Supplemental Material for Edge-aware Bidirectional Diffusion for Dense Depth Estimation from Light Fields

Numair Khan[1]
http://cs.brown.edu/~nkhan6/
Min H. Kim[2]
http://vclab.kaist.ac.kr/
James Tompkin[1]
https://www.jamestompkin.com

[1] Brown University
Providence, USA
[2] KAIST
Daejeon, South Korea

1 Supplemental Material

We include additional discussion covering the benefits over naive diffusion, consistency over views within the 4D light field, tolerance to depth label errors and edge blur, robustness to hyperparameter variation, details of dataset preprocessing, and an example of textures within dark backgrounds in the Stanford dataset (Section 1.1). Next, we present error maps comparing reprojection loss versus our bidirectional diffusion approach (Section 1.3), and error maps versus ground truth for the HCI dataset (Section 1.4). Finally, we show additional qualitative results on the Stanford dataset (Section 1.2) and an additional editing example (Figure 20).

1.1 Additional Discussion

Naive Diffusion. In Figure 10, we demonstrate visually that naively diffusing disparity labels can be problematic because edge localization is ambiguous.

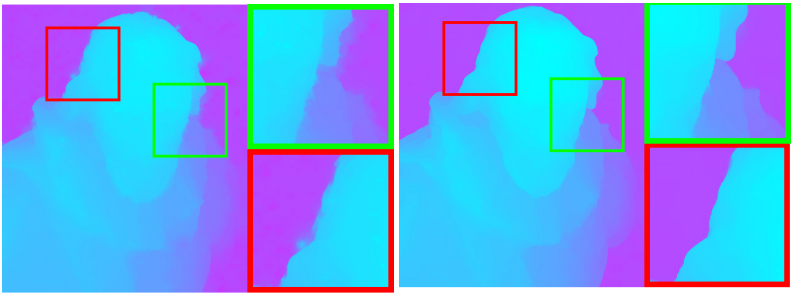

Figure 10: *Left*: Naïvely diffusing disparity labels causes artifacts around edges due to ambiguity in the localization of labels around edges. *Right*: Estimating the diffusion gradient removes this ambiguity and yields sharp depth edges.

Multi-view Depth and Error. As ground truth disparity is only provided for the central view of the HCI data set, and as the Stanford data set has no ground truth depth, we did not include quantitative error evaluation across '4D' views. Qualitatively, our method tends to produce results that are consistent across views (Fig. 11).

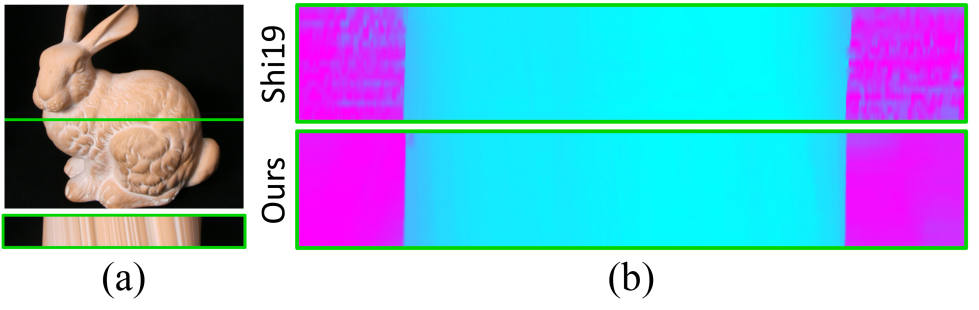

Figure 11: **(a)** We visualize depth consistency for the highlighted epipolar line. **(b)** Our results are more consistent than Shi et al. [4] across views (EPIs are scaled vertically for clarity).

Disparity Noise and Blur Tolerance. To show our robustness, we evaluate our method on noisy disparity labels (Fig. 12) and low-gradient edges (Fig. 13). Our method provides greater robustness to disparity errors than naive diffusion, and provides greater robustness via MSE to low-gradient (or blurry) edges than two learning-based baselines.

