Supplmentary Material:
L-CAD: Language-based Colorization with Any-level Descriptions using Diffusion Priors

Zheng Chang[#1] **Shuchen Weng**[#2,3] **Peixuan Zhang**[1] **Yu Li**[4] **Si Li**[*1] **Boxin Shi**[2,3]
[1] School of Artificial Intelligence, Beijing University of Posts and Telecommunications
[2] National Key Laboratory for Multimedia Information Processing
School of Computer Science, Peking University
[3] National Engineering Research Center of Visual Technology
School of Computer Science, Peking University
[4] International Digital Economy Academy
{zhengchang98,pxzhang,lisi}@bupt.edu.cn
{shuchenweng, shiboxin}@pku.edu.cn liyu@idea.edu.cn

6 Appendix

6.1 Robustness for contour estimation

We leverage a referring segmentation model to roughly estimate object contours mentioned in the description, which enables us to perform the instance-aware sampling strategy. To further demonstrate the robustness of our model, we manually annotate a sequence of contours ranging from coarse to fine and visualize the corresponding colorization results. As shown in Figure 8, our model presents a remarkable ability to produce condition-consistent colorization results even using imprecise contours. This is because the sampling is performed in the latent space using downsampled contours and the compression decoder in the pixel space could adaptively fix color bleeding issues.

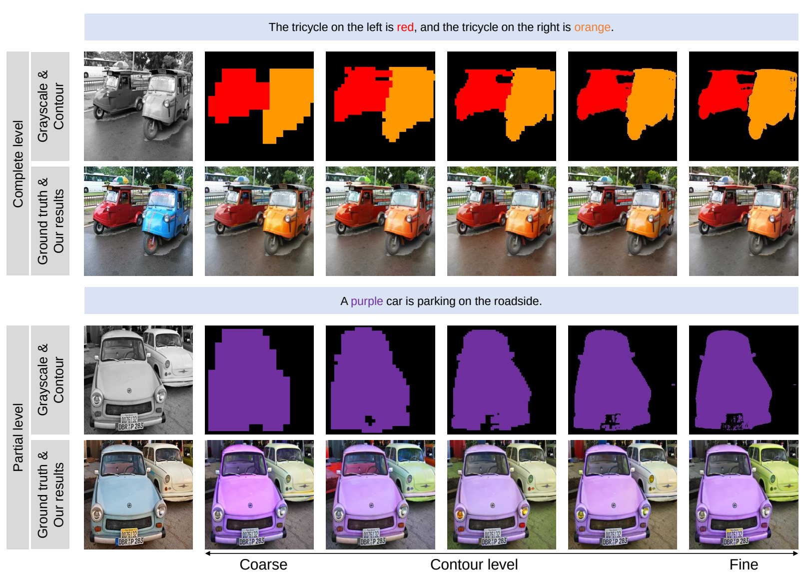

Figure 8: Visualization of colorization results by applying contours from coarse to fine.

# Equal contribution. * Corresponding author

37th Conference on Neural Information Processing Systems (NeurIPS 2023).

6.2 Additional comparison results

As presented in Sec. 4.1 of the main paper, we comprehensively evaluate our method on language-based colorization datsets, where we make comparisons with language-based colorization methods (*e.g.*, LBIE [3], ML2018 [6], Xie2018 [14], L-CoDe [12], L-CoDer [1], and L-CoIns [2]) using complete-level and partial level descriptions, and comparisons with automatic colorization methods (*e.g.*, CIC [15], InstColor [9], ChromaGAN [10], BigColor [5], DISCO [13], and CT^2 [11]) using scarce-level descriptions.

Following the evaluation protocol on ImageNet dataset [7], we evaluate colorization results at the more common resolution of 256×256, instead of 224×224 resolution in previous works [1, 2, 12]. This higher resolution increases the difficulty of the colorization, resulting in slightly lower scores for the quantitative metrics (see Tab. 1 of the main paper), compared to those reported in previous works [1, 2, 12]. Additionally, we provide more qualitative comparison results with language-based colorization methods and automatic colorization methods in Fig. 9 and Fig. 10, respectively.

