Supplementary Materials for Time-Travel Rephotography

XUAN LUO, University of Washington
XUANER (CECILIA) ZHANG, Adobe Inc.
PAUL YOO, University of Washington
RICARDO MARTIN-BRUALLA, Google Research
JASON LAWRENCE, Google Research
STEVEN M. SEITZ, University of Washington and Google Research

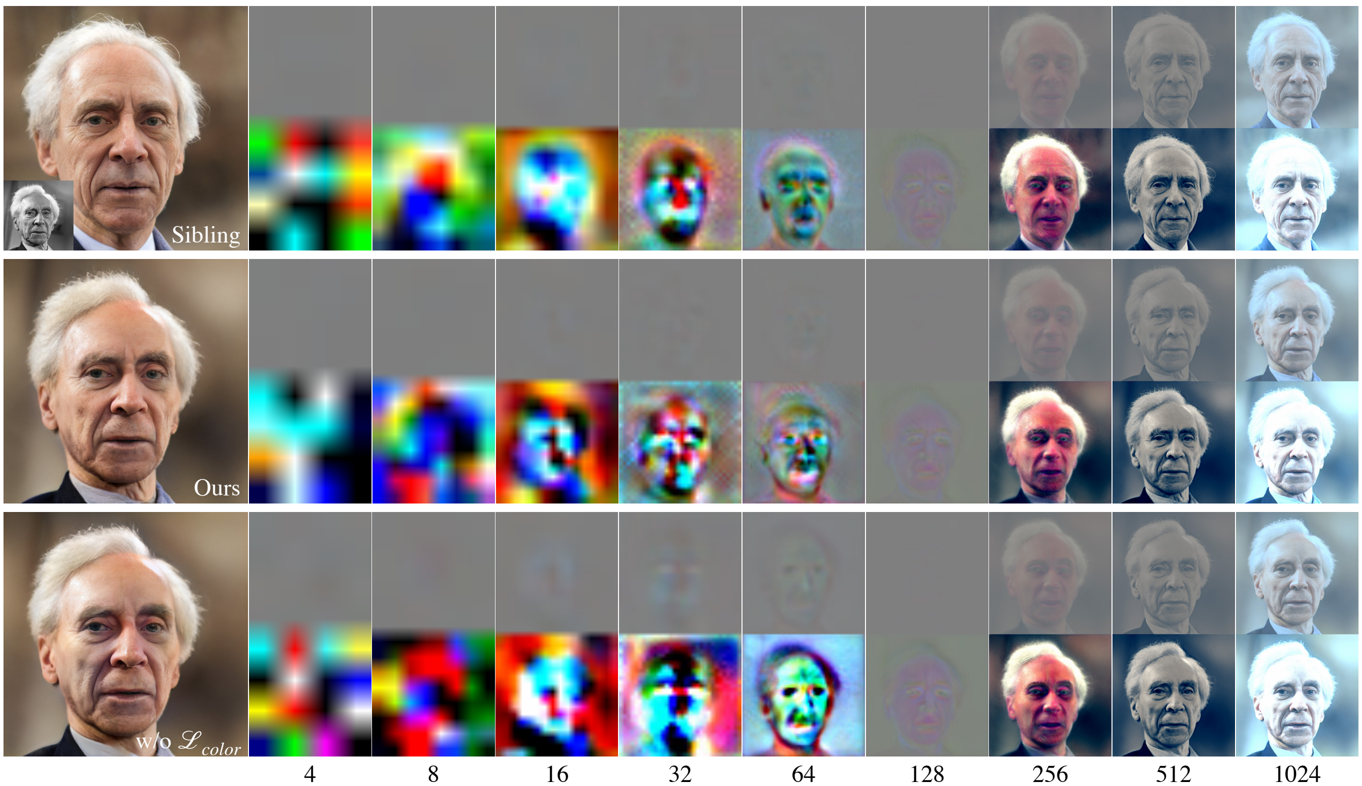

Fig. 1. Effect of the *color transfer loss*, $\mathcal{L}_{color}$. The *upper image* of each row visualizes the ToRGB layer outputs of different resolutions before adding the constant bias. A gray image implies zero values. We scale the pixel values up in the *bottom image* of each row for ease of visualization. With $\mathcal{L}_{color}$, the ToRGB outputs of the coarse layers have small values similar to the sibling. Without $\mathcal{L}_{color}$, the ToRGB pixel values are too large, causing low-frequency unnatural color variation in the output. For example, the nose region is unnaturally purple and the cheeks show too much yellow, which are introduced by the extreme values in layer 16 and 32. Input image: Bertrand Russel (1872 - 1970) from BBC Photo Library.

