Supplementary Information: Learned Large Field-of-View Imaging With Thin-Plate Optics

YIFAN PENG*, Stanford University, USA
QILIN SUN*, King Abdullah University of Science and Technology, Saudi Arabia
XIONG DUN*, King Abdullah University of Science and Technology, Saudi Arabia
GORDON WETZSTEIN, Stanford University, USA
WOLFGANG HEIDRICH, King Abdullah University of Science and Technology, Saudi Arabia
FELIX HEIDE, Princeton University, USA

ACM Reference Format:
Yifan Peng, Qilin Sun, Xiong Dun, Gordon Wetzstein, Wolfgang Heidrich, and Felix Heide. 2019. Supplementary Information: Learned Large Field-of-View Imaging With Thin-Plate Optics. *ACM Trans. Graph.* 38, 6, Article 219 (November 2019), 15 pages. https://doi.org/10.1145/3355089.3356526

1 ADDITIONAL DETAILS ON ZEMAX OPTIMIZATION

The surface parameters of our two prototype lenses are described in Table 1, with their optical layouts and spot diagrams shown in Figure 1 and Figure 2, respectively.

We use the surface equation of even aspherical surfaces in Zemax for both designs. We observe that the peak intensity of PSF of the dual-surface lens design (referred to as A) is higher than that of the single-surface lens design (referred to as B). This is reasonable since using two surfaces theoretically increases the designs degree-of-freedom for aberration correction.

The thickness deviation of the two prototype lenses, i.e. 10 mm v.s. 3 mm, is mainly because we require the solid substrate support in the prototyping process for dual optical surfaces. Theoretically, the planar substrate shows a very minimal effect on the lens quality. As mentioned in the main text, our design can be applied to multiple surfaces configuration which offers more degrees of freedom for diverse lens designs.

Table 1. Surface parameters of our two prototype lenses.

	3 mm design		**10 mm design**	
	front surface	back surface	front surface	back surface
Radius (mm)	infinity	-21.399	226.656	-23.164
Thickness (mm)	3	-	10	-
Material	PMMA	-	PMMA	-
K	0	0	0	0
A4	0	1.080e-5	1.492e-5	1.696e-5
A6	0	-1.828e-8	1.139e-7	2.631e-8
A8	0	2.981e-10	-5.886e-10	1.240e-10
A10	0	-7.834e-13	0	-2.977e-13

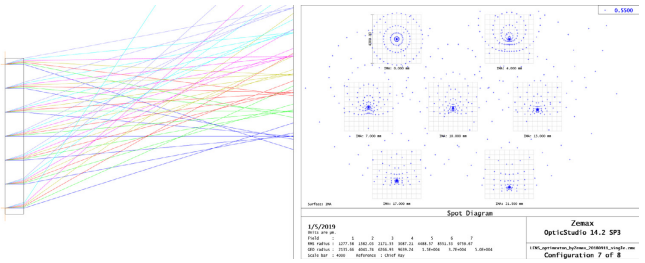

Fig. 1. Optical layout (left) and spot diagrams (right) of the prototype lens with a substrate thickness of 3 mm.

2 FABRICATION DETAILS

Photographs of our two prototype lenses are shown in Figure 3. As mentioned in the main text, the fabrication accuracy of our deep Fresnel lens surface depends on the specifications of the diamond turning tip. Illustrated in Figure 4, the head of the turning tip has a rounded region that results in an imperfect deep valley for each ring. Although the rounded radius has only a size of 16 μm, a part of light rays directly travels through it without being altered by the correct phase modulation. This is a crucial source of contrast loss of the measurement images.

3 ADDITIONAL DETAILS ON DECONVOLUTION

In this section, we provide additional detail on the proposed image reconstruction method. Recent advances in image deconvolution have the goal of including statistical prior knowledge, such as

*joint first authors.

Authors' addresses: Yifan Peng, evanpeng@stanford.edu, Stanford University, Stanford, CA, USA; Qilin Sun, qilin.sun@kaust.edu.sa, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia; Xiong Dun, xiong.dun@kaust.edu.sa, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia; Gordon Wetzstein, gordon.wetzstein@stanford.edu, Stanford University, Stanford, CA, USA; Wolfgang Heidrich, wolfgang.heidrich@kaust.edu.sa, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia; Felix Heide, fheide@cs.princeton.edu, Princeton University, Princeton, NJ, USA.