6.3 Additional ablation results

To demonstrate the effectiveness of our proposed luminance-guided image compression, semantic-aligned latent representation, and instance-aware sampling strategy (details in Sec. 4.3 of the main paper), we create three baselines by disabling corresponding modules. Additional qualitative ablation study results are shown in Fig. 11.

6.4 Additional application results

We demonstrate our generalization capability by showing more colorization results on legacy black-and-white photos in Fig. 12, where results are presented sequentially from left to right using descriptions at the complete, partial, and scarce levels.

6.5 Diverse colorization results

By leveraging the inherent stochasticity of diffusion models [4, 8], which sample noise from a Gaussian distribution at each step of the denoising process, our method could effectively generate diverse colorization results for unmentioned objects in descriptions. We show our diverse colorization results with partial-level and scarce-level descriptions in Fig. 13.

Furthermore, we present more challenging results of our L-CAD using complete-level, partial-level, and scarce-level descriptions in Fig. 14, Fig. 15, and Fig. 16, respectively. These demonstrate that our method could produce high-quality colorization results for diverse and complex scenarios.

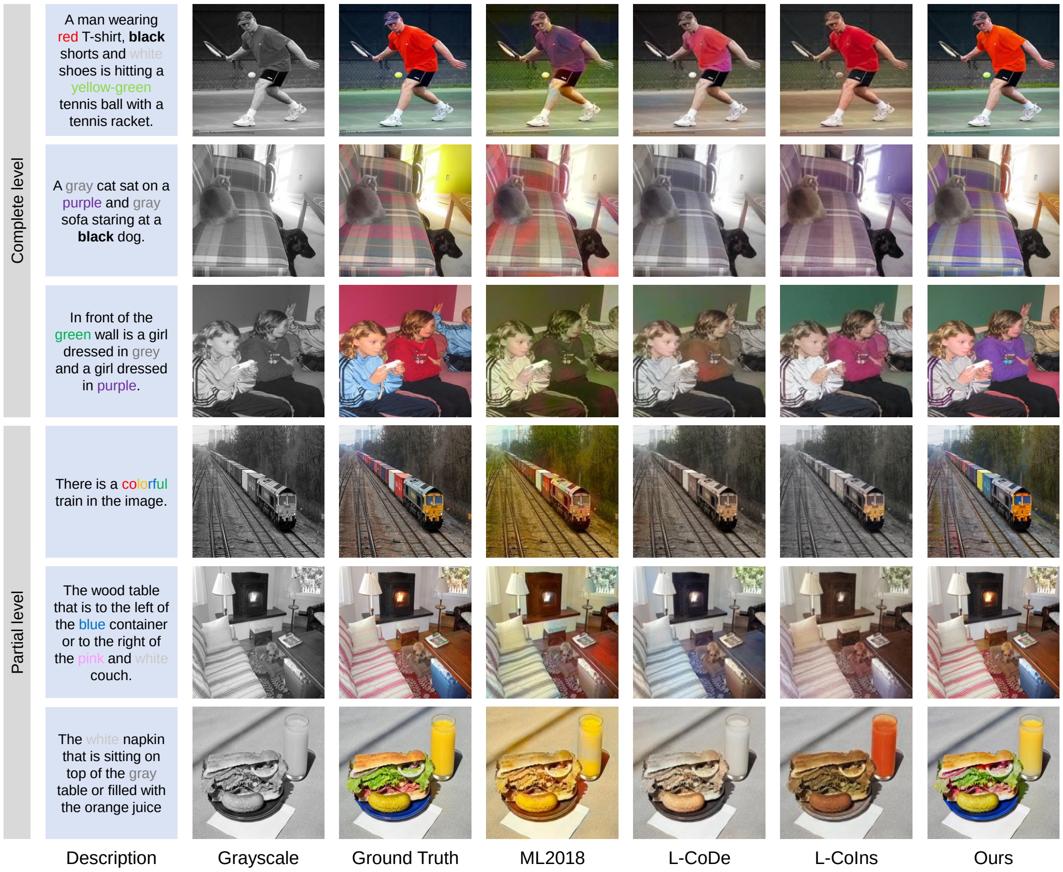

Figure 9: More comparison results with language-based colorization methods.

