Supplementary Material: Copy-and-Paste Networks for Deep Video Inpainting

Sungho Lee
Yonsei University

Seoung Wug Oh
Yonsei University

DaeYeun Won
Hyundai MNSOFT

Seon Joo Kim
Yonsei University

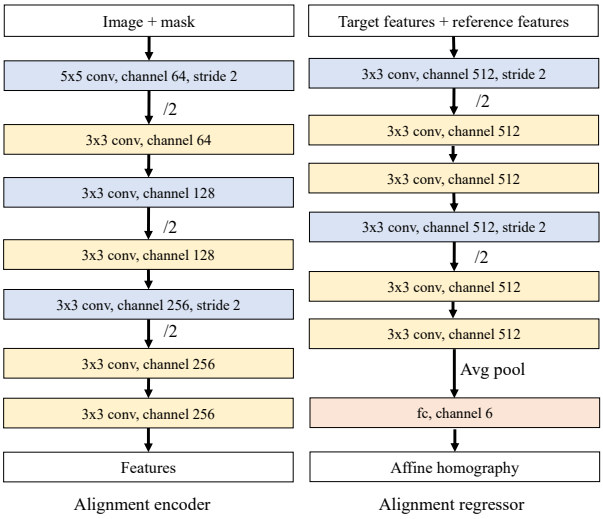

Figure 1: Alignment Network

1. Network Architectures

Fig. 1, 2 illustrate the detailed network architectures of our copy-and-paste networks described in Sec.3 of the main paper. The top and the bottom of each network represent inputs and outputs, respectively. $\boldsymbol{C}_{\text{out}}$, $\boldsymbol{C}_{\text{mask}}$ are outputs of the context matching module described in Sec.3.2. In Fig. 1, "Avg Pool" denotes the average pooling over all location $h \times w$. All the convolutional layers are followed by ReLU [3] for the non-linearity except for the last layer of decoder(Fig. 2).

2. More Results

In addition to Fig.7 in the main paper, we present more qualitative video results[1] for the video object removal task. We compare our results with the methods in [1, 2]. The input videos and masks used in the experiments are DAVIS datasets[4, 5] with shadow annotations provided by [1].

We also present more application results (Fig. 3) for the restoration of under/over-exposed images. The results show that our method improves the texture and the color of the images.

[1] Qualitative video results: https://youtu.be/BKdxR9bQQMU

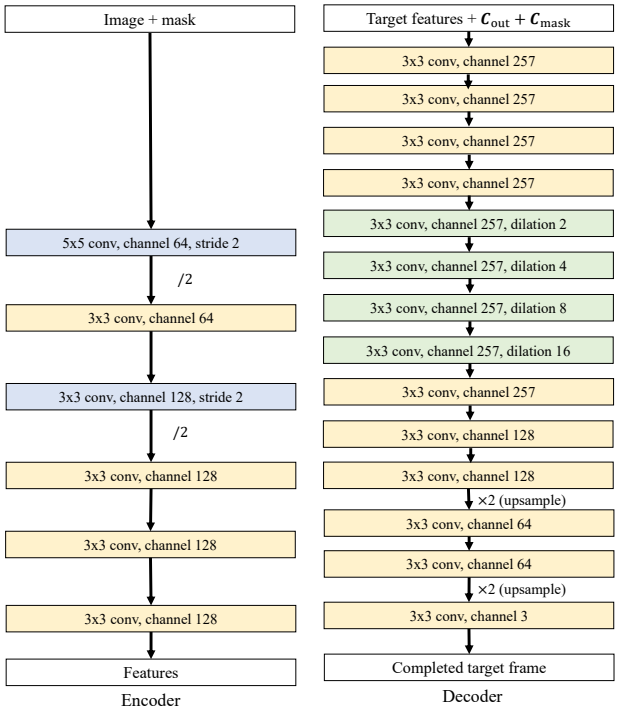

Figure 2: Copy-and-paste Network

References

[1] Jia-Bin Huang, Sing Bing Kang, Narendra Ahuja, and Johannes Kopf. Temporally coherent completion of dynamic video. *ACM Transactions on Graphics (TOG)*, 35(6), 2016. 1

[2] Dahun Kim, Sanghyun Woo, Joon-Young Lee, and In So Kweon. Deep video inpainting. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5792–5801, 2019. 1

[3] Vinod Nair and Geoffrey E Hinton. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the*