0730-0301/2019/11-ART219 $15.00
https://doi.org/10.1145/3355089.3356526

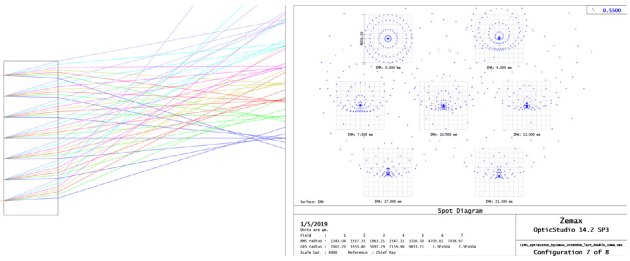

Fig. 2. Optical layout (left) and spot diagrams (right) of the prototype lens with a substrate thickness of 10 mm.

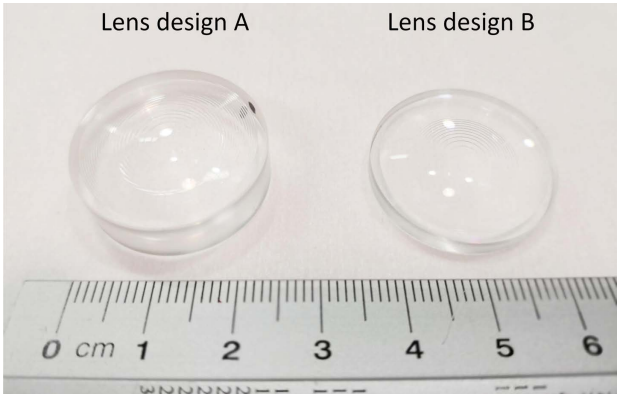

Fig. 3. Photographs of our prototype lenses with dual optical surface (left) and single optical surfaces (right).

gradient statistics [Chan et al. 2011; Goldluecke and Cremers 2011; Krishnan and Fergus 2009]. At its core, the proposed deconvolution approach also relies on such prior knowledge – however, learned from captured image data.

Patch-based Processing. In our approach, we assume the blur is more dependent on field-of-view rather than depth in our case. Inspired by the prior work [Yue et al. 2015], we divide the image radially to a few rings, and then unwrap each ring to a rectangular image patch, except for the central round region. To this end, the blur becomes more spatial invariant on each image patch such that we could drastically increase the efficiency of deconvolution, compared to the regular square block partitioning. After the processing, we wrap the patches and then stitch them back to yield the final reconstruction result. We will release the code for our implementation including all processing steps and hyperparameters.

Cross-channel Deconvolution. As mentioned in the main text, traditionally, camera systems preserve color fidelity by designing a complex stack of refractive and diffractive optics. To compare our method in terms of color fidelity, we compare it against the cross-channel prior which enforces gradient consistency of edges on the images of different color channels [Heide et al. 2016; Peng et al. 2015; Sun et al. 2017]. For this baseline method, for each individual patch, the optimization problem of resolving the latent image i_c given the

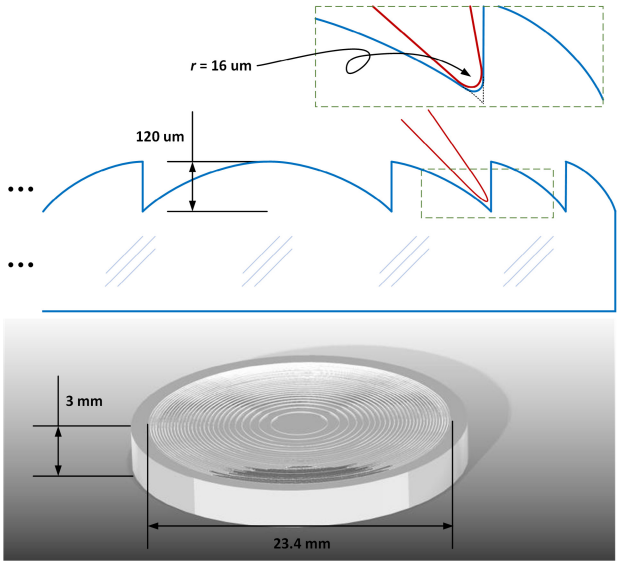

Fig. 4. Illustration of the rounded head region of the turning tip (top) and the 3D layout of our prototype lens (bottom).

blurry input b_c and the estimated kernel p_c, is described as:

$$\mathrm{i}_c = \mathrm{argmin}_{\mathrm{i}_c} \frac{\mu}{2}\|\mathrm{b}_c - \mathrm{p}_c\mathrm{i}_c\|_2^2 + \beta\|\mathrm{Di}_c\|_1 + \gamma \sum_{l\neq c} \|\mathrm{Di}_c - \mathrm{Di}_l\|_1, \quad (1)$$

where the first term is the data fitting term, μ_c is its weight for channel c. D is an operator that takes gradients of the image. Thus, the second term is a total variation regularization while the third term is a cross-channel regularization, with their weights of β and γ, respectively. Both terms enforce gradient priors of natural images.