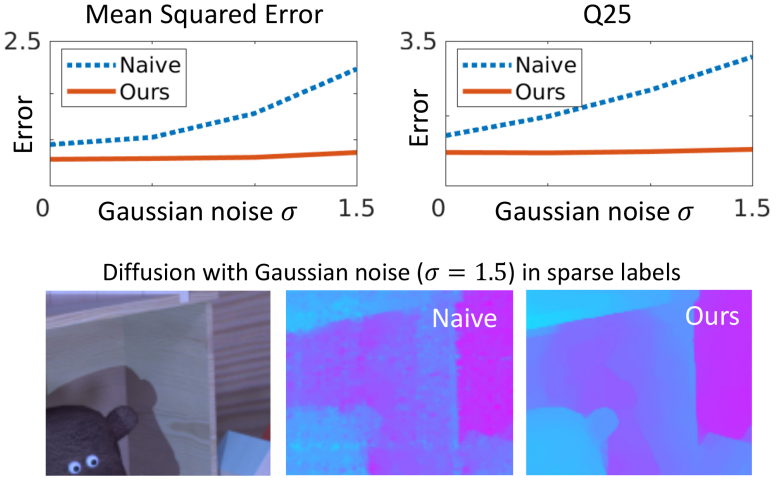

Figure 12: Robustness of our method to noise in disparity labels (*Dino* light field; we compare with naive diffusion.).

Hyperparameter variation. Figure 14 demonstrates the variation in error as hyperparameter values change. Across all parameters, our approach is stable around our declared values.

Lenslet Distortion and EPFL Lytro Dataset. The Lytro light fields in the EPFL dataset are provided decoded as MATLAB files. In general, while our method can handle small amounts of

Figure 13: Robustness of our method to low-gradient edges (*Dino* light field; we compare to the methods of Zhang et al. [5] and Jiang et al. [2] which have the best MSE and Q25 performance on this light field, respectively).

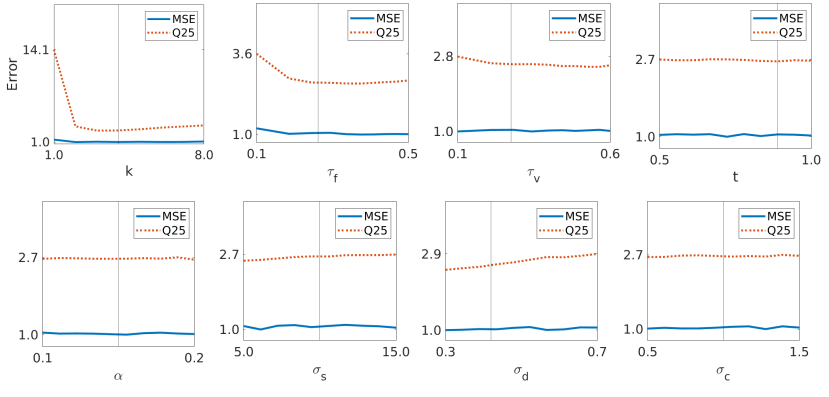

Figure 14: Effect of hyperparameter values on the MSE and Q25, averaged across the HCI dataset: k and τ_f (Eqn. 1), τ_v (Eqn. 2), t and α (entropy-based refinement), and σ_s, σ_d and σ_c (Eqn. 4). The vertical lines indicate our chosen values. The stochasticity of our algorithm means the chosen values may not be optimal in all cases. However, the method is stable to variation around these values.

distortion, the EPI-based edge detection stage expectedly fails when EPI features are no longer linear. This is true for the edge views of Lytro light fields. As such, we only use the central 7×7 views of the EPFL scenes for all experiments.

Black Backgrounds and Stanford Dataset. Our EPI edge detector aggregates information from all three channels in CIE LAB color space, which allows it to detect even faint edges. Thus, it captures the subtle background texture on the black cloth in the Stanford dataset examples of

Figure 15: *Top:* On the Stanford Bunny scene, enhanced image contrast shows the texture of the cloth in the seemingly black background. *Bottom:* In EPI space (scaled vertically for clarity) the texture appears as sloped lines, providing background disparity to methods that can exploit this subtle information.

single objects; typically, this detail is not visible to the naked eye. This feature of our work also explains why we do not incorrectly detect false edges in the Lego Technic Plow scene, as shown in Figure 7 of the main paper.