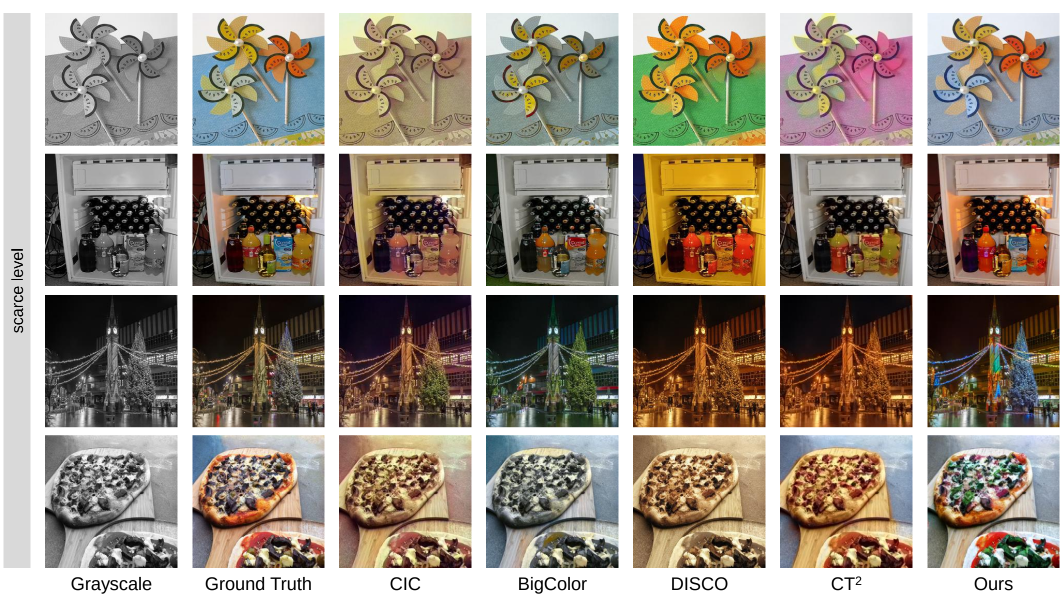

Figure 10: More comparison results with automatic colorization methods.

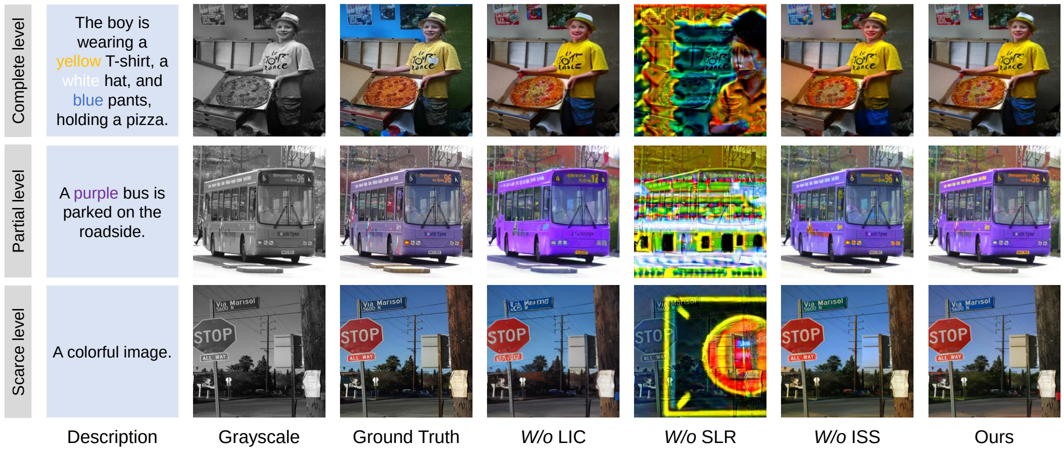

Figure 11: More ablation study results.