This cross-channel deconvolution can be implemented in both blind and non-blind manner, that at each step, the regularizations result in a non-linear optimization problem that can be solved by introducing slack variables for the ℓ_1 terms and using proximal operators [Boyd et al. 2011] to turn the ℓ_1 terms into shrinkage operators. We refer the readers to [Heide et al. 2016] for details.

4 ADDITIONAL COMPARISON OF GENERATIVE END-TO-END DESIGN

In this section, we show additional results demonstrating that an end-to-end training with the proposed reconstruction framework is not feasible with existing methods. As a result, the following experiments validate the proposed co-design relying on an engineered intermediate metric.

As existing end-to-end design methods, such as [Sitzmann et al. 2018], only support optics where the paraxial approximation holds, we cannot directly apply such methods to our wide FOV design problem. Instead, we compare the proposed GAN reconstruction framework applied to the task of a full-spectrum focusing lens. Here, we consider the problem of designing a phase plate which focuses light for the visible spectrum. In particular, we adopt the setting from Sitzmann et al. [2018] and optimize for a lens with a lens-to-sensor distance of 25 mm and an aperture of 5 mm. Instead of

optimizing for the full spectrum, we optimize for the wavelengths 460 nm, 550 nm, and 640 nm. We approximate a sensor readout noise with standard distribution drawn uniformly between 0.001 and 0.02. We use a Zernike parametrization [Sitzmann et al. 2018] to regularize the optimization.

For reproducibility by readers, we note that the code from the official repository is incorrect, and requires the PSF to be flipped using *tf.reverse* in the implementation of the Wiener deconvolution.

In the following, we describe five experiments with and without the GAN reconstruction approach from the main draft. In particular, we compare end-to-end training with a low-parameter Wiener filter against end-to-end training with the proposed GAN recovery method, and against fine-tuning the GAN on pre-optimized PSFs. The GAN reconstruction experiments are performed as follows. At each iteration of training, the Zernike lens model produces a 1,356×1,356 sensor image. Then random 256×256 crops are taken of this image, along with the corresponding 256x256 crop of the ground truth, and these are fed as training input into the GAN. If we are training end-to-end then losses are back-propagated to both the GAN and the lens model. If we are fine-tuning the GAN then losses are only back-propagated to the GAN and the lens model stays fixed. For the optimization, we use Adam with a learning rate of 10^{-4} and momentum of 0.99.

We used the high-resolution images from [Jegou [n. d.]] for optimizing the phase masks presented below (Table 2).

Experiment 1 = Wiener deconvolution end-to-end. This approach is the method proposed in [Sitzmann et al. 2018]. The lens model is jointly trained with a Wiener deconvolution operation, whose only trainable parameter is γ. The PSF of the lens model is convolved with the input image and random noise is added in. This simulates the image captured by the sensor. To recover the true image, Wiener deconvolution uses the known PSF of the lens model along with an estimate of the noise (determined by γ) to deconvolve the sensor image.

The pipeline: Input Image → Lens (Trainable) + Noise → Sensor Image → Wiener Deconvolution (Trainable) → Output Image

Experiment 2 = GAN reconstruction end-to-end. The lens model is jointly trained with GAN reconstruction, which has orders of magnitudes more trainable parameters than the Wiener recovery. The generator receives a loss penalty from the discriminator (critic) and the perceptual loss.

This experiment had poor results as output images still looked very blurry. Inspection of the PSF shows that the lens model has not been optimized at all, and it seems that the pipeline is relying solely on the GAN.

The pipeline: Input Image → Lens (Trainable) + Noise → Sensor Image → GAN Reconstruction (Trainable) → Output Image

Experiment 3 = Fine-tuning the GAN using PSF from Experiment 1. The lens model is taken from Experiment 1 and is fixed. Only the GAN is being trained in this experiment.

We note that this experiment had the best results of the five experiments. Output images were sharp and clearer than all of the other experiments.

The pipeline: Input Image → Lens (Fixed) + Noise → Sensor Image → GAN Reconstruction (Trainable) → Output Image

Experiment 4 = Fine-tuning the GAN using PSF from Experiment 2. The lens model is taken from Experiment 2 and is fixed. Only the GAN is being trained in this experiment.

