Supplement to Deep Optics for Single-shot High-dynamic-range Imaging

Christopher A. Metzler   Hayato Ikoma   Yifan Peng   Gordon Wetzstein
Stanford University
{cmetzler, hikoma, evanpeng, gordon.wetzstein}@stanford.edu

This supplement contains additional reconstruction results, both from simulated and experimental data, and an image illustrating the spatially variant nature of the PSF. Additionally, the included "HDR_Reconstructions" directory contains both ground-truth images at multiple exposures and high-dynamic-range reconstructions of our experimental scenes.

Note that both in the paper and the supplement, when computing the PSNR we first normalized by the reference and the reconstruction by the maximum of the reference image. We then used the following definition for PSNR,

$$\mathrm{PSNR}(u, v) = 10 \log_{10}\left(\frac{1}{\|\mathbf{u} - \mathbf{v}\|_2^2 / n}\right),$$

where n is the length of $\mathbf{u}$ and $\mathbf{v}$.

In computing the HDR-VDP-2 loss we use 30 pixels per degree and the 'sRGB-display' option. The other parameters are set to their default values.

References

[1] Gabriel Eilertsen, Joel Kronander, Gyorgy Denes, Rafał K Mantiuk, and Jonas Unger. Hdr image reconstruction from a single exposure using deep cnns. *ACM Transactions on Graphics (TOG)*, 36(6):178, 2017.

[2] Mushfiqur Rouf, Rafał Mantiuk, Wolfgang Heidrich, Matthew Trentacoste, and Cheryl Lau. Glare encoding of high dynamic range images. In *Proc. CVPR 2011*, pages 289–296, 2011.

