE
GroundingDINO	54.45	137.12	54.16	157.87
CountGD (ours)	**7.10**	**26.08**	**5.74**	**24.09**

Table 3: **CountGD, LOCA, and PseCo's counting performance on FSC-147.** Here, we report the best results from each paper. LOCA and PseCo use exemplars only to achieve their best results. CountGD achieves its best results and significantly better counting accuracy than both PseCo and LOCA by accepting *both* text and exemplars. *Lower errors are better.* MAE := Mean Absolute Error, RMSE := Root Mean Squared Error.

	Val. MAE	Val. RMSE	Test MAE	Test RMSE
PseCo	15.31	68.34	13.05	112.86
LOCA [9]	10.24	32.56	10.79	56.97
CountGD (ours)	**7.10**	**26.08**	**5.74**	**24.09**

Figure 1: **Text is sometimes not enough to specify the object to count.** In (a), given only text, CountGD accurately estimates the number of crystals. In (b), CountGD cannot accurately estimate the number of crystals in the X-ray image using text alone, since they look unfamiliar. In (c), providing an additional visual exemplar alleviates the issue. Input images are in the top row. Detected instances from CountGD are shown in the bottom row.

Table 4: **Influence of the proposed SAM Test-Time Normalization (SAM TT-Norm) on CountGD's counting accuracy.** The SAM TT-Norm provides a small improvement to CountGD's counting accuracy. Adaptive cropping is applied here, unlike in Tab. 4 of the appendix. Note: CountGD still achieves state-of-the-art accuracy without the SAM TT-Norm. *Lower errors are better.* MAE := Mean Absolute Error, RMSE := Root Mean Squared Error.

SAM TT-Norm	Val. MAE	Val. RMSE	Test MAE	Test RMSE
✗	7.79	28.70	7.03	26.74
✓	**7.10**	**26.08**	**5.74**	**24.09**

Table 5: **Sensitivity of CountGD's counting accuracy to $\boldsymbol{\lambda_{loc} : \lambda_{cls}}$ on FSC-147 given *both* text and exemplars.** Decreasing $\boldsymbol{\lambda_{loc}/\lambda_{cls}}$ improves the val. errors more than increasing it does. Deviating $\boldsymbol{\lambda_{loc}/\lambda_{cls}}$ from one worsens the test errors with increasing it harming the test accuracy the most. We choose $\lambda_{loc} : \lambda_{cls} = 1 : 5$ as this achieves the lowest validation set MAE. *Lower errors are better.* MAE := Mean Absolute Error, RMSE := Root Mean Squared Error.

λ_{loc}	λ_{cls}	Val. MAE	Val. RMSE	Test MAE	Test RMSE
1	1	8.64	44.71	**5.62**	**21.58**
5	1	8.55	35.65	8.01	82.55
1	**5**	**7.10**	**26.08**	5.74	24.09

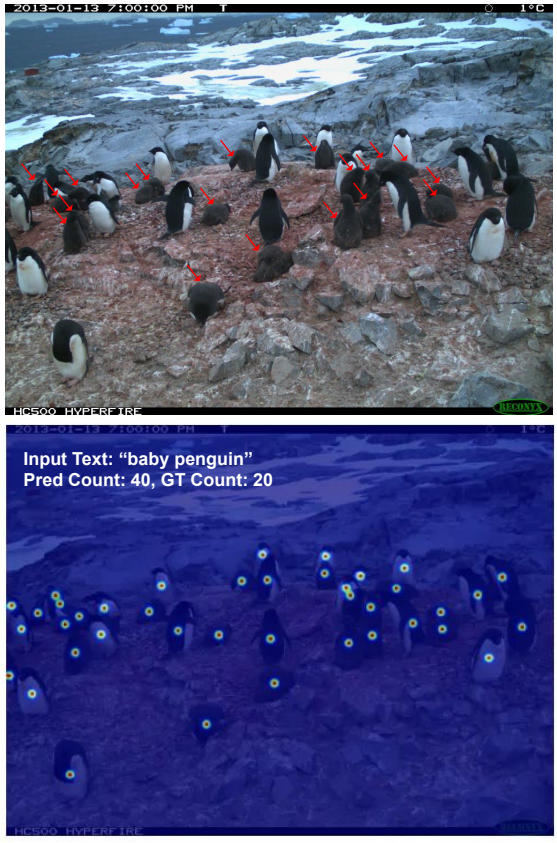

Figure 2: **Very fine-grained counting can be challenging.** CountGD cannot distinguish between the baby penguins (pointed to with red arrows in the top image) and the adult penguins. Given the text "baby penguin," CountGD counts all of the penguins in the input image. The adult and baby penguins look very similar. They have similar colors (mostly black) and shapes. The input image is in the top row. Detected instances from CountGD are shown in the bottom row.