Appendix

Figure 7: *Fine-tuning the diffusion model resolves the domain gap between ImageNet (collected almost two decades ago) and images generated by the stable diffusion model (trained recently).* **A.** ImageNet examples of the class “cellphone” show devices that were popular in 2006 when ImageNet was collected. **B.** Prompting the pretrained stable diffusion model (here: CLIP PROMPTS) generates images depicting newer cellphones used in recent times. **C.** Fine-tuning the DM (here: PSEUDOWORD+DM) closes this domain gap, as generated images show cellphones akin to the ImageNet samples in panel A.

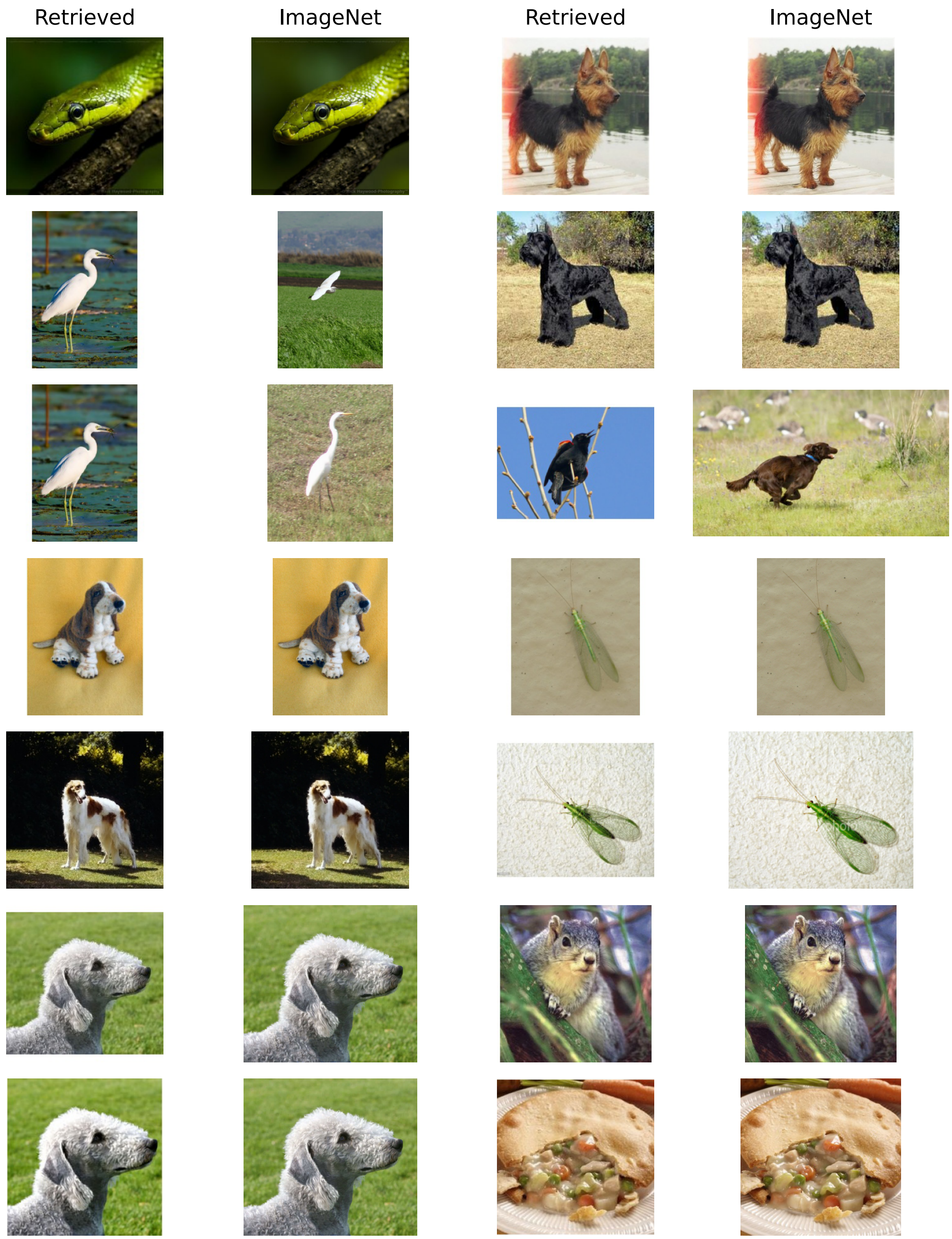

Figure 8: *Duplicate candidates found by comparing perceptual image hashes of retrieved images to our ImageNet test-split.*

