Supplementary Materials: Stylization-Based Architecture for Fast Deep Exemplar Colorization

Here we provide supplementary materials for our paper titled "Stylization-Based Architecture for Fast Deep Exemplar Colorization". We intend to provide more details that are not involved in the main body and show more experiment results. Specifically, we

- further compare the two feature extracting strategies, i.e., ours pre-colorization and He *et al.*'s gray VGG19, in the transfer sub-net visually and statistically;
- present the influence of AdaIN operation added in skip connection modules and show more results in photorealistic image stylization task with our transfer sub-net directly;
- explain detailed network architecture and sampling method of the colorization sub-net;
- show more colorization results and extend our model to multi-reference colorization;
- present some failure colorization examples.

1. Transfer sub-net

1.1. Gray VGG19 vs Pre-colorization

Table 1. Classification accuracies calculated on ImageNet validation dataset.

Method VGG19	Top-1 Class Acc(%)	Top-5 Class Acc(%)
Original VGG19 tested on color images	73.10	91.24
Original VGG19 tested on gray images	59.68	80.56
Our VGG19 tested on gray images	68.78	85.64

The intermediate outputs of VGG19 are usually took as feature representation due to its good extracting ability. However, the original VGG19 is trained on color images and has a degraded accuracy on gray image recognition task. He *et al.* [2] propose to train a gray VGG19 and achieve acceptable top-5 accuracy. Our solution is to use the original color VGG19 by utilizing a pre-trained colorization network to give the gray target image pre-color. It increases the top-5 accuracy of gray VGG19 from 59% to 68% and approaches that of the original VGG19 (73%) evaluated on color images (see Table 1).

We utilize the intermediate features extracted by the gray VGG19 for our transfer sub-net and show the results in Fig.1. Compared with our pre-colorization method, the results of gray VGG19 is under-saturated, which shows its poor representational ability.

1.2. AdaIN

We have compared the effect of AdaIN operations added in different parts of the transfer sub-net, which proves that AdaIN has a good performance in encoder and decoder modules of our transfer sub-net. In fact, we can also add it into skip connection modules to further enhance the transfer effect without expensive cost. In Fig. 2, the boxes in yellow show the minor differences when additional AdaIN operations are added in skip connection modules.

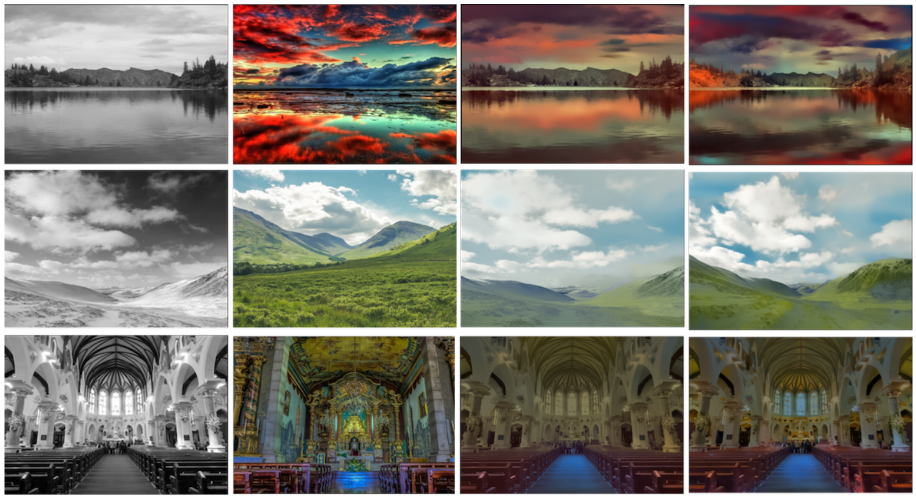

Figure 1. **Comparisons between gray VGG-19 and pre-colorization with the tranfer sub-net.**

Figure 2. **The effect of AdaIN added in skip connection modules.**

1.3. Photorealistic Stylization Comparison

Our transfer sub-net can be applied to photorealistic image stylization task directly and solve most problems in current approaches. Classical methods in photorealistic transfer are mostly matching the color of the images, which causes severe distortion and undesired artifacts. Luan *et al.* [5] propose to augment traditional Neural Style algorithm with an additional regularization term, which requires heavy computation to solve the regularized optimization problem. Since max-pooling operation reduces spatial information and fails to recover detailed structures of the input images, Li *et al.* [4] replace it with the unpooling layer [6]. Although PhotoWCT alleviates distortion and has much fewer structural artifacts, semantically similar regions in PhotoWCT-stylized result are often inconsistent. To tackle this issue, additional smooth operation is added, which, however, leads to over-smooth result. Our transfer sub-net can address these problems simultaneously using a progressive way. The details of the transfer sub-net are present in the main body. More comparison results can be seen in Fig. 3.