Figure 12: More colorization results of legacy black-and-white photos.

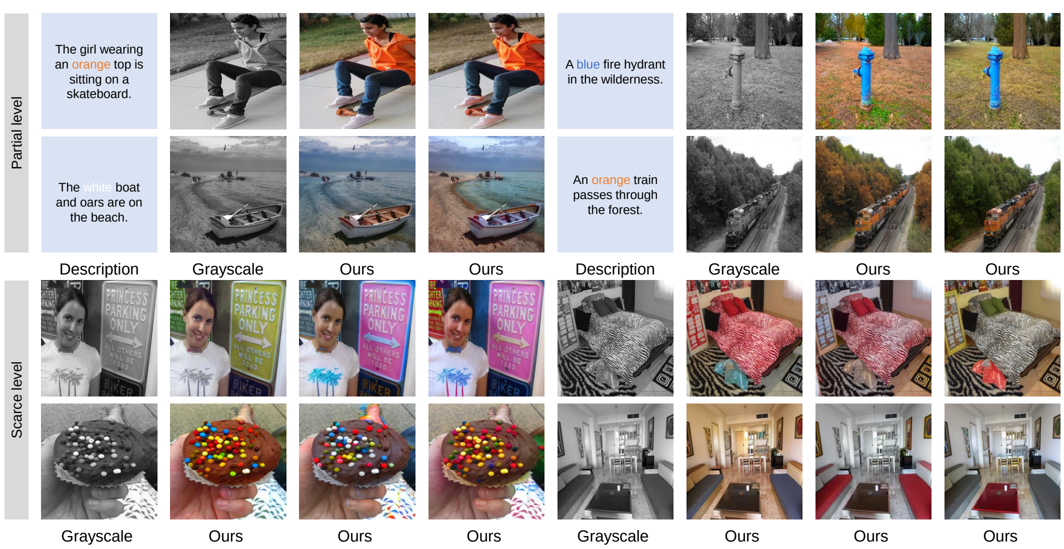

Figure 13: Diverse colorization results.

