Cut-and-Paste Neural Rendering

Appendix

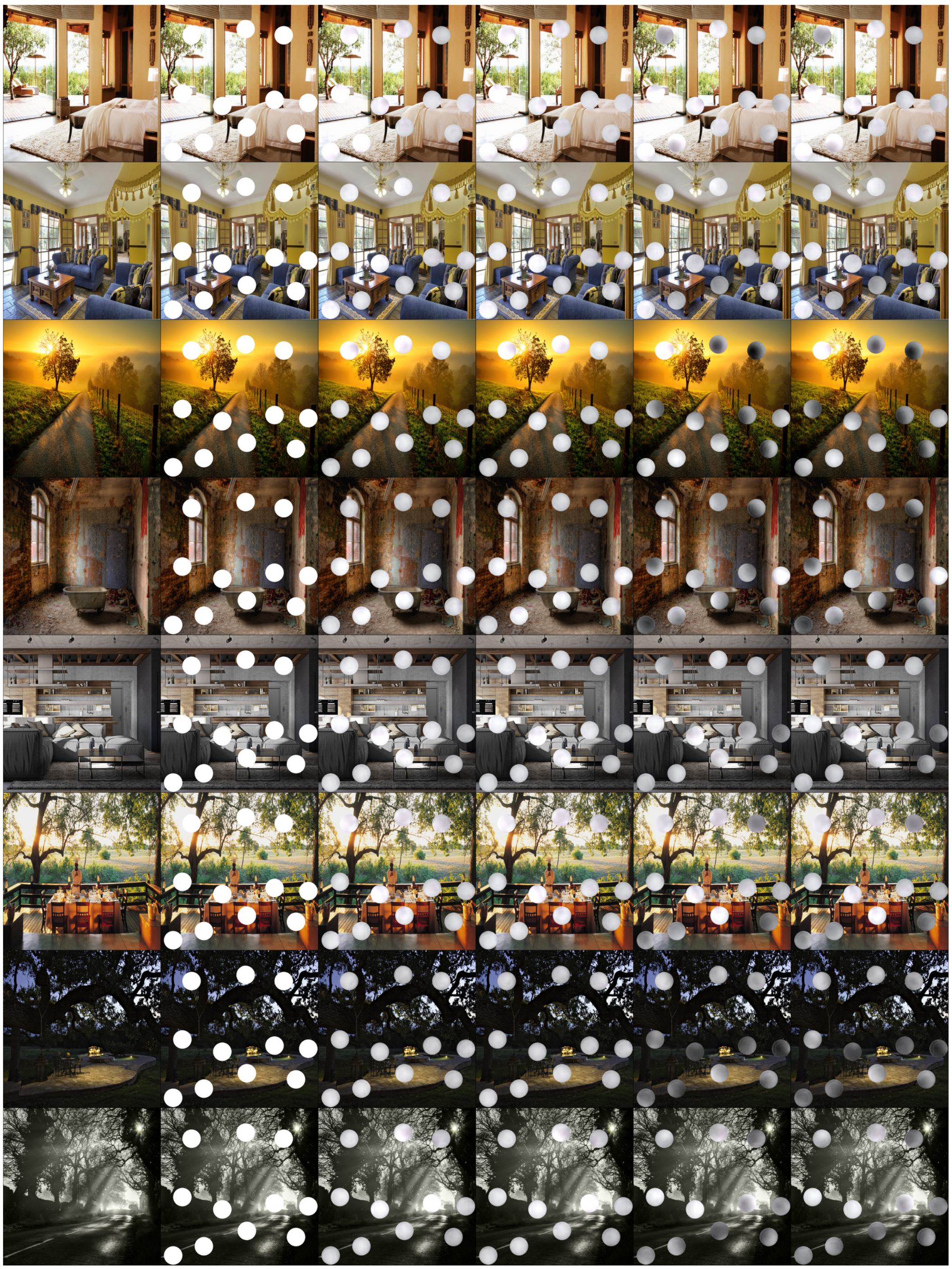

Figure 11: **More examples for ablation.** We show more renderings of spheres from circle cut-outs (see Fig 6) in scenes exhibiting complex lighting. Results improve moving from left to right.

Figure 12: Samples from iharmony dataset (Cong et al., 2020). Image Harmonization methods mostly deals with albedo change.