Besides, Table 2 shows the runtime comparison with aforementioned methods. These approaches are implemented on a PC with Intel E5 2.5GHz CPU and a single NVIDIA 1080Ti GPU. OOM denotes Out Of Memory exception. The experiments show that our transfer sub-net obtains the best transferring results with the least runtime.

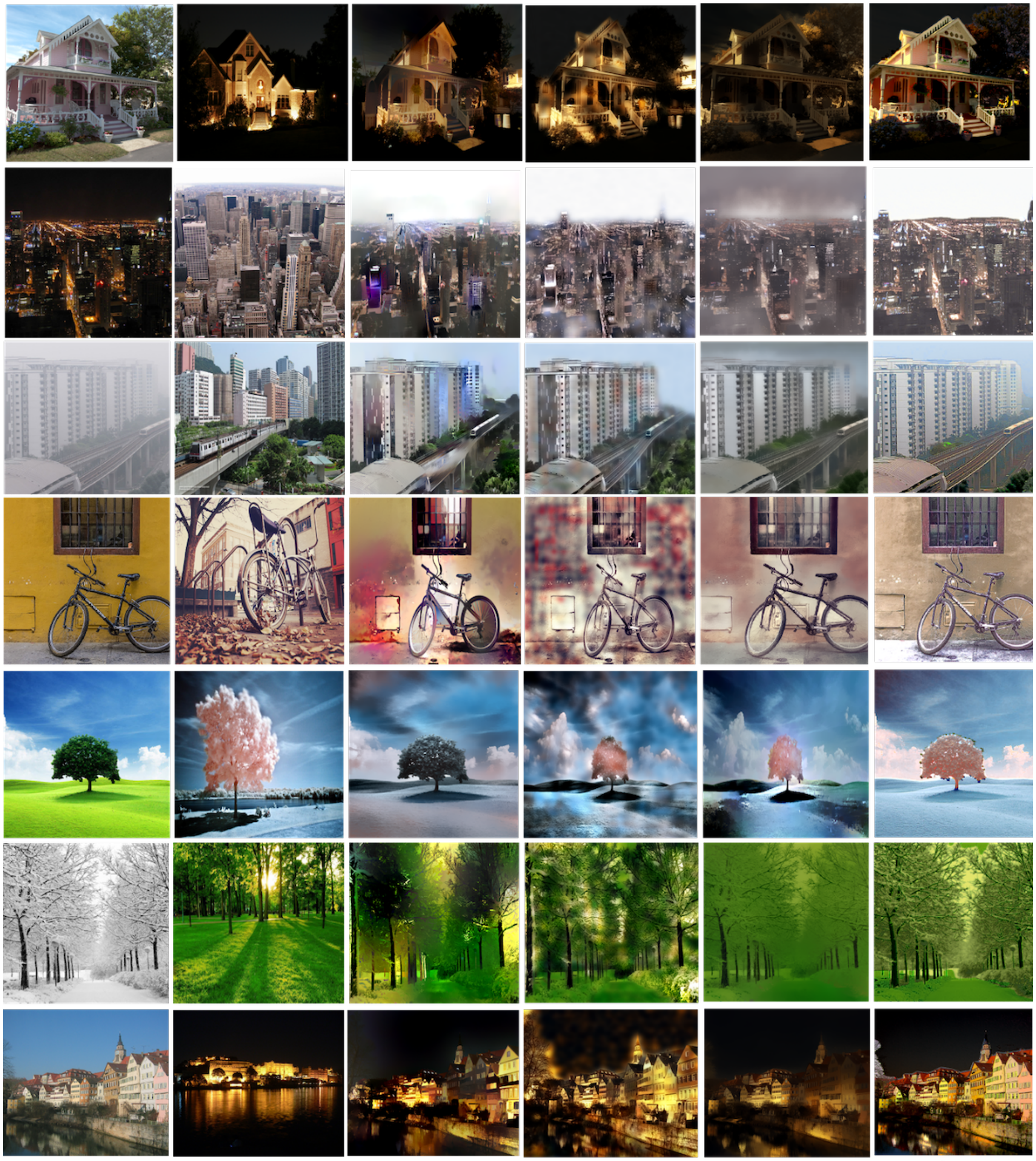

Figure 3. **Comparisons between current methods in photorealistic image stylization task.**

Table 2. Runtime comparisons on photorealistic image stylization task.

Image Size	256×256	512×512	1024×1024
DPST	234.92	803.76	3508.46
PhotoWCT	2.12+3.01	2.19+8.51	OOM
Ours(WCT)	0.68	1.22	2.18
Ours(AdaIN)	0.04	0.14	0.59

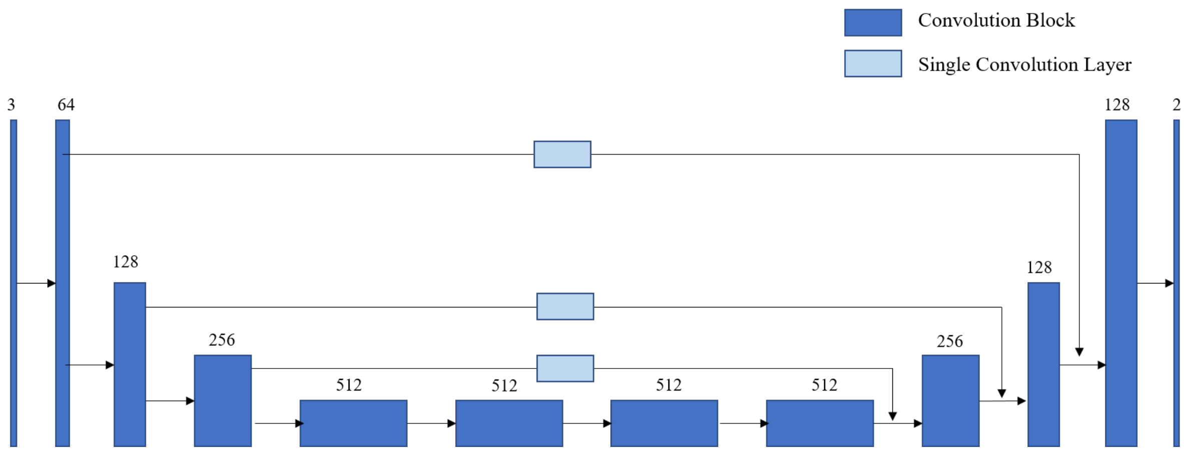

Figure 4. Colorization sub-net architecture.

2. Colorization sub-net

2.1. Network architecture

We use an analogous U-Net [8] architecture in Fig.4, which is formed by ten feature blocks ($Conv1 - 10$) and one output block ($Conv\_out$). Specially, every feature block consists of 2-3 $Conv - Relu$ pairs and ends with a BN layer. The output block contains a $Conv$ layer and a final $tanh$ layer like most image generation tasks [3]. In $Conv1 - 4$, features are progressively halved spatially while doubling in dimension. In bottleneck layers ($Conv5 - 6$), dilated convolution with factor 2 is used to get broader receptive field and keep additional information. Then spatial resolution is recoved while feature dimensions are halved in $Conv7 - 10$, Symmetric shortcut connections are added to help the network to recover spatial information and also enables easy accessibility to low-level information for later layers, for example, the $Conv2$ is connected to the $Conv8$ block.