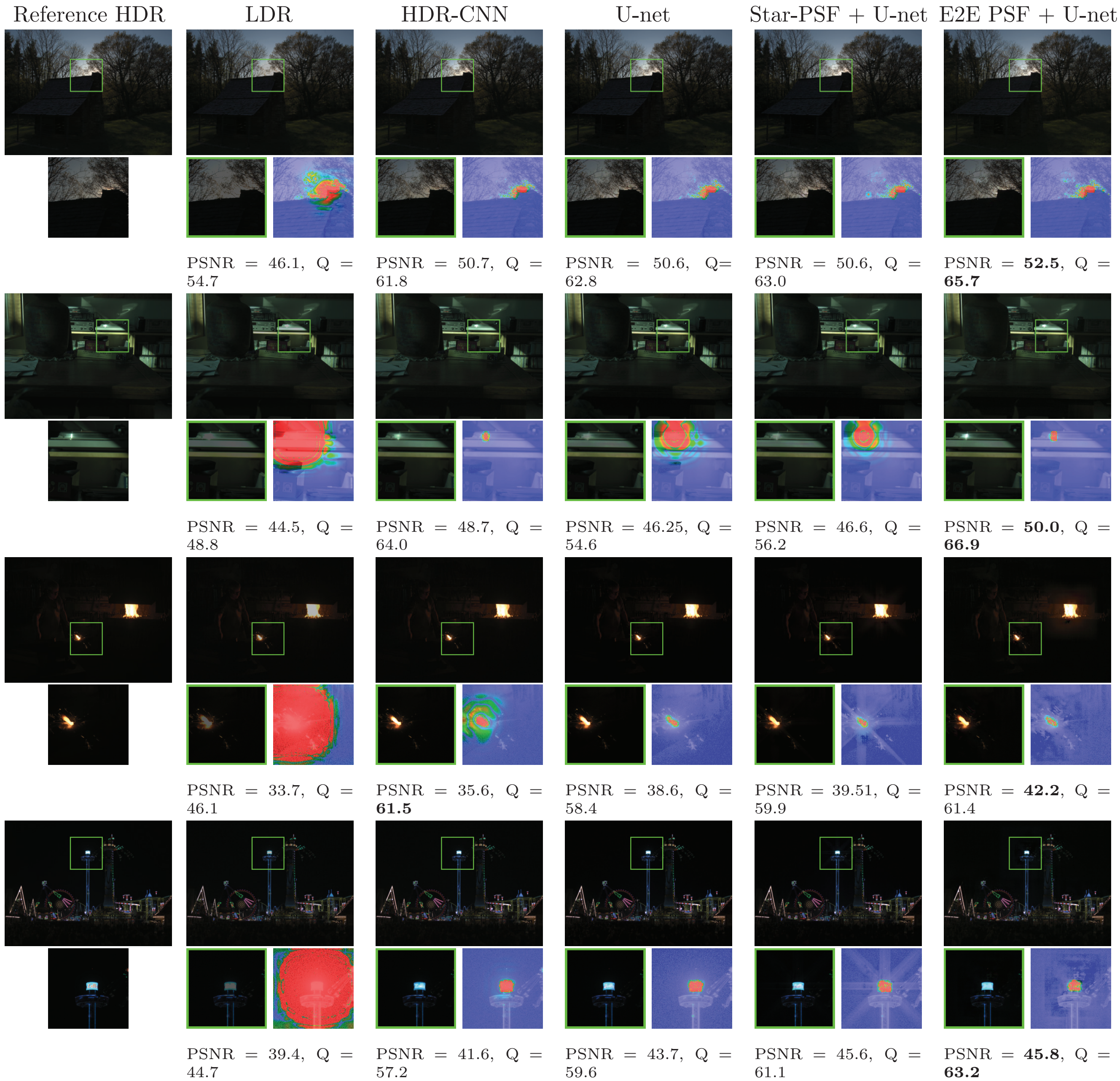

Figure 1: Comparison of various single-shot HDR imaging approaches. In all examples, the whole images are displayed at -1 stop and regions of interest at -3 stops. In the columns, we show the ground truth HDR image, a corresponding LDR image, Eilertsen et al.'s CNN-based reconstruction applied to the LDR image (HDR-CNN) [1], a slightly simpler U-Net applied to the LDR image, the star-shaped PSF proposed by Rouf et al. [2] with a U-Net reconstruction, and our end-to-end deep optics approach (E2E). In all three examples, the peak signal-to-noise ratio (PSNR) and also observed image quality (see insets) are best for our approach. Color-coded insets also show probabilities of perceiving the difference between the reconstructions and the ground truth HDR images, as computed by the HDR-VDP-2 visible differences predictor. Again, our approach qualitatively and quantitatively (evaluated with HDR-VDP-2 Q value) outperforms other approaches. Note: The top three scenes are from test set while the bottom scene is from the validation set. All results in the main text are from the test set.