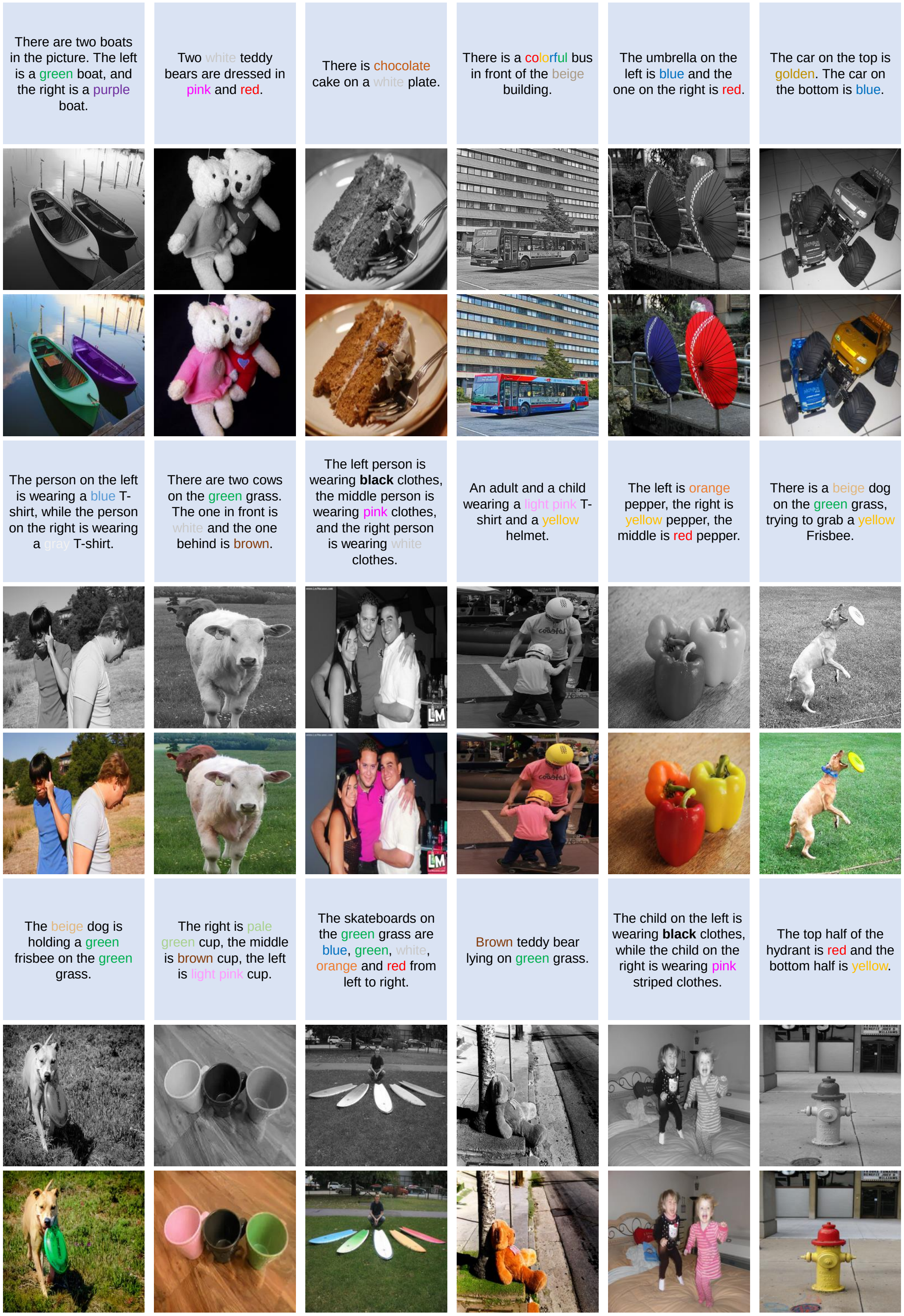

Figure 14: More results of our L-CAD using complete-level descriptions.