2.2. Sample Principles

It is hard to refine the coarse ab map by feeding it to the colorization sub-net directly. We bypass the direct way via sampling patches from the ab map flexibly and propagating these sampled patches to the semantically related regions. The same sample principles are used in [10]. Patches size is drawn uniformly from size 1×1 to 9×9, then the average ab value within the patch is fed to the sub-net. We expect more patches are sampled in the center of the images. Hence, each patch location is sampled from a 2-D Gaussian with $\mu = \frac{1}{2}[H, W], \sum = diag([(\frac{H}{4})^2, (\frac{W}{4})^2])$ and the number of patches are drawn from a geometric distrbution with $p = \frac{1}{8}$. In Fig. 6, we compare the colorization results with different samples. It can be seen that the sampled ab actually can be propagated to semantically related regions. If we just input the entire coarse ab map to the colorization sub-net, the sub-net just simple copies the ab value to the output, which is shown in Fig. 6(e). Although the sampling idea is based on intuition, it is proven to work well in practice.

3. Experiment Results

3.1. Colorization Comparison

we have compare our method with four existing exemplar-based colorization methods[9, 1, 7, 2] in the main body. More colorization comparison results can be found in Fig.5. Our method still has the best colorization performance,

3.2. Multi-Reference Colorization

Reference selection is crucial to achieve satisfactory results in exemplar-based colorization. Current methods always fail to obtain a natural appearance in the regions where no proper color guidance can be found in the reference. This problem can be addressed by introducing multiple references. Pixel matching approach in [2] is based on global search and hard to map objects in different reference images. Although we do not have to deal with such problem exclusively, our model is well suited for multi-reference colorization. We interpolate between feature maps with M reference images $r_1, r_2, ..., r_M$

