Supporting Information

GEOMScope: Large Field-of-view 3D Lensless Microscopy with Low Computational Complexity

Feng Tian[1], Junjie Hu[1], and Weijian Yang[1]*

[1]*Department of Electrical and Computer Engineering, University of California, Davis, CA 95616, USA*
**wejyang@ucdavis.edu*

Supporting Section S1-S5.

Supporting Table S1.

Supporting Figure S1-S12.

Section S1. Particle clustering algorithm

In this section, we discuss the detailed procedure of background suppression through a particle clustering algorithm (Figure S2b). This algorithm is the second step in the hybrid object reconstruction algorithm. After the first step of object reconstruction through the pixel back projection, the reconstructed object could contain light from the focused object, and background from the defocused object, ghost object and system noise (e.g. stray light, camera readout noise). The particle clustering algorithm aims to pick out the focused object. This algorithm is suitable for sparse, discrete point objects distributed in a 3D space. The clustering algorithm contains two steps, and we illustrate it with an example in Figure S3.

Step 1: Separate and cluster the reconstructed objects into isolated particle groups (Figure S3b).
In this step, we aim to distinguish different particle groups that do not connect with each other, by a combined operation of thresholding and clustering.

We first perform an intensity thresholding to remove the ghost object and system noise. The ghost object originates from the "fake" projections in the pixel back projection, due to the "one-to-many mapping" nature of microlens array. When the object is sparse, the ghost objects have a much smaller intensity than the correctly reconstructed object (including both focused and defocused). The same applies to the system noise. We can thus use a threshold to remove these two background components. The threshold is chosen based on the mean and standard deviation of all pixels' intensity. This step essentially cleans up the background between the particle objects far from (or isolated from) each other.

To distinguish the different particles groups, we perform clustering on the thresholded 3D volumetric data based on the pixel connectivity. From the graph perspective, each group has internal connections through overlapped focused and defocused light from particles that are close to each other within the group. No connection occurs between different groups.

Step 2: Subdivide each clustered particle group into smaller particle clusters (Figure S3c-h).
As each group could contain multiple particles, we aim to further divide them into individual smaller particle clusters. We note that the different particles within each group could have different intensity. The defocused light of the brighter particles may overlap with particles with weaker intensity. Thus, the clustered particle group could contain intensity peaks for the brighter particles, and intensity sub peaks from smaller particles. For each group, we can use thresholding to remove the low intensity light between the sub peaks and perform clustering again to separate them into smaller particle groups. This can be considered as a reiteration of step 1, but it is now performed for each group. We find a threshold setting as 5-20% of the peak intensity of each group (i.e. only keeping those voxels with intensity beyond this threshold) can typically yield desirable results. For each small cluster, we find the voxel that has peak intensity to represent the position of the particle in the 3D volume.

As a small threshold (e.g. 5%) may lead to clustering two close particles with high intensity into one and a large threshold (e.g. 20%) may remove particles with small intensity (Figure S3g), to maximize the number of clusters, we select multiple threshold (e.g. 5%, 10%, 15%, 20%). For each threshold, we perform the clustering (Figure S3c-g). We then unionize/superimpose the results together to achieve a final clustering result (Figure S3h).

Section S2. Generation of training data for convolutional neural network

The second approach of background suppression is through a U-net based convolutional neural network (Figure S2c). This method is suitable for 3D objects consisting of continuously connected features. To train the neural network, we generated artificial datasets using polygons. In each dataset, we generated polygons with random intensity and at random positions across 10 slices in an imaging stack. Each slice contains in-focus polygons and out-of-focus light from polygons at other slices, with a 5% additive noise. The out-of-focus light is modeled by a Gaussian blur process of polygons at other slices. The Gaussian blur kernel expands spatially with reduced weight when the slices are further away. The advantage of using Gaussian blur process is its low computational cost, so the training dataset can be generated fast.

We validated the Gaussian blur process can indeed model the defocused objects. As the pixel back projection correctly models the physical process, we compared the intensity profile of the defocused light artificially generated by Gaussian blur process with that from pixel back projection of a snowflake object (Figure S4). The two sets of intensity profiles, and their depth dependent mean intensity, agree well with each other across all defocused distances (Figure S4b-c).

Section S3. Derivation of the depth of field

To calculate the depth of field of a single lens unit, we define the circle of confusion at the image space, and then project it into the object space with thin lens equation [1].

