PalGAN: Image Colorization with Palette Generative Adversarial Networks Supplementary Material

Yi Wang[1], Menghan Xia[2], Lu Qi[3], Jing Shao[4], and Yu Qiao[1]

[1]Shanghai AI Laboratory [2]Tencent AI Lab [3]UC Merced [4]SenseTime Research

We describe the following contents in this supplementary file:

- Specific design of our given PalGAN.
- More colorization results on ImageNet, COCOStuff, and real-world legacy images.
- The discussion about limitations and failure cases.

Code and models will be released at https://github.com/shepnerd/PalGAN.

1 Network Architectures

PalGAN contains Palette Generator, Palette Assignment Generator, and Color Discriminator. We detail their topology and configuration below. For clarity, we denote Conv(k, s, c) indicates a convolutional operation whose kernel size, stride size, and output channel number of the used convolution are k, s, and c, respectively. The dilation ratio and padding size of Conv(k, s, c) are both set to 1. FC(c_{in},c_{out}) denotes a fully-connected operation where c_{in} and c_{out} stands for input channel and output one, respectively. $\uparrow$ and $\otimes$ denote a $2\times$ bilinear upsampling and concatenation (along with the channel dimension) operations, respectively. ResBlock(k, s, c) denotes a residual block as used in [3,5].

Palette Generator : $\mathbf{L} \rightarrow \text{Conv}_1(3,2,32) \rightarrow \text{LReLU} \rightarrow \text{Conv}_2(3,2,64) \rightarrow \text{LReLU} \rightarrow \text{Conv}_3(3,2,128) \rightarrow \text{LReLU} \rightarrow \text{Conv}_4(3,2,256) \rightarrow \text{LReLU} \rightarrow \text{Conv}(3,2,256) \rightarrow \text{LReLU} \rightarrow \text{Conv}(3,2,256) \rightarrow \text{LReLU} \rightarrow \text{FC}(4096,256) \rightarrow \text{Sigmoid} \rightarrow \hat{\mathbf{h}}$.

Palette Assignment Generator : $\mathbf{L} \downarrow \otimes\mathbf{M}_4 \rightarrow \text{Conv}_1(3,2,512) \rightarrow \text{LReLU} \rightarrow \text{ResBlock}(3,1,512) \rightarrow \text{ResBlock}(3,1,512) \rightarrow\uparrow [\otimes\mathbf{M}_3] \rightarrow \text{ResBlock}(3,1,256) \rightarrow \text{ResBlock}(3,1,256)$ $(\mathbf{S}) \rightarrow\uparrow [\otimes\mathbf{M}_2] \rightarrow \text{ResBlock}(3,1,128) \rightarrow \text{ResBlock}(3,1,128) \rightarrow\uparrow [\otimes\mathbf{M}_1] \rightarrow \text{ResBlock}(3,1,128) \rightarrow \text{ResBlock}(3,1,64) \rightarrow \text{CA}(\mathbf{S},\mathbf{L}) \rightarrow\uparrow \rightarrow \text{Conv}(3,1,2) \rightarrow \tanh \rightarrow \hat{\mathbf{I}}$,
where $\mathbf{M}_i$ indicates the output from Conv_i in Palette Generator.

Color Discriminator : $\hat{\mathbf{C}}\oplus\hat{\mathbf{I}}$ or $\mathbf{C}\oplus\mathbf{I} \rightarrow \text{Conv}(3,2,32) \rightarrow \text{LReLU} \rightarrow \text{Conv}(3,2,64) \rightarrow \text{LReLU} \rightarrow \text{Conv}(3,2,128) \rightarrow \text{LReLU} \rightarrow \text{Conv}(3,2,256) \rightarrow \text{LReLU} \rightarrow \text{Conv}(3,2,256) \rightarrow \text{LReLU} \rightarrow \text{FC}(128,256) \rightarrow \text{proj}(\mathbf{h}) \rightarrow \text{T or F}$,
where proj is the conditional projection (Eqn. 9) in the paper.

1.1 Model Capacity

The inference part of PalGAN (Palette Generator+Palette Assignment Generator) has only 22M parameters, smaller than that of UGColor [9] (34M), ColTrans [4] (74M), GPColor [8] (≥71M), and InstColor [6] (84M).

2 More Experimental Results and Analysis

We give more visual comparisons on ImageNet [2] (Figure 1), and COCO-Stuff [1] (Figure 2). Also, more reference-based results (Figure 3), and legacy photo colorization (Figure 4) are presented.