This experiment had poor results as the network was not able to recover the large blur due to poor focusing of the optimized lens.

The pipeline: Input Image → Lens (Fixed) + Noise → Sensor Image → GAN Reconstruction (Trainable) → Output Image

Experiment 5 = Fine-tuning the GAN using initial guess PSF. The lens model is the initial guess for the Zernike system, from [Sitzmann et al. 2018].

We note that this experiment achieved the same results as Experiment 4, with the end-to-end designed lens.

The pipeline: Input Image → Lens (Fixed) + Noise → Sensor Image → GAN Reconstruction (Trainable) → Output Image

Assessment. Experiment 2 validates that joint training with GAN reconstruction is more difficult than joint training with Wiener deconvolution. This is because of the large number of trainable parameters that a GAN has compared to that of Wiener filter– single trainable parameter, so not enough training signal passes through the GAN to the lens model. Furthermore, experiments 4 and 5 highlight the importance of having both a good optimized lens model and a high-quality deconvolution method. Despite the GANs strength, it cannot perform well if the lens model is poor.

However, experiment 3 shows that good results can be obtained by first optimizing the lens model using the Wiener filter as a proxy that can be efficiently optimized, then replacing the Wiener filter with a GAN and optimizing the GAN while fixing the obtained lens model. Using the Wiener filter forces the lens model to be optimized because the Wiener filter is a less powerful deconvolution method than the GAN.

5 ADDITIONAL ASSESSMENTS

Comparison against alternative reconstruction algorithms. In addition to the discussion in Section 8 of the main text, we present the full resolution images reconstructed by U-net, pixel-to-pixel and our method, respectively, shown in Figure 7. We observe that the pix2pix outperforms U-net on real-world data, while still suffering from severe artefacts, e.g. the sky region. The proposed method mitigates these artifacts.

Comparison against off-the-shelf well-corrected lens. Figure 8 shows comparisons of our prototype lens compared against a well-corrected commercial SONY FR 1.4/50 compound lens. The proposed method achieves high image quality, when compared to this compound lens, while showing a larger depth-of-field. This is due to the fact our effective aperture that contributes to the central peak of PSF is smaller than that of a regular lens under the same clear aperture setting. Furthermore, we are training our method on 2D images displayed and captured on an LCD monitor, without providing depth information. However, our recovery shows robustness on this semi-synthetic data acquisition process, and in turn, shows the possibility of extending depth of field.

Table 2. End-to-end training for various training configurations for an RGB collimator realized with a single diffractive surface, including full end-to-end training and fine-tuning using proxy models or the initial guess (Fresnel lens centered on green channel).

	Experiment 1 Wiener End-to-End	Experiment 2 GAN End-to-End	Experiment 3 Fine-tune GAN on (1)	Experiment 4 Fine-tune GAN on (2)	Experiment 5 Fine-tune GAN on Zernike Init.
PSNR [dB]	23.9	22.8	25.7	21.9	21.9

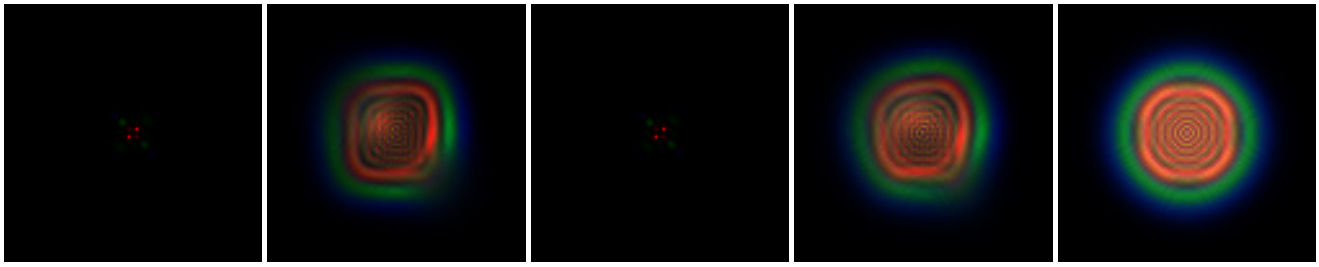

Fig. 5. PSFs for **Experiments 1-5** for the end-to-end design comparisons, see text.