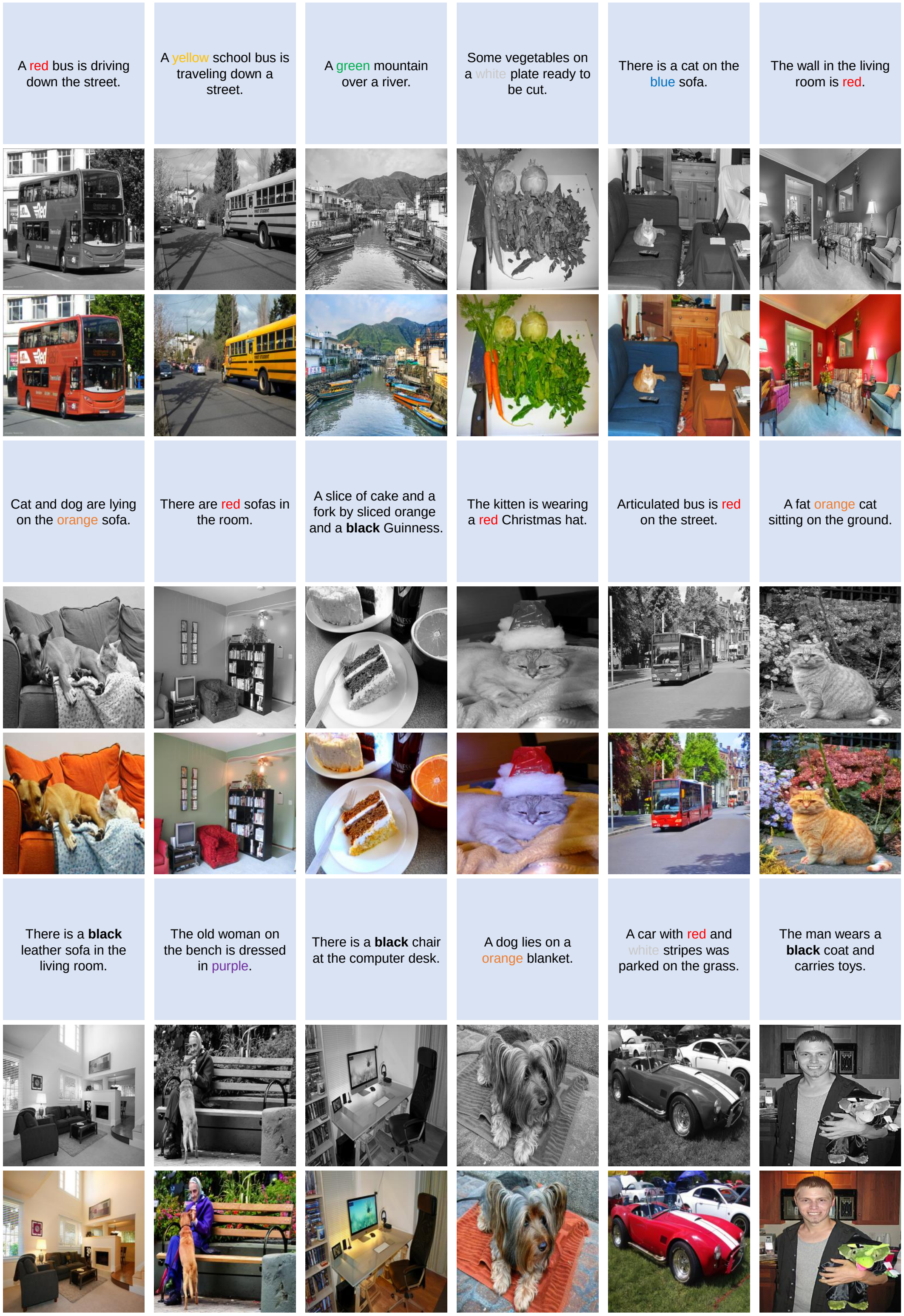

Figure 15: More results of our L-CAD using partial-level descriptions.