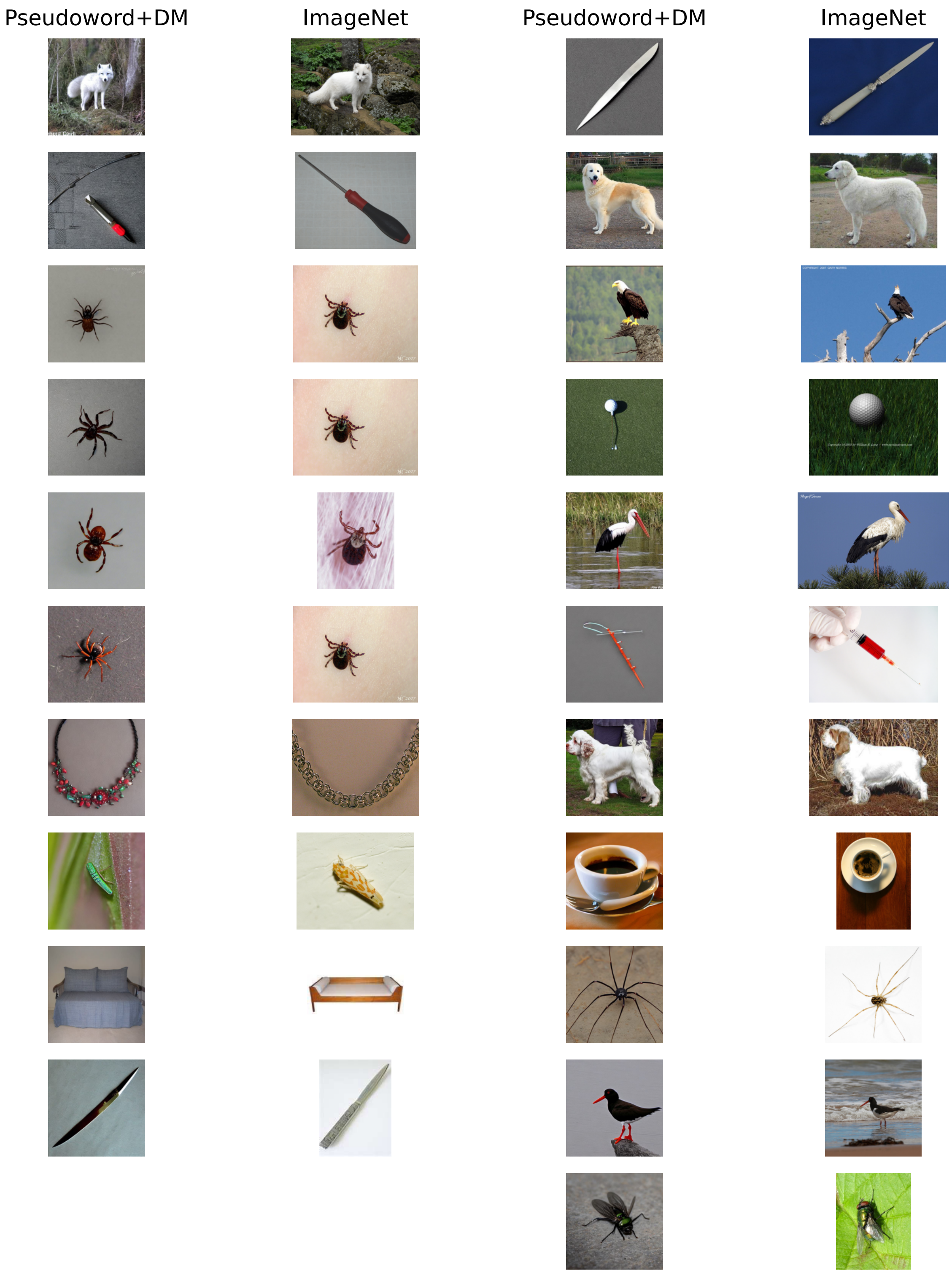

Figure 9: *Duplicate candidates found by comparing perceptual image hashes of generated images* (PSEUDOWORD+DM) *to our ImageNet test-split.*