2.1 Limitations and Failure Cases

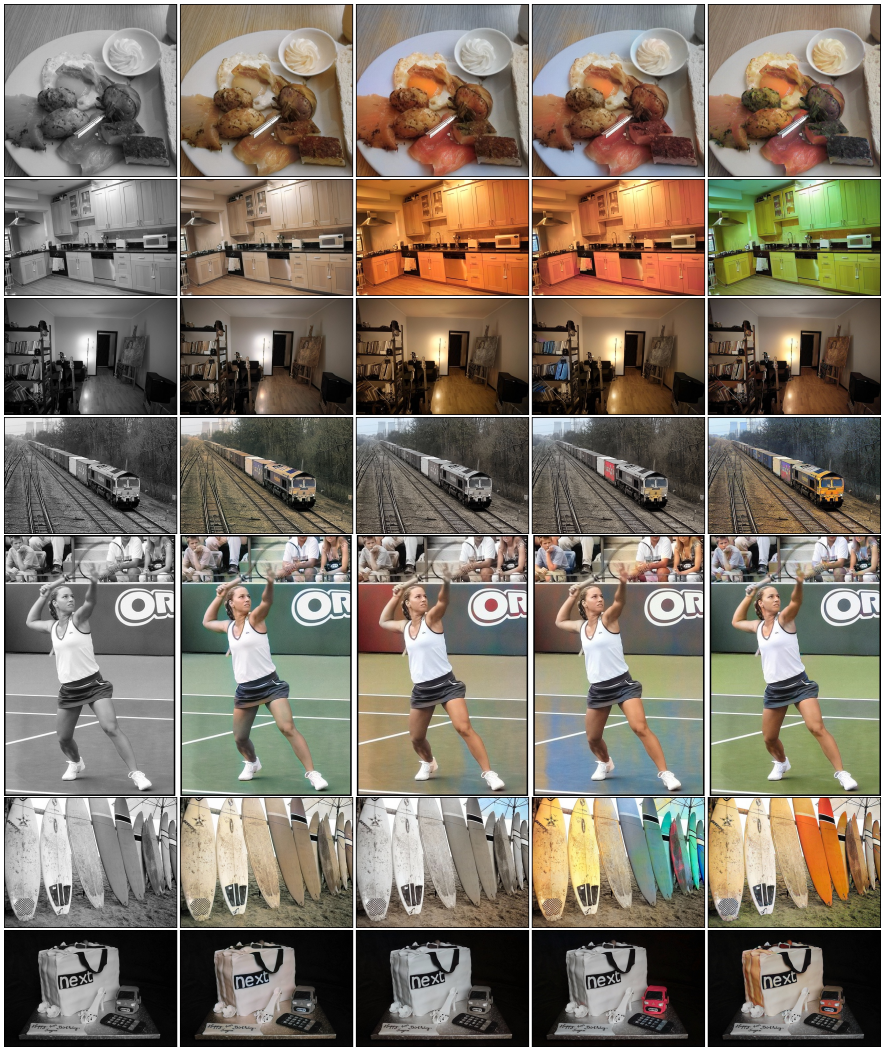
As discussed in the paper, PalGAN cannot handle small objects or structures (especially for high-resolution inputs) as chromatic attention is incapable of well representing these targets with small-scale feature maps, as shown in Figure 5. For crowds (top row in Figure 5) or small-scale buildings in the background (bottom row in Figure 5), all the employed methods fail to paint consistent and delicate colors according to the semantics.

2.2 Future Work

In the current design, we explicitly disentangle the color distribution from a spatial domain. Undoubtedly, predicting palettes locally will grant more controllability to users as we can regionally change the color distribution. It remains one of future works. Besides, considering the mentioned limitations on small-scale object colorization, using multi-scale attention mechanism or maintaining high-resolution representation as in [7] might address it, at the cost of large computational consumption.

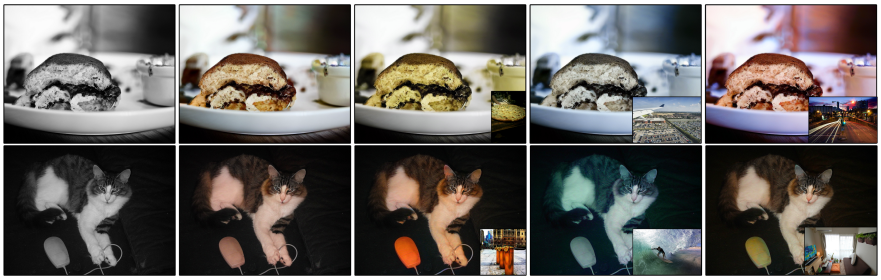
(1) Input (2) Deoldify (3) UGColor (4) InstColor (5) Ours

Fig. 1: Visual comparison on ImageNet.



(1) Input (2) Deoldify (3) UGColor (4) InstColor (5) Ours

Fig. 2: Visual comparison on COCO-Stuff.



Input Automatic Colorization Reference-based Colorization

Fig. 3: Our reference-based colorization results.

Fig. 4: Our legacy photo colorization results. First row: legacy images, and second row: our predictions.

(1) Input (2) Deoldify (3) UGColor (4) InstColor (5) Ours

Fig. 5: Failure cases. The given methods are unable to handle small objects and structures in the give inputs, leading to obvious color bleeding and inconsistency.

References

1. Caesar, H., Uijlings, J., Ferrari, V.: Coco-stuff: Thing and stuff classes in context. In: CVPR. pp. 1209–1218 (2018) 2
2. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: CVPR. pp. 248–255 (2009) 2
3. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: CVPR. pp. 1125–1134 (2017) 1
4. Kumar, M., Weissenborn, D., Kalchbrenner, N.: Colorization transformer. arXiv preprint arXiv:2102.04432 (2021) 2
5. Park, T., Liu, M.Y., Wang, T.C., Zhu, J.Y.: Semantic image synthesis with spatially-adaptive normalization. In: CVPR. pp. 2337–2346 (2019) 1
6. Su, J.W., Chu, H.K., Huang, J.B.: Instance-aware image colorization. In: CVPR. pp. 7968–7977 (2020) 2
7. Wang, J., Sun, K., Cheng, T., Jiang, B., Deng, C., Zhao, Y., Liu, D., Mu, Y., Tan, M., Wang, X., et al.: Deep high-resolution representation learning for visual recognition. IEEE transactions on pattern analysis and machine intelligence (2020) 2
8. Wu, Y., Wang, X., Li, Y., Zhang, H., Zhao, X., Shan, Y.: Towards vivid and diverse image colorization with generative color prior. In: ICCV. pp. 14377–14386 (2021) 2
9. Zhang, R., Zhu, J.Y., Isola, P., Geng, X., Lin, A.S., Yu, T., Efros, A.A.: Real-time user-guided image colorization with learned deep priors. arXiv preprint arXiv:1705.02999 (2017) 2