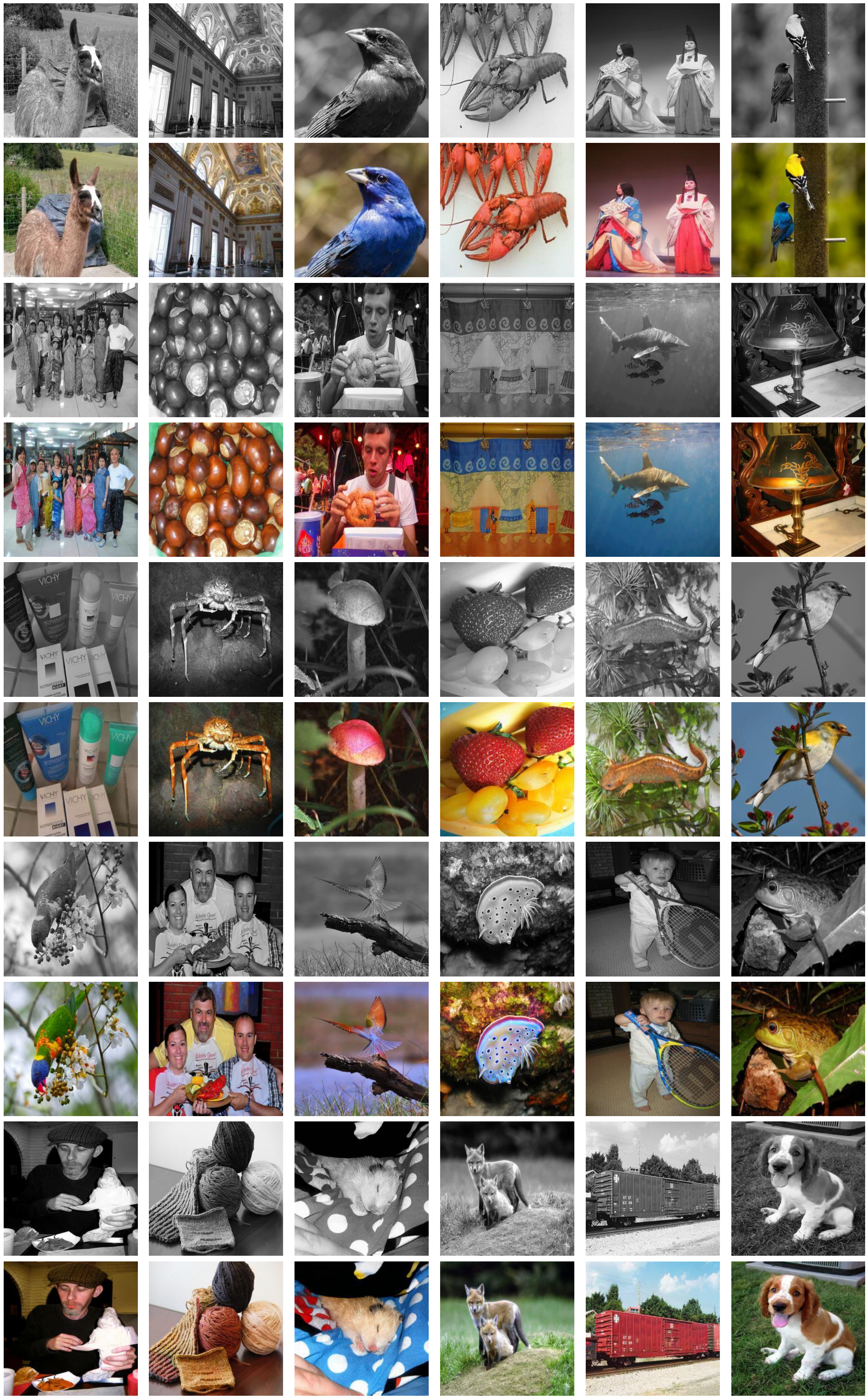

Figure 16: More results of our L-CAD using scarce-level descriptions.

References

[1] Z. Chang, S. Weng, Y. Li, S. Li, and B. Shi. L-CoDer: Language-based colorization with color-object decoupling transformer. In *ECCV*, 2022.

[2] Z. Chang, S. Weng, P. Zhang, Y. Li, S. Li, and B. Shi. L-CoIns: Language-based colorization with instance awareness. In *CVPR*, 2023.

[3] J. Chen, Y. Shen, J. Gao, J. Liu, and X. Liu. Language-based image editing with recurrent attentive models. In *CVPR*, 2018.

[4] J. Ho, A. Jain, and P. Abbeel. Denoising diffusion probabilistic models. In *NIPS*, 2020.

[5] G. Kim, K. Kang, S. Kim, H. Lee, S. Kim, J. Kim, S.-H. Baek, and S. Cho. BigColor: Colorization using a generative color prior for natural images. In *ECCV*, 2022.

[6] V. Manjunatha, M. Iyyer, J. Boyd-Graber, and L. Davis. Learning to color from language. In *NAACL*, 2018.

[7] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, et al. Imagenet large scale visual recognition challenge. *IJCV*, 2015.

[8] J. Song, C. Meng, and S. Ermon. Denoising diffusion implicit models. In *ICLR*, 2021.

[9] J.-W. Su, H.-K. Chu, and J.-B. Huang. Instance-aware image colorization. In *CVPR*, 2020.

[10] P. Vitoria, L. Raad, and C. Ballester. ChromaGAN: Adversarial picture colorization with semantic class distribution. In *WACV*, 2020.

[11] S. Weng, J. Sun, Y. Li, S. Li, and B. Shi. CT^2: Colorization transformer via color tokens. In *ECCV*, 2022.

[12] S. Weng, H. Wu, Z. C. Chang, J. Tang, S. Li, and B. Shi. L-CoDe: Language-based colorization using color-object decoupled conditions. In *AAAI*, 2022.

[13] M. Xia, W. Hu, T.-T. Wong, and J. Wang. Disentangled image colorization via global anchors. *TOG*, 2022.

[14] Y. Xie. Language-guided image colorization. Master's thesis, ETH Zurich, Departement of Computer Science, 2018.

[15] R. Zhang, P. Isola, and A. A. Efros. Colorful image colorization. In *ECCV*, 2016.