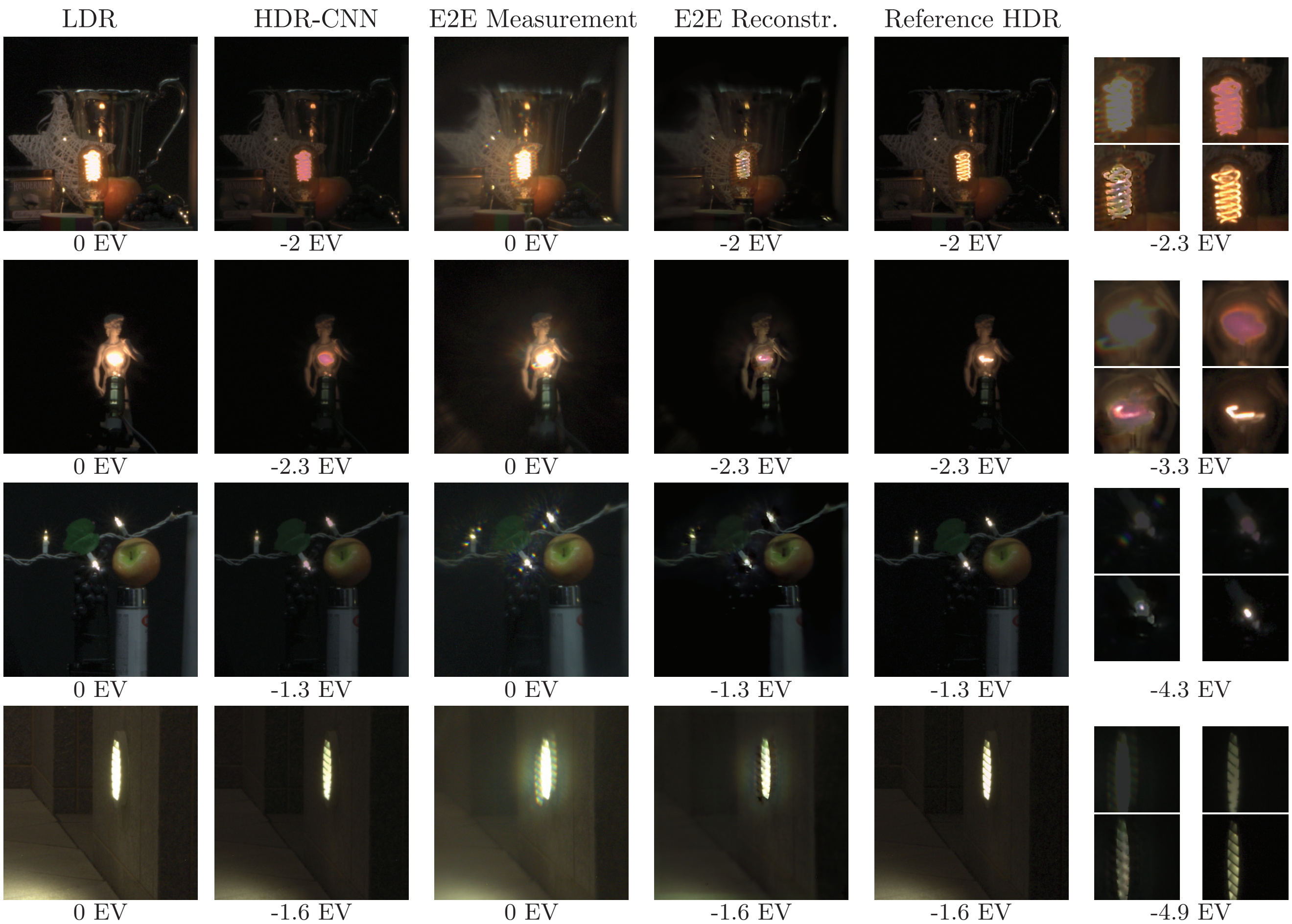

Figure 2: Experimental data of three indoor scenes (top three rows) and one outdoor scene at night (bottom row). The limited dynamic range of the sensor loses details in the brighter parts of the captured LDR image (first column) as compared with the reference HDR image (fifth column). A CNN operating directly on the LDR images hallucinates brighter content in these saturated parts, but it is missing the detail (second column). The measurements captured with our prototype camera optically encode this detail in the image by superimposing several scaled and shifted copies of the image on itself (third column). This information is used by our CNN to recover the missing parts of the scene while digitally removing the image copies (fourth column). The closeups show that deep optics is more successful in recovering bright detail of HDR scenes than other single-shot HDR imaging approaches (right column).

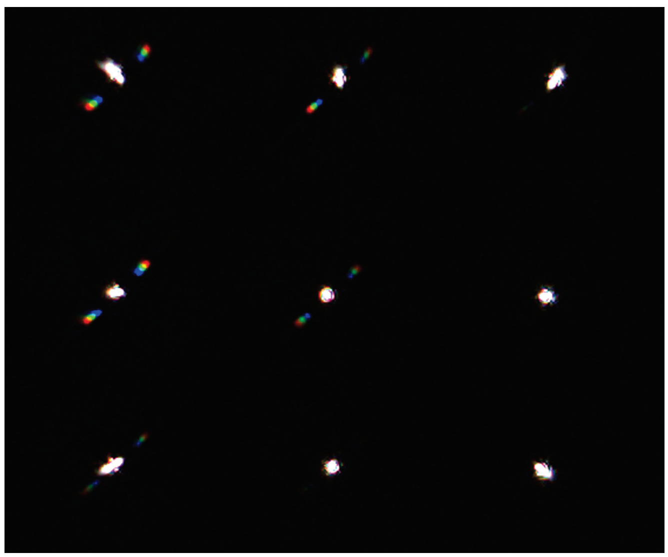

Figure 3: Image of nine point sources from different regions in the camera's field of view (FoV). To deal with the spatially variant PSF, we crop the FoV to the center and train our network on the PSF there.