Appendix

Efficient Discovery and Effective Evaluation of Visual Perceptual Similarity: A Benchmark and Beyond

A. Further evaluation

A.1. Additional models

In order to demonstrate that even without being a generator, it is possible to obtain good discovery results, we provide results for *8 additional* pre-trained models on our annotated GT in Tab. 7 and Tab. 8. We evaluated supervised top models: ResNet50 [20] and ConvNext (CN) [26] pretrained on ImageNet1K, ViT B-16 [12] pretrained on ImageNet21K and SwAG [37] weakly supervised pretrained through hashtags. Self-supervised top methods: MoCo [19], SwAV [4], MAE [18], and NoisyStud (NS) [44].

We observe that the results obtained by the newly added models are on par with the GT generators (Tab. 1 in the main paper). Specifically, SwAG produces SoTA performance on the discovery task. An exception is MAE which performs poorly compared to the other models. This can be attributed to its necessity for finetuning with a non-linear head before transferring to downstream tasks.

B. Optimization details

Our supervised finetuned baselines are complemented by a linear classification head (at the top of the backbone) that matches the dimension of the number of IDs and/or categories. By utilizing the attributed ID and/or category information of each gallery item, we minimize the categorical cross-entropy loss for 50 epochs. It is used with Adam optimizer, with weight decay of $w = 5 \cdot 10^{-5}$, and no momentum. The size of the mini-batches is set to be 64.

	HR	MRR	ROC-AUC		PR-AUC	
	@5	@5	Micro	Macro	Micro	Macro
RN50	6.0	6.7	64.4	71.1	54.3	66.0
MoCo	8.5	9.8	62.1	67.3	52.0	63.1
SwAV	8.3	9.6	62.4	68.1	52.1	63.5
MAE	0.6	0.7	47.4	51.7	39.8	51.7
CN	5.5	6.3	65.9	71.9	56.1	66.7
B-16	8.0	9.4	67.8	74.3	58.8	69.8
NS	9.0	10.4	70.8	76.5	62.0	70.5
SwAG	**10.0**	**11.1**	**72.4**	**78.5**	**63.4**	**72.7**

Table 8. Image in the wild discovery performance.

	HR	MRR	ROC-AUC		PR-AUC	
	@5	@5	Micro	Macro	Micro	Macro
RN50	26.5	30.9	67.4	80.7	93.4	93.6
MoCo	32.9	39.0	63.8	78.7	92.5	93.0
SwAV	41.0	**48.0**	63.9	79.0	92.6	93.1
MAE	13.3	17.0	59.6	77.1	91.1	92.5
CN	24.6	29.0	66.1	79.7	93.1	93.4
B-16	34.3	40.5	68.9	81.7	94.0	94.1
NS	24.0	28.6	69.0	82.2	93.8	94.2
SwAG	**40.3**	47.3	**75.4**	**85.0**	**95.4**	**95.2**

Table 7. Closed-catalog discovery performance.