Authors' addresses: Xuan Luo, xuanluo@cs.washington.edu, University of Washington; Xuaner (Cecilia) Zhang, Adobe Inc., cecilia77@berkeley.edu; Paul Yoo, University of Washington, yoosehy@cs.washington.edu; Ricardo Martin-Brualla, Google Research, rmbrualla@google.com; Jason Lawrence, Google Research, jdlaw@google.com; Steven M. Seitz, University of Washington and Google Research, seitz@cs.washington.edu.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
© 2021 Association for Computing Machinery.
0730-0301/2021/12-ART213 $15.00
https://doi.org/10.1145/3478513.3480485

ACM Reference Format:
Xuan Luo, Xuaner (Cecilia) Zhang, Paul Yoo, Ricardo Martin-Brualla, Jason Lawrence, and Steven M. Seitz. 2021. Supplementary Materials for Time-Travel Rephotography. *ACM Trans. Graph.* 40, 6, Article 213 (December 2021), 2 pages. https://doi.org/10.1145/3478513.3480485

1 EFFECT OF THE COLOR TRANSFER LOSS

StyleGAN2 [Karras et al. 2020] produces an image by summing up multiple ToRGB layer outputs. Fig. 1 visualizes the ToRGB outputs, ψ_l, for each layer l before adding its constant bias term. The coarse layer ψ_l's of the in-domain sibling image have negligible values, and hence contributes little to the final color output. Without the color transfer loss ($\mathcal{L}_{color}$), ψ_l goes out of typical range for an in-domain face, producing color artifacts as shown in Fig. 1(bottom) and Fig. 3 of the main paper. Adding $\mathcal{L}_{color}$ shifts the values to within range and helps its distribution to be close to that of the sibling.

2 IMPLEMENTATION DETAILS

Losses. We compute the contextual loss on the `relu1_2`, `relu2_2`, and `relu3_4` of the pretrained VGG19 network [Simonyan and Zisserman 2014]. For the reconstruction loss, we use layers `conv1_1`, `conv2_1`, `conv3_1`, and `conv4_1` of the pretrained VGG-Face network [Parkhi et al. 2015], and use layers `relu1_2`, `relu2_2`, `relu3_3`, `relu4_3` of the pretrained VGG16 network [Simonyan and Zisserman 2014].

*Sibling Encoder (**E**).* We train three sibling encoders for the three negative film types: *blue-sensitive*, *orthochromatic* and *panchromatic*. The grayscale conversion for blue-sensitive and orthochromatic negatives is unknown and can differ from one photo to another due to different emulsion chemicals. Besides, we want our sibling prediction to be less sensitive to the brightness of the photo since the blue and green channels are often darker than the red channel, and the photos may have been manipulated during the digitization process. Therefore, we augmented the data with color jitter of random brightness factor in $[0.8, 1.8]$, random contrast factor in $[0.8, 1.2]$ and random hue factor in $[-0.03, 0.03]$. We train the three encoders using the Adam [Kingma and Ba 2014] optimizer with a learning rate of 0.0005 and batch size of 4. We train the blue-sensitive and orthochromatic models for 100 epochs and the panchromatic model for 70 epochs.

Camera Response Initialization. Initializing our camera response fitting technique with an identity curve alone (Sec. 4.3) is inadequate when the input and the sibling have substantially different contrast or exposure. We found this to be often true for blue-sensitive and low-contrast photos. To bridge this gap, we first adjust the input image to better match the sibling using *histogram matching* [Burger and Burge 2016]. Specifically, we first convert the sibling to grayscale based on the film type $\mathcal{G}(\tilde{I}_s)$. We weigh the face region more during histogram matching between the input and sibling. We extract the face regions in the input and grayscale sibling images using face parsing [zll 2019] and compute a histogram transform function that brings these two regions into their closest alignment. We then apply the resulting transform on the entire input image to get I', which is then used as the target for the reconstruction losses.

REFERENCES

Wilhelm Burger and Mark J Burge. 2016. *Digital image processing: an algorithmic introduction using Java.* Springer.
Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. 2020. Analyzing and Improving the Image Quality of StyleGAN. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.*
Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).
Omkar M. Parkhi, Andrea Vedaldi, and Andrew Zisserman. 2015. Deep Face Recognition. In *British Machine Vision Conference.*
Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).
zll. 2019. face-parsing.PyTorch. https://github.com/zllrunning/face-parsing.PyTorch