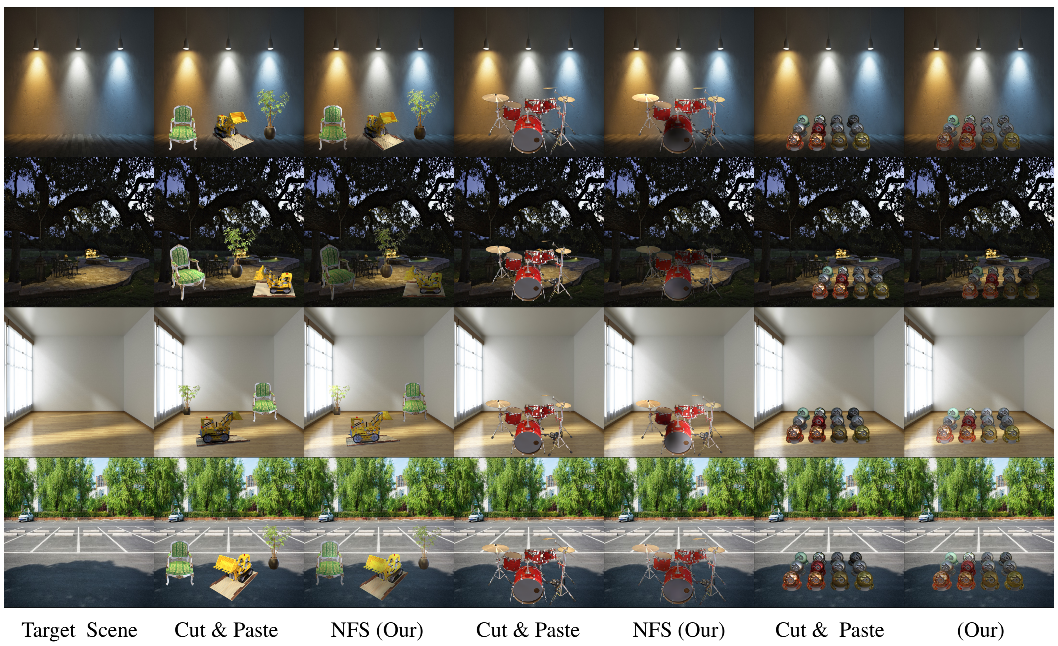

Figure 13: Reshading Complex Materials and Geometry. Our DIP renderings for a lego crane, chair and plant in the third column; drum-kit in the fifth column and a set of 16 materials in the last column. Our method can reshade complicated shapes and materials only from a single image under spatially-varying lighting without requiring 3D geometry of the inserted objects or environment parameters.

Figure 14: **Rendering different poses.** Cut-and-Paste in the top row and our reshaded in the bottom row. DIP reshades chipmunks quite convincingly in a scene exhibiting complex illumination. Note that all frames are rendered simultaneously via the same network and *not* one at a time.

Target Scene | Cut & Paste | Our | Cut & Paste | Our

Figure 15: **Multiple Objects.** We show multiple objects added to a scene at different location.

Target Scene | Cut & Paste | Our | Cut & Paste | Our

Figure 16: **Reshading Cars.** Two different cars under spatially-varying lighting scenes.

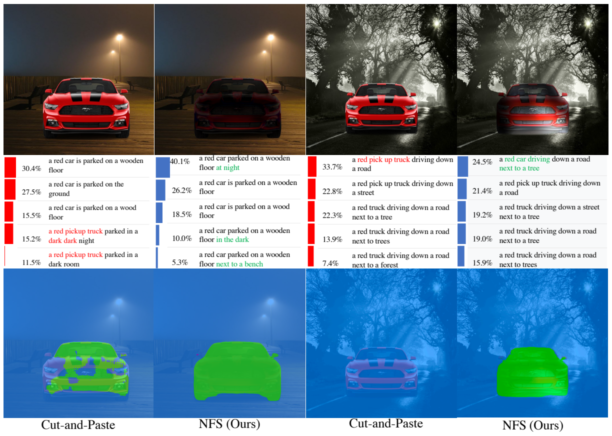

Figure 17: **Consistent Inferences on Downstream Tasks.** DIP images produces consistent and accurate captions (middle row) and segmentation maps (last row). In both scenes DIP cars appear convincingly realistic as if they were already part of the capture. Notice how the segmentation network has completely failed to detect or segment the second car. Also notice the accurate change in specular highlights for both the cars rendered from DIP.

PARADIGMS

This paradigm based decomposer works as follows. An encoder, with a resnet structure, produces an image code. This is decoded (again with a resnet structure) into an albedo field, a shading field and a gloss field. The albedo field is colored. Shading and gloss field are colorless. The training process accepts albedo, shading and gloss paradigms (generated as samples in advance and cached) and real images from Zhou et al. (2017). The real images are given without any of their ground truth. We use them to ensure our image decomposition inferences when recomposed back are able to produce an image that very much looks like real images. We update our decomposer with this residual loss after training with about 50k paradigm samples.

There are four losses. The decomposer must decompose a fake image (constructed out of randomly selected albedo, shading and gloss paradigms) into its correct, known paradigms (mixed L1/L2 loss). The albedo, shading and gloss fields constructed from a real image using our decomposer must combine into that real image (mixed L1/L2 loss). The remaining two losses are adversarial. The albedo field constructed from a real image must fool a classifier trained to distinguish between fake albedo fields and those constructed from real images. Similarly, the shading field constructed from a real image must fool a classifier trained to distinguish between fake shading fields and those constructed from real images.

The Paradigm I constructs albedo, shading and gloss filed. The Paradigm II constructs albedo and shading only without gloss filed.