Comparison against Fresnel lens and aspherical lens. Using the same setup as in comparison above, Figure 9 shows results captured by an aspherical lens, and our prototype lens. Note, the 100 mm Fresnel lens is fabricated using state-of-the-art photolithography [Peng et al. 2015]. The 50 mm lens is realized by stacking two 100 mm Fresnel lenses together. This is because the fabrication resources available to us can only realize a pixel pitch down to 1 μm, which is far from sufficient to create a well-defined Fresnel lens with an aperture diameter of 8 mm at a focal length of 50 mm at a principle wavelength of 550 nm.

Although featuring ultra-thin form factors, conventional Fresnel lenses suffer from severe chromatic aberration because of diffraction. Moreover, a conventional aspherical lens suffers from the most severe off-axis aberration among all evaluated lenses.

Outlier patches for hallucination experiments. As mentioned in Section 8 of main text, Figure 10 presents additional outlier image patches that are selected from the validation set. Note again, these image patches are considered as outlier images, i.e., worst case scenarios. We do not observe noticeable artifacts, beyond blur and lack of contrast in some regions, indicating that our learn reconstruction does not hallucinate unwanted details.

Reconstruction comparison of different perceptual loss settings. Figure 11 shows comparisons of results reconstructed from the same network but with slightly different settings regarding perceptual loss. Note that with applying VGG loss of single layer the reconstructed result visually show higher image contrast and more aggressive color tone. There may appear some residual artifacts in dark regions. Instead, with applying VGG loss of multiple layers the reconstructed result of same input visually show more gentle tone mapping with less residual artifacts in dark regions. Our learned reconstruction has been validated robust under both settings to produce high quality results.

Additional Imaging Results. Figure 12 shows additional comparisons against the pinhole imaging at different exposure levels. These results further validate the discussion in Section 9.2 of the main text. Refer to the caption of each figure for details.

More measurement and reconstruction results of our prototype lens are shown in Figure 13, 14, 15, and 16.

Additional Web-based Viewer. In addition to the selected image crops presented in the main text and this document, we have provided an interactive web-page that users can browse and switch between the measurement and the visualization result via mouse-over. Please refer to the **.html** file in the package.

6 SUPPLEMENTARY VIDEO

This manuscript includes a supplementary video (**.mpeg**) that provides a brief overview of this work.

REFERENCES

Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. 2011. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine Learning* 3, 1 (2011), 1–122.

Stanley H Chan, Ramsin Khoshabeh, Kristofor B Gibson, Philip E Gill, and Truong Q Nguyen. 2011. An augmented Lagrangian method for total variation video restoration. *IEEE Transactions Image Processing (TIP)* 20, 11 (2011), 3097–3111.

Bastian Goldluecke and Daniel Cremers. 2011. Introducing total curvature for image processing. In *Computer Vision (ICCV), IEEE International Conference on*. IEEE, 1267–1274.

Felix Heide, Qiang Fu, Yifan Peng, and Wolfgang Heidrich. 2016. Encoded diffractive optics for full-spectrum computational imaging. *Scientific Reports* 6 (2016).

Herve Jegou. [n. d.]. INRIA Holidays dataset. http://lear.inrialpes.fr/~jegou/data.php

Dilip Krishnan and Rob Fergus. 2009. Fast image deconvolution using hyper-Laplacian priors. In *Advances in Neural Information Processing Systems*. NIPS, 1033–1041.

Yifan Peng, Qiang Fu, Hadi Amata, Shuochen Su, Felix Heide, and Wolfgang Heidrich. 2015. Computational imaging using lightweight diffractive-refractive optics. *Optics Express* 23, 24 (2015), 31393–31407.

Vincent Sitzmann, Steven Diamond, Yifan Peng, Xiong Dun, Stephen Boyd, Wolfgang Heidrich, Felix Heide, and Gordon Wetzstein. 2018. End-to-end optimization of optics and image processing for achromatic extended depth of field and super-resolution imaging. *ACM Transactions on Graphics (TOG)* 37, 4 (2018), 114.

Tiancheng Sun, Yifan Peng, and Wolfgang Heidrich. 2017. Revisiting Cross-channel Information Transfer for Chromatic Aberration Correction. In *Proc. International Conference on Computer Vision (ICCV)*. 3248–3256.

Tao Yue, Jinli Suo, Jue Wang, Xun Cao, and Qionghai Dai. 2015. Blind optical aberration correction by exploring geometric and visual priors. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 1684–1692.

