A. Performance Across Object Quantities

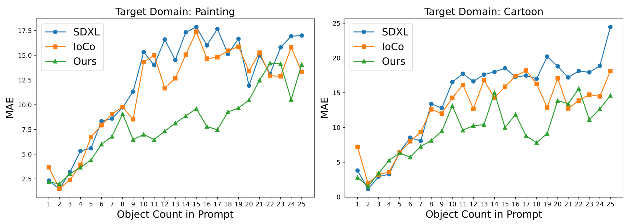

Figure 6. MAE curves across object quantities for *Painting* and *Cartoon* domains. QUOTA significantly reduces error compared to SDXL [23] and IoCo [35], especially in the common range of 5–20 objects.

Table 7 and Table 6 present the MAE results for 19 classes across two target domains: *Painting* and *Cartoon*, covering object quantities from 1 to 25. The results provide a detailed comparison between SDXL [23], IoCo [35], and QUOTA.

In the *Painting* domain, QUOTA achieves the lowest MAE in 20 out of 25 quantities, consistently outperforming both IoCo [35] and SDXL [23]. Similarly, in the *Cartoon* domain, QUOTA ranks first in 19 out of 25 cases, demonstrating its robustness in counting objects within stylized visual contexts. These findings highlight the strong performance of QUOTA in capturing complex visual variations and maintaining accurate object quantification across diverse classes and quantity ranges, further validating its effectiveness in real-world applications requiring precise object counting.

Figure 6 complements these tables by visualizing the per-count MAE curves for the two target domains. The curves reveal that QUOTA's largest gains concentrate in the practically most frequent range of 5–20 objects. In this interval, our method reduces the MAE by an average of 42.4% over SDXL [23] and 38.3% over IoCo [35] in the *Painting* domain, and by 36.0% and 28.8% respectively in the *Cartoon* domain. Below five objects, all three models already achieve very low errors, offering limited potential for further gains. In contrast, for counts above 20, the MAE remains high for all methods, likely due to increased visual complexity. These results indicate that QUOTA is particularly effective in the count range that matters most for real-world applications, while maintaining comparable performance on simpler cases.

B. Robust Generalization to Unseen Classes

As shown in Figure 7, we train the models using the 'apples' class in the painting target domain and evaluate their generalization on unseen classes with semantic similarity: 'tomatoes,' 'oranges,' and 'strawberries.' While IoCo [35] performs well on the seen class during training, it fails to generalize effectively to unseen classes, leading to inaccurate quantification across all tested classes. Similarly, SDXL [23] struggles to adapt, producing inconsistent object counts and exhibiting misalignment with textual prompts.

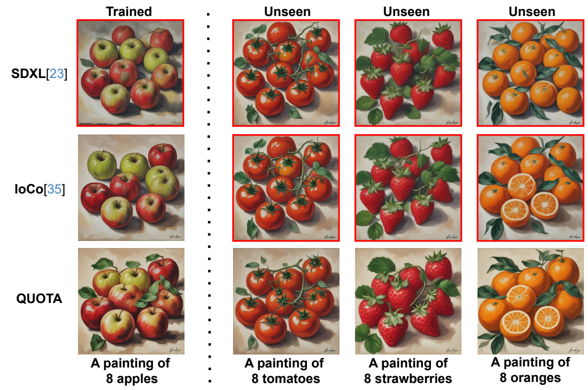

Figure 7. Generalization to unseen classes. Comparison of object quantification between SDXL [23], IoCo [35], and our QUOTA, evaluated on both trained and unseen classes.

In contrast, our method, QUOTA, exhibits strong generalization, with accurate quantification across unseen classes. This highlights QUOTA 's ability to adapt to unseen classes and preserve accuracy, even across stylistically distinct domains. These results highlight the effectiveness of QUOTA in addressing the challenges of cross-domain and cross-class generalization.

Models	1	2	3	4	5	6	7	8	9	10	11	12	13
SDXL [23]	3.80	**1.13**	**2.93**	**3.27**	6.47	8.53	8.07	13.40	12.80	16.53	17.73	16.63	17.60
IoCo [35]	7.20	1.93	3.20	3.60	6.40	8.00	9.33	12.60	12.00	14.27	16.13	12.67	16.80
QUOTA	**2.80**	1.53	3.40	5.27	**6.33**	**5.73**	**7.27**	**8.13**	**9.47**	**13.13**	**9.60**	**10.27**	**10.40**
Models	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**Avg**
SDXL [23]	18.00	18.53	17.27	17.47	17.00	20.20	18.80	17.20	18.13	17.93	18.87	24.47	14.10
IoCo [35]	**14.27**	15.87	17.40	18.20	16.27	12.87	17.07	**12.73**	**13.87**	14.73	14.47	18.13	12.40
QUOTA	15.00	**10.00**	**11.87**	**8.80**	**7.80**	**9.13**	**13.87**	13.40	15.60	**11.13**	**12.67**	**14.60**	**9.48**

Table 6. MAE comparison across 19 classes in the *cartoon* target domain for varying object counts. This table evaluates SDXL [23], IoCo [35], and QUOTA on object quantification for the object count ranging from 1 to 25, averaged across 19 classes in the *cartoon* target domain. QUOTA achieves the best performance in 19 out of 25 quantities, highlighting its superior accuracy and robustness in this target domain.

Models	1	2	3	4	5	6	7	8	9	10	11	12	13
SDXL [23]	2.33	**1.47**	3.20	5.33	5.60	8.33	8.6	9.73	11.33	15.33	14.00	16.60	14.53
IoCo [35]	3.67	1.53	**2.40**	3.93	6.73	7.93	9.07	9.77	8.53	14.33	15.00	11.67	12.67
QUOTA	**2.20**	2.00	3.00	**3.67**	**4.40**	**6.01**	**6.80**	**9.07**	**6.47**	**7.00**	**6.47**	**7.33**	**8.13**
Models	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**Avg**
SDXL [23]	17.33	17.87	16.00	17.67	15.13	16.67	11.93	15.00	13.13	15.80	16.93	17.00	12.27
IoCo [35]	15.07	17.40	14.67	14.80	15.47	15.87	13.40	15.27	**12.93**	**12.87**	15.80	**13.33**	11.36
QUOTA	**8.87**	**9.60**	**7.80**	**7 .47**	**9.27**	**9.67**	**10.47**	**12.47**	14.20	14.13	**10.53**	14.07	**8.04**

Table 7. MAE comparison across 19 classes in the *Painting* target domain for varying object counts. This table evaluates SDXL [23], IoCo [35], and QUOTA on object quantification for the object count ranging from 1 to 25, averaged across 19 classes in the *Painting* target domain. QUOTA achieves the best performance in 20 out of 25 quantities, highlighting its superior accuracy and robustness in this target domain.