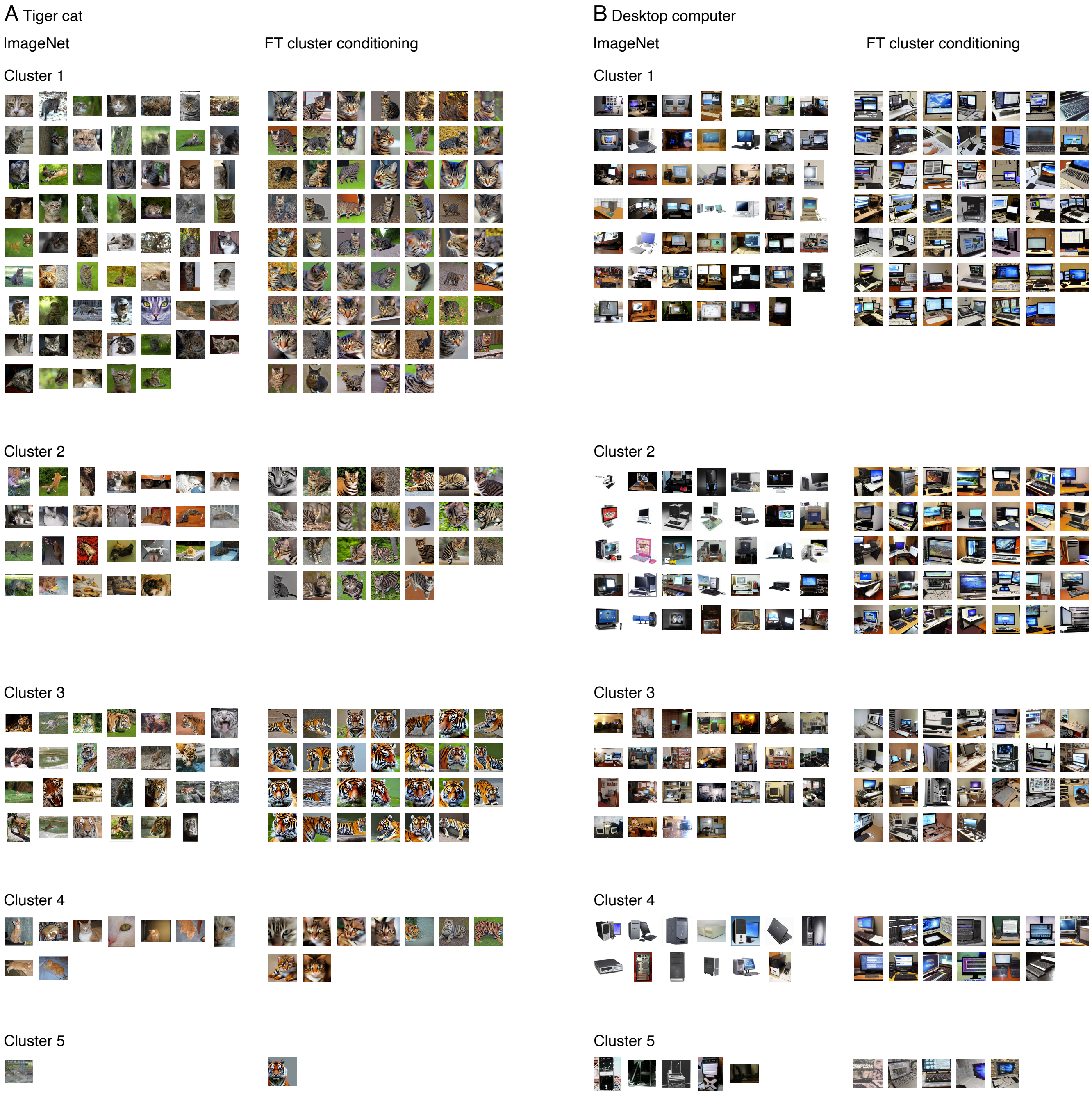

Figure 10: FT CLUSTER CONDITIONING *with $k = 5$ clusters compared to ImageNet.* Semantically similar ImageNet images are clustered together and one conditioning is learned for each cluster to reconstruct the training images (see section 3 for details). We exclude images resembling human faces to preserve data privacy. **A.** Examples for the class "tiger cat" which is ambiguous in ImageNet itself (left column). **B.** Examples for the class "desktop computer". Best viewed when zoomed in.

Figure 11: CLIP PROMPTS *examples for each CLIP text template.*

Figure 12: *examples for each* TEXTUAL INVERSION *text template.*

Figure 13: *Examples of* IMAGIC *optimization for various epochs.*