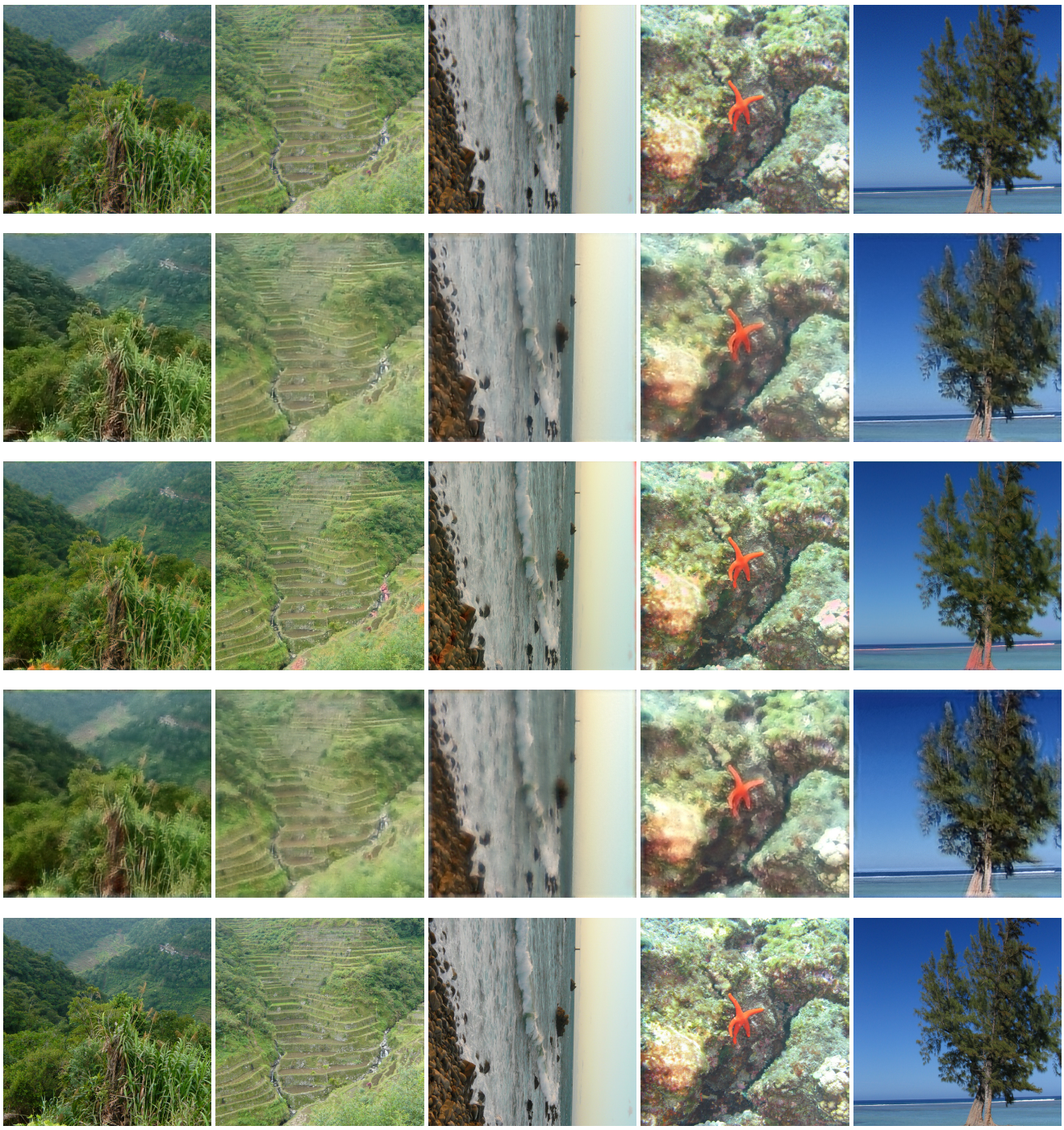

Fig. 6. Reconstruction results for **Experiments 1-5** for the end-to-end design comparisons. The images for each experiment are shown in rows 1-5, with each row corresponding to the respective experiment as described in the text. The bottom row represents the ground truth images.

Fig. 7. Comparison results reconstructed by U-net, Pixel-to-pixel, and ours. The pix2pix outperforms U-net on real world data, while still suffers from non-trivial visual artifacts, e.g. in the sky region. However, our proposed reconstruction mitigates these artifacts.

Fig. 8. Comparison results of our prototype lenses (left) and the SONY FE 1.4/50 lens (right) with a slightly different field of view due to the deviation of focal lengths. We show both measurement and reconstruction of ours side-by-side. With this numerical aperture, the SONY lens shows noticeable depth-dependent defocus effect. Note, our lens is a single element thin plate while the counterpart lens has an optical stack of more than a dozen of optical elements.