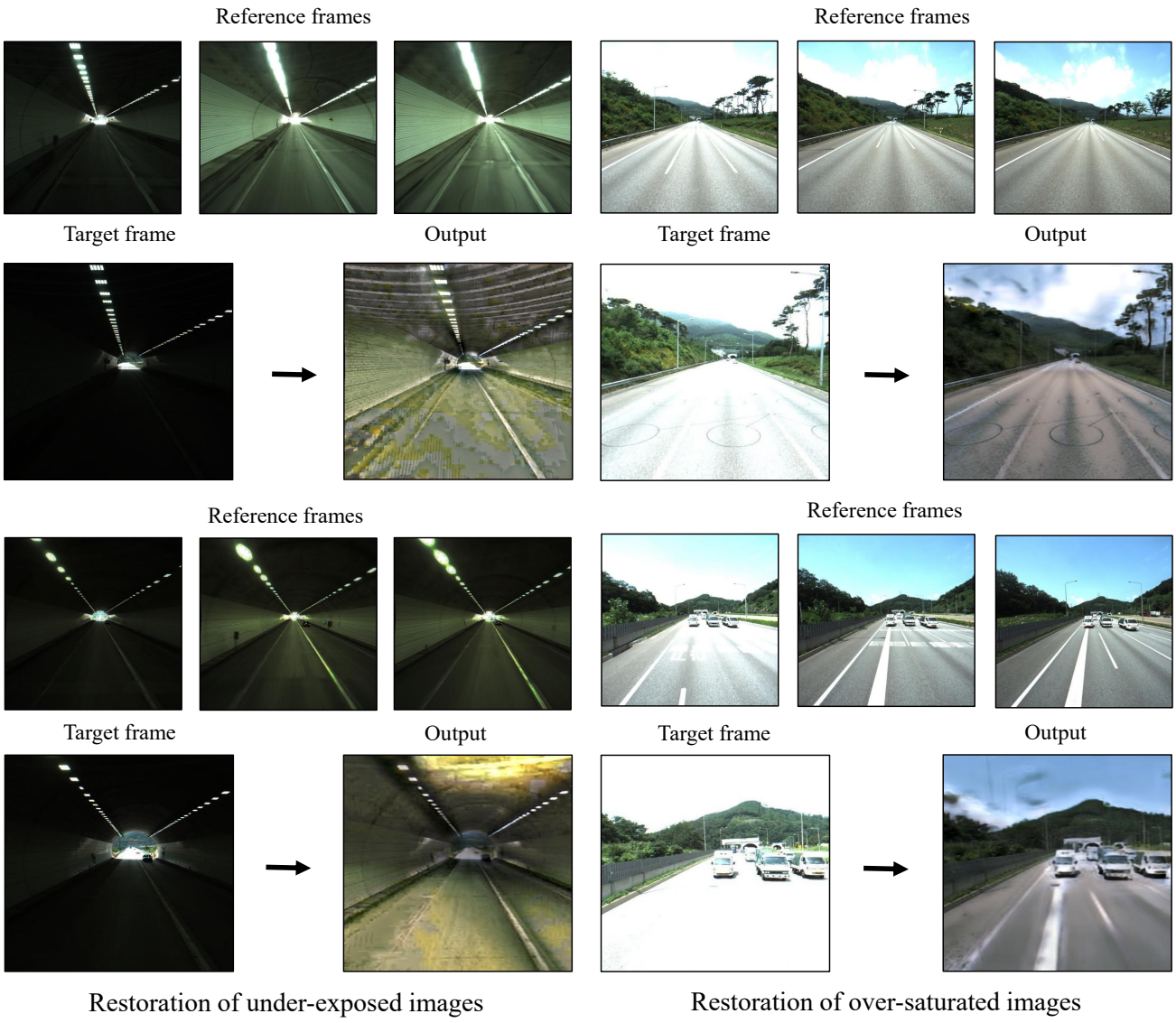

Figure 3: More results of our application for the restoration of under/over-exposed images

27th international conference on machine learning (ICML-10), pages 807–814, 2010. 1

[4] Federico Perazzi, Jordi Pont-Tuset, Brian McWilliams, Luc Van Gool, Markus Gross, and Alexander Sorkine-Hornung. A benchmark dataset and evaluation methodology for video object segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 724–732, 2016. 1

[5] Jordi Pont-Tuset, Federico Perazzi, Sergi Caelles, Pablo Arbeláez, Alex Sorkine-Hornung, and Luc Van Gool. The 2017 davis challenge on video object segmentation. *arXiv preprint arXiv:1704.00675*, 2017. 1