1.2 Expanded Results

We present qualitative results on the HCI dataset in Figure 16, and expanded results on the real-world light fields of the Stanford dataset in Figure 19. Our method produce stronger depth edges compared to the baselines, and our smoothness regularization (Equation 10, main paper) leads to fewer artifacts in textureless regions.

1.3 Diffusion Gradients as Self-supervised Loss

As in main paper Section 4, we compare our method to a reprojection error loss. In Figure 17, to complement the quantitative MSE numbers in the main paper, we demonstrate the qualitative improvement from our bidirectional diffusion gradient approach in comparison.

1.4 Error Maps

We visualize the absolute disparity error of all baselines and our method in Figure 18. The baseline methods produce larger errors around depth edges compared to our approach. This can be seen in the fewer regions of red for our method compared to the baselines. The corresponding dense disparity maps are shown in Figure 16. Qualitatively, our results are comparable to the learning-based baselines [2, 3, 4] with fewer extreme errors around edges.

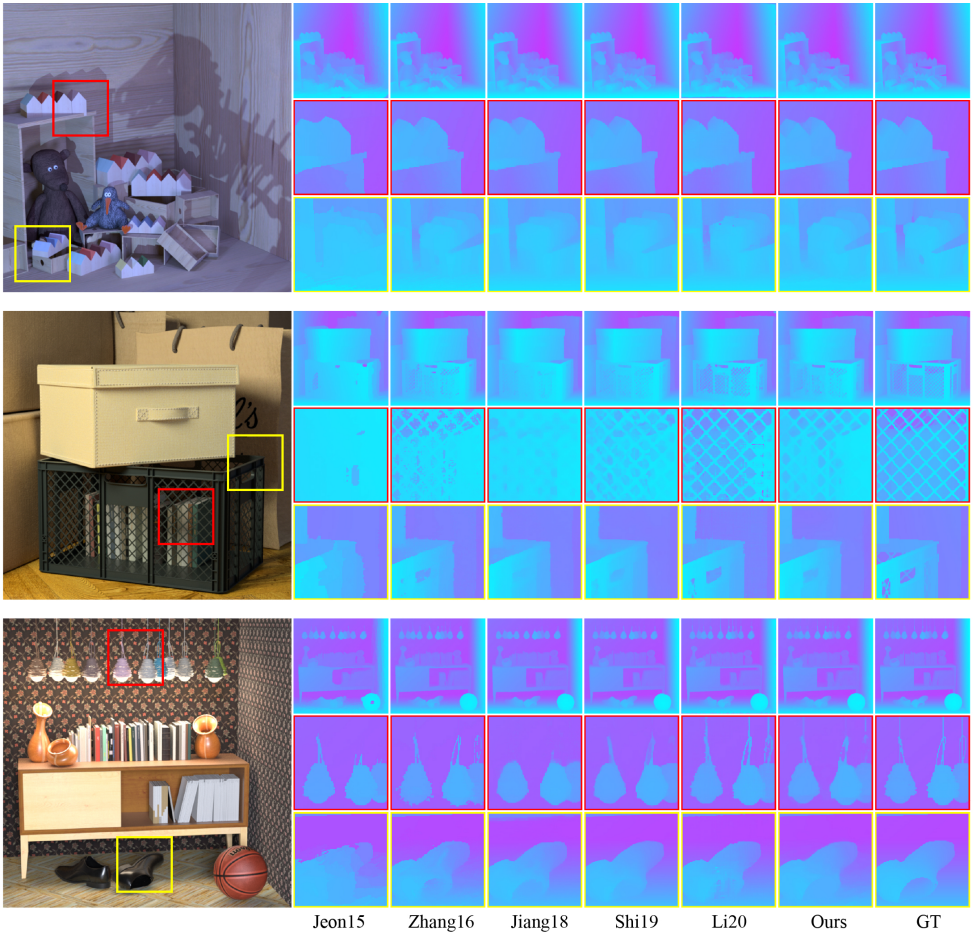

Figure 16: Results on the synthetic light fields of the HCI dataset. *Left to right:* Jeon et al. [1], Zhang et al [5], Jiang et al. [2], Shi et al. [4], Li et al. [3], our method, and finally, the ground truth. Qualitatively, our results are comparable to the learning-based baselines [2, 3, 4] with fewer extreme errors around edges.