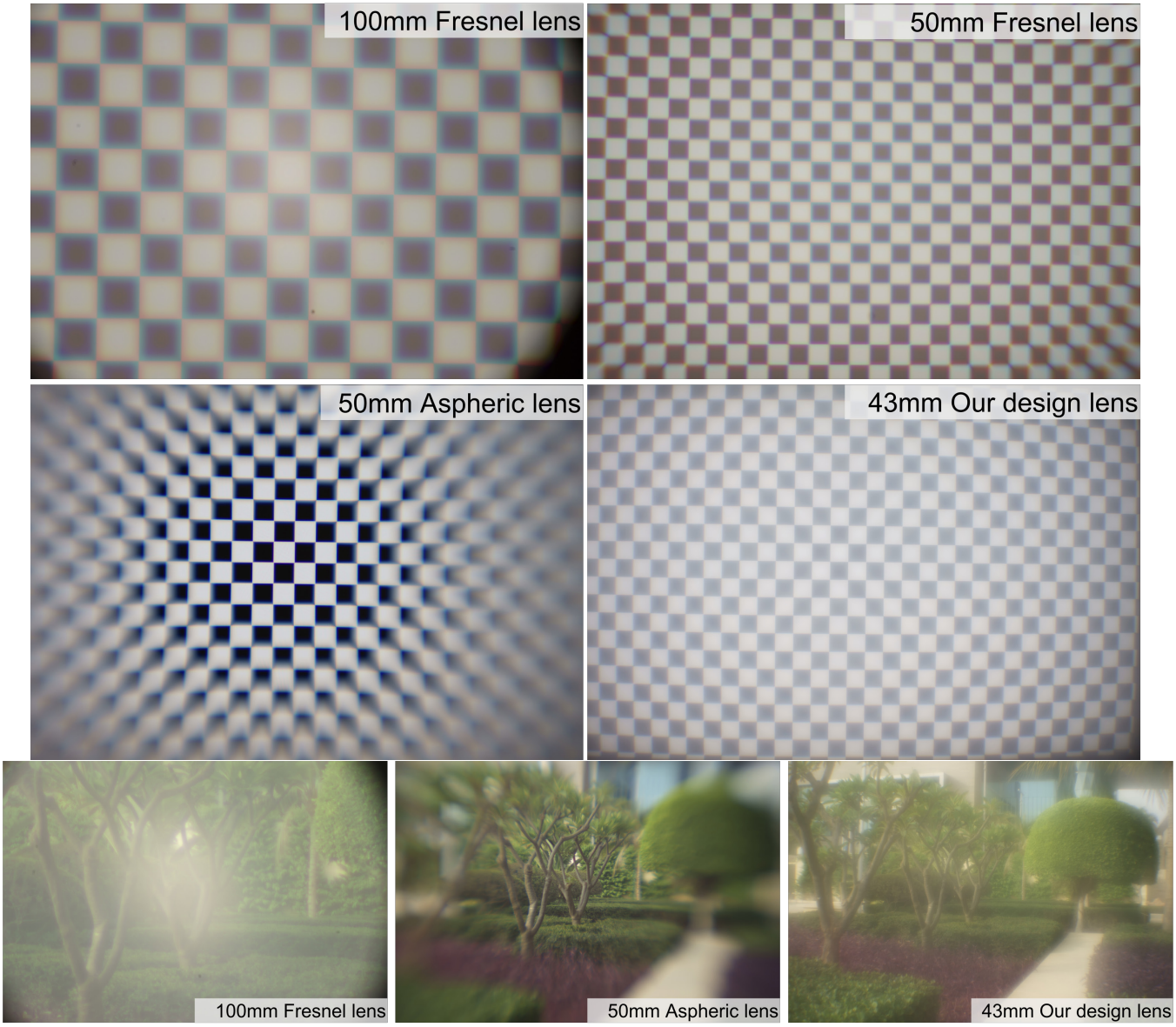

Fig. 9. Comparison results of a chessboard target (top) and a wild scene (bottom) captured by Fresnel lenses (aperture 8mm), an aspherical lens(aperture 23.5mm), and our prototype lens(aperture 23.5mm).