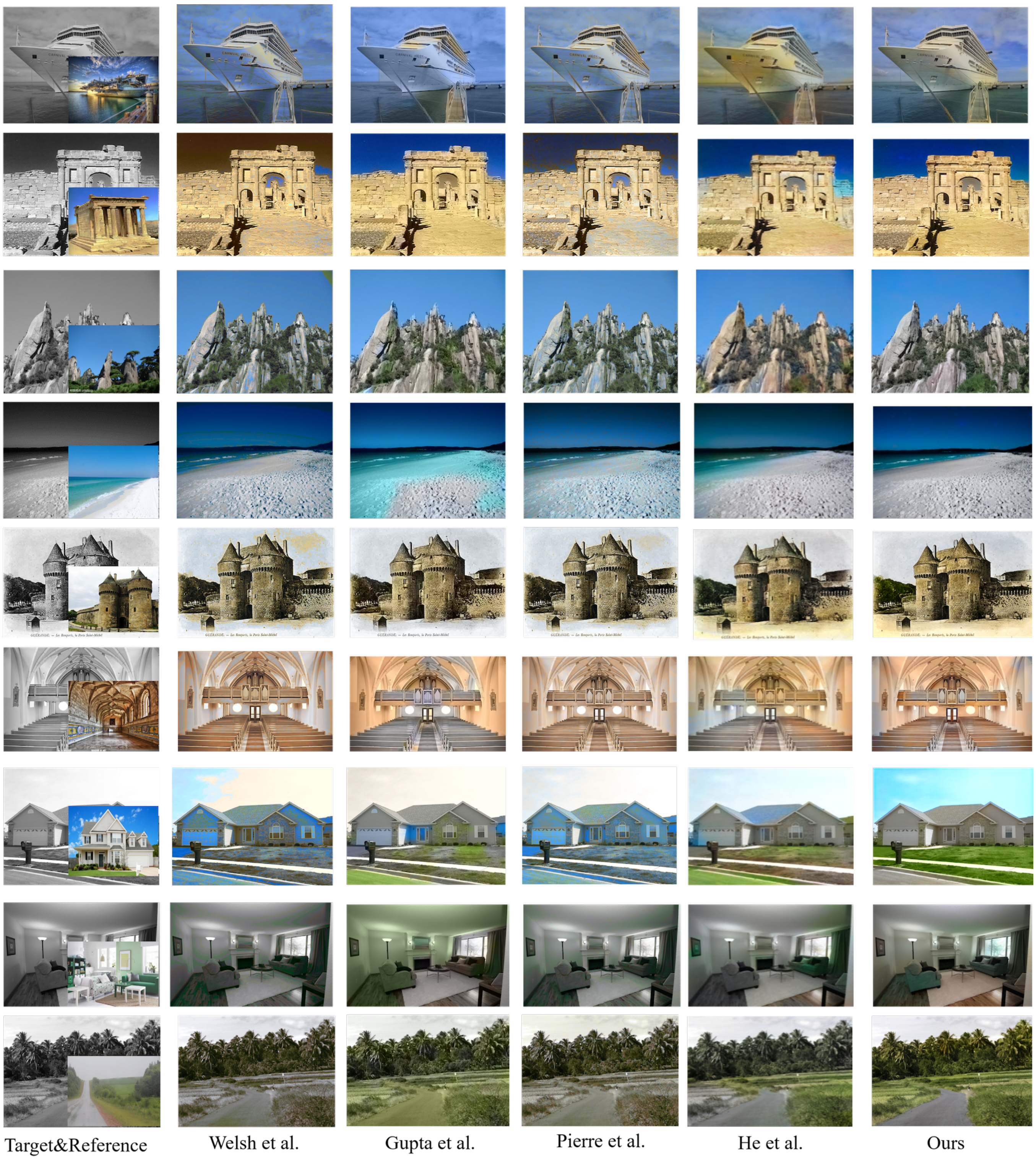

Figure 5. Comparison results between current exemplar colorization methods.

and corresponding weights $w_1, w_2, ..., w_M$.

$$T(t, (r_i, w_i)_{i=1}^{M}) = g\left(\sum_{m=1}^{M} w_m AdaIn(f(t), f(r_m))\right) \tag{5}$$

Figure 6. The first two rows contain target, sampled images and their corresponding colorization results. The third row present the results of other exemplar colorization methods.

Figure 7. Multi-reference colorization.

The semantic map can be used for more accurate transferring. Unmatched object will be neglected and less erroneous information will be affected in the final multi-reference colorization. The result of multi-reference colorization can be found in Fig 7.

Figure 8. Failure examples

3.3. Failure Examples

Although our model can obtain promising results in most cases, there is still plenty of room for improvement. For example, some colors in the coarse ab map are hardly propagated by the colorization sub-net. We show two failure examples in Fig. 8. One can see that is difficult to refer the sky with red color even if the transfer sub-net provides a credible map (see the first row of Fig. 8). He *et al.*'s method is also based on deep neural network and generates the similar results with ours. The reason may lie in that images in the training dataset we used are unevenly distributed. Exploring more advanced network or constructing special dataset are expected to ensure more reliable propagation while rendering meaning colors for content-unrelated objects in the future.

References

[1] Raj Kumar Gupta, Alex Yong-Sang Chia, Deepu Rajan, Ee Sin Ng, and Huang Zhiyong. Image colorization using similar images. In *Proceedings of the 20th ACM international conference on Multimedia*, pages 369–378. ACM, 2012. 4

[2] Mingming He, Dongdong Chen, Jing Liao, Pedro V Sander, and Lu Yuan. Deep exemplar-based colorization. *ACM Transactions on Graphics (TOG)*, 37(4):47, 2018. 1, 4

[3] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *IEEE CVPR*, pages 1125–1134, 2017. 4

[4] Yijun Li, Ming-Yu Liu, Xueting Li, Ming-Hsuan Yang, and Jan Kautz. A closed-form solution to photorealistic image stylization. In *IEEE ECCV*, pages 453–468, 2018. 2

[5] Fujun Luan, Sylvain Paris, Eli Shechtman, and Kavita Bala. Deep photo style transfer. In *IEEE CVPR*, pages 4990–4998, 2017. 2

[6] Hyeonwoo Noh, Seunghoon Hong, and Bohyung Han. Learning deconvolution network for semantic segmentation. In *Proceedings of the IEEE international conference on computer vision*, pages 1520–1528, 2015. 2

[7] Fabien Pierre, J-F Aujol, Aurélie Bugeau, Nicolas Papadakis, and V-T Ta. Luminance-chrominance model for image colorization. *SIAM Journal on Imaging Sciences*, 8(1):536–563, 2015. 4

[8] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2015. 4

[9] Tomihisa Welsh, Michael Ashikhmin, and Klaus Mueller. Transferring color to greyscale images. 21(3):277–280, 2002. 4

[10] Richard Zhang, Jun-Yan Zhu, Phillip Isola, Xinyang Geng, Angela S Lin, Tianhe Yu, and Alexei A Efros. Real-time user-guided image colorization with learned deep priors. *ACM Transactions on Graphics*, 36(4):1–11. 4