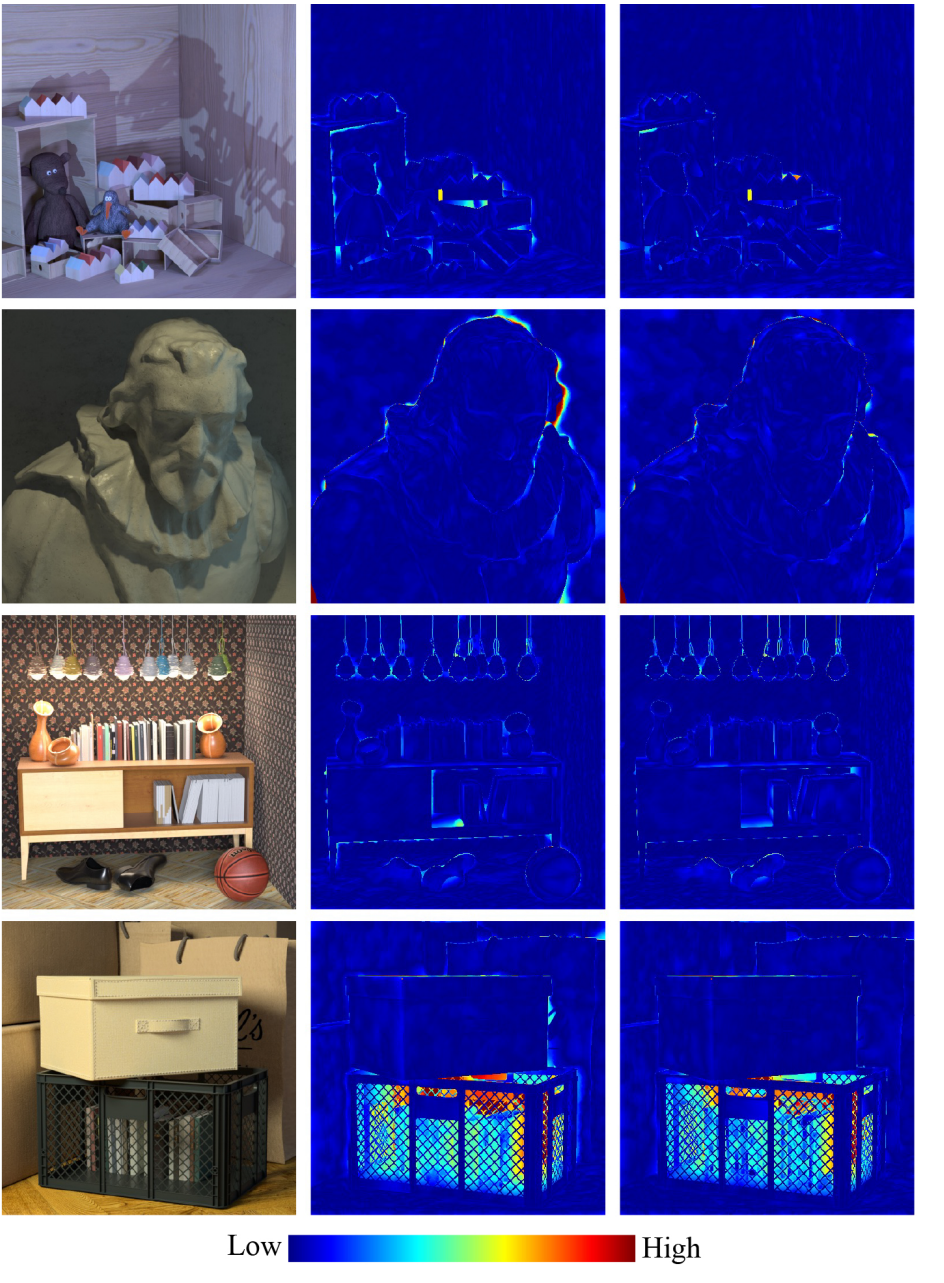

Figure 17: Multiview reprojection error (*center*) as self-supervised loss for depth edge localization, compared to our bidirectional diffusion gradients (*right*). We show absolute disparity error. Our method has lower error around edges.