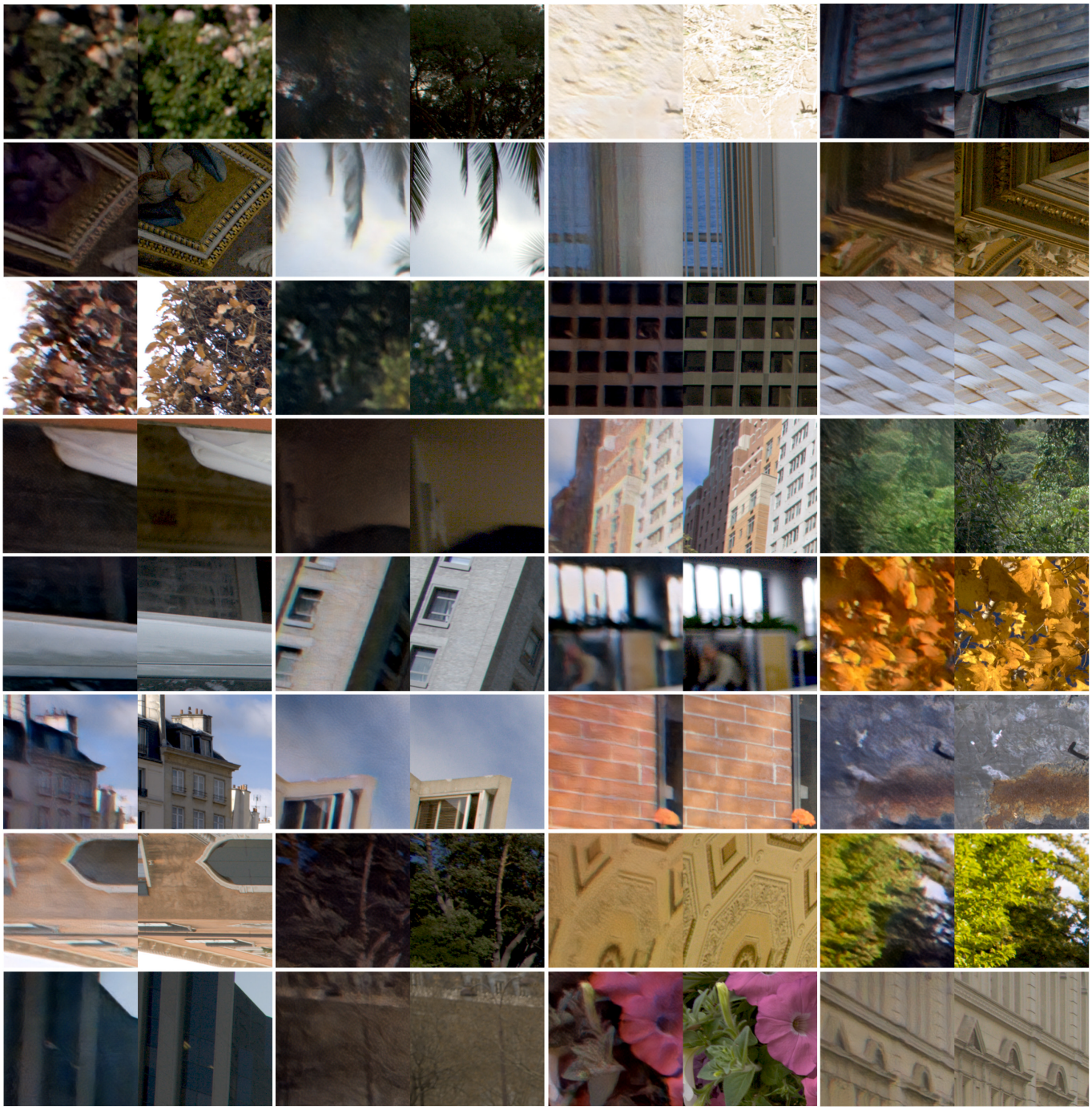

Fig. 10. Outlier image patches from validation set for hallucination assessment. For each pair, we show the recovered image patch on the left while the ground truth image patch on the right.

Fig. 11. Comparison results reconstructed using our network but with different settings regarding perceptual loss functions.

Fig. 12. Additional comparison results captured by our lens and a pinhole for low light imaging scenarios. For each example we present the exposure time and ISO for readers' information.

Fig. 13. Additional results captured by our prototype lenses (1-4). For each example we present side-by-side the measurement and the recovery images.

Fig. 14. Additional results captured by our prototype lenses (2-4). For each example we present side-by-side the measurement and the recovery images.

Fig. 15. Additional results captured by our prototype lenses (3-4). For each example we present side-by-side the measurement and the recovery images.

Fig. 16. Additional results captured by our prototype lenses (4-4). For each example we present side-by-side the measurement and the recovery images.