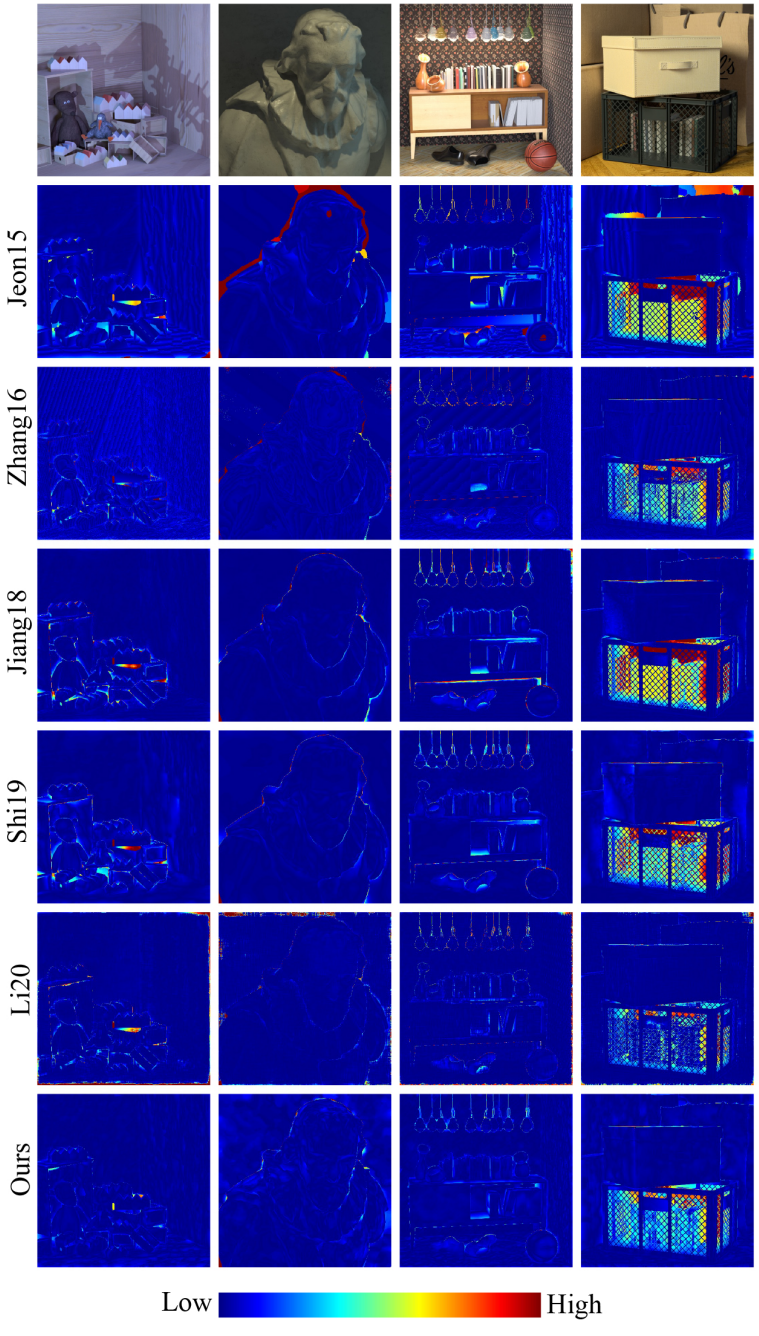

Figure 18: A visualization of the absolute disparity error for all baselines. *Top to bottom:* Jeon et al. [11], Zhang et al [56], Jiang et al. [22], Shi et al. [49], Li et al. [33], and our method.

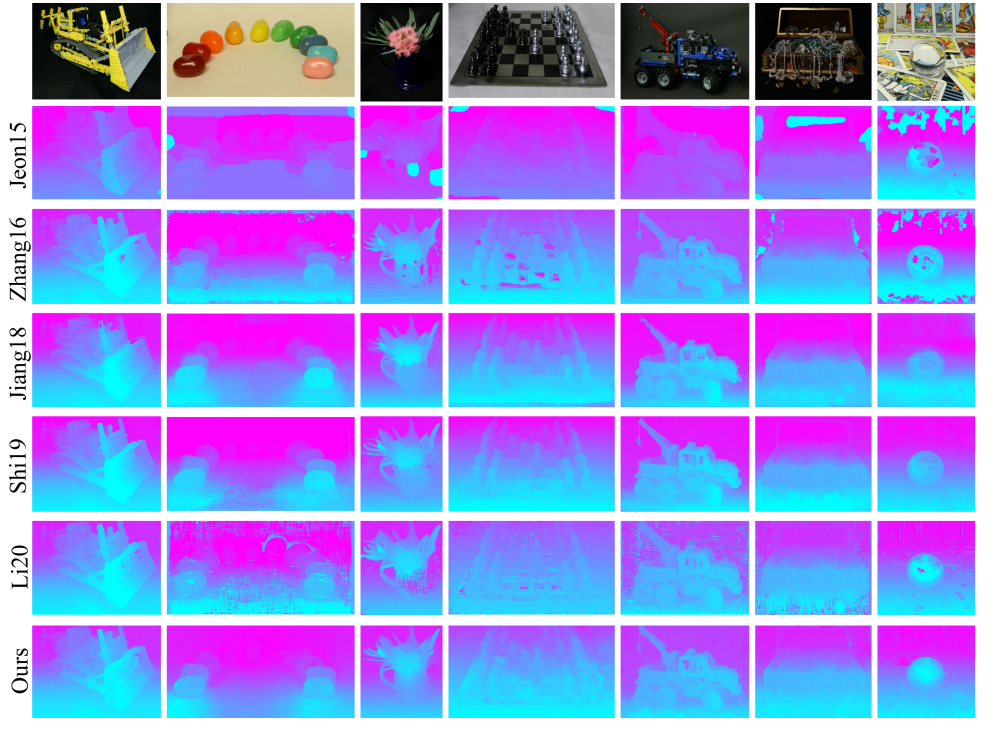

Figure 19: Results on light fields from the Stanford dataset. *Top to bottom:* Jeon et al. [1], Zhang et al [5], Jiang et al. [2], Shi et al. [4], Li et al. [3] and our method.

Figure 20: Additional light field editing result. *Left:* input scene. *Center:* Our editing results. *Right, clockwise from top-left*: Detail of the unmodified light field image, Zhang et al. [5]'s editing result, Shi et al. [4]'s editing result, and our result with fewer artifacts.

References

[1] Hae-Gon Jeon, Jaesik Park, Gyeongmin Choe, Jinsun Park, Yunsu Bok, Yu-Wing Tai, and In So Kweon. Accurate depth map estimation from a lenslet light field camera. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1547–1555, 2015.

[2] Xiaoran Jiang, Mikaël Le Pendu, and Christine Guillemot. Depth estimation with occlusion handling from a sparse set of light field views. In *2018 25th IEEE International Conference on Image Processing (ICIP)*, pages 634–638. IEEE, 2018.

[3] Kunyuan Li, Jun Zhang, Rui Sun, Xudong Zhang, and Jun Gao. Epi-based oriented relation networks for light field depth estimation. *British Machine Vision Conference*, 2020.

[4] Jinglei Shi, Xiaoran Jiang, and Christine Guillemot. A framework for learning depth from a flexible subset of dense and sparse light field views. *IEEE Transactions on Image Processing*, 28(12):5867–5880, 2019.

[5] Shuo Zhang, Hao Sheng, Chao Li, Jun Zhang, and Zhang Xiong. Robust depth estimation for light field via spinning parallelogram operator. *Computer Vision and Image Understanding*, 145:148–159, 2016.