Given the diameter of the circle of confusion c, lens unit diameter D, lens focal length f, object distance z and image distance v, and following the properties of similar triangles in Figure S6, we have

$$\frac{v_N - v}{v_N} = \frac{c}{D} \tag{S1}$$

The thin lens equation expresses as,

$$\frac{1}{z} + \frac{1}{v} = \frac{1}{f} \tag{S2}$$

$$\frac{1}{z_N} + \frac{1}{v_N} = \frac{1}{f} \tag{S3}$$

Expressing v_N using v in Equation S1, we have

$$v_N = \frac{Dv}{D - c} \tag{S4}$$

Substituting v_N into Equation S3, we get expression of z_N as

$$z_N = \frac{fDv}{Dv - fD + fc} \tag{S5}$$

Expressing v by Equation S2, we have

$$v = \frac{fz}{z - f} \tag{S6}$$

Thus, the distance of the nearest scene in good focus is

$$z_N = \frac{fD\frac{fz}{z-f}}{D\frac{fz}{z-f} - fD + fc} = \frac{f^2Dz}{f^2D + zfc - f^2c} = \frac{f^2z}{f^2 + z\,F_{\#}c - f\,F_{\#}c} \tag{S7}$$

where we replace f/D with the lens F-number $F_{\#}$.

Similarly, for distance of the furthest scene in good focus, we have the following from similar triangles,

$$\frac{v - v_F}{v_F} = \frac{c}{D}, v_F = \frac{Dv}{D + c} \tag{S8}$$

Applying the thin lens equation, we get

$$\frac{1}{z_F} + \frac{1}{v_F} = \frac{1}{f} \tag{S9}$$

$$z_F = \frac{fDv}{Dv + fD - fc} \tag{S10}$$

Substituting Equation S8 leads to,

$$z_F = \frac{fD\frac{fz}{z-f}}{D\frac{fz}{z-f} + fD - fc} = \frac{f^2Dz}{f^2D - zfc + f^2c} = \frac{f^2z}{f^2 - z\,F_{\#}c + f\,F_{\#}c} \tag{S11}$$

The depth of field can thus be expressed as

$$DOF = z_F - z_N \cong \frac{2F_{\#}cz^2f^2}{\left(f^4 - {F_{\#}}^2c^2z^2\right)} \tag{S11}$$

Section S4. Performance of the individual lens units

The nominal diameter D and the focal length f of an individual lens unit is 2 mm and 4.65 mm respectively. We simulate the 3D PSF in OpticStudio (Figure S7a). The lateral span of the focusing spot increases from the post focus position to pre focus position, due to spherical aberration of the lens. This translates to an enlarged reconstruction profile, and thus a degraded image resolution when an object point moves closer to the microlens array (Figure S9).

To evaluate the off-axis performance of the lens unit, we simulate the dependence of Strehl ratio (Figure S7b) and FWHM of the lateral PSF (Figure S7c) on the lateral displacement of a point object from the optical axis. Setting a Strehl ratio above half of its peak value, we determine that the diameter of the effective field of view of each lens unit is ~6 mm.

Section S5. Comparison between hybrid reconstruction method (pixel back projection followed by background suppression) and iterative optimization algorithms at different data scale

We compared the reconstruction quality and the required computation time of the hybrid reconstruction method (pixel back projection followed by background suppression) and two representative global optimization solvers: ADMM and Richard-Lucy (R-L) deconvolution. Here, we used a convolutional forward model to project objects (either featureless objects containing discrete, isolated point sources or objects with continuous features) onto an image. The PSF was set to be sparse random points distributed across the image plane to mimic the random microlens array imaging scenario. While we chose a spatial invariant PSF here for simplicity, we note that similar conclusion can be reached in the spatial variant PSF case as the reconstruction algorithms follow the same principles.

We performed the comparison for two classes of objects: featureless objects containing discrete, isolated point sources on a 2D plane (using particle clustering as background suppression algorithm), and 3D objects with continuous features (using convolutional neural network as background suppression algorithm). For each class of objects, we varied the data scale by controlling the pixel-voxel pairs, i.e. the pixel numbers in the measured image and the voxel number to be reconstructed in the object space. Across different data scale of pixel-voxel pair numbers, we kept the same pixel-voxel ratio. For the featureless objects containing discrete, isolated point sources on a 2D plane, the reconstructed objects are shown in Figure S11 with the ground truth for reference. For the 3D objects with continuous features, the reconstructed objects are shown in Figure S12 with the ground truth for reference. The results on the computation time and metrics of reconstruction quality are shown in Figure 7 in the main text.

Table S1

Device	Effective sensor pixels	Lateral res.	Axial res.	Demonstrated field of view	Depth of field	Working distance	Reported /Derived maximum number of resolvable points	Object reconstruction time for max. resolvable points	Max. pixel-voxel pair counts
FlatScope [2]	1300 × 1000	$\sim 2\mu m$	$\sim 15\mu m$	2.25 × 2.8 mm^2	$210\mu m$	$< 1\ mm$	2.2×10^7	15 min. NVIDIA Tesla GK210 GPU, MATLAB	2.87×10^{13}
Diffuser-Cam [3]	1280 × 1080	$\sim 60\ \mu m$	$\sim 448\ \mu m$	> 10 × 10 mm^2	$\sim 8\ mm$	$\sim 20\ mm$	5.3×10^8	26 min. Workstation/ 144-core CPU, 85 GB RAM MATLAB	7.4×10^{14}
Microlens diffuser [4]	968 × 548	$\sim 8\ \mu m$	$\sim 50\ \mu m$	~5.6 × 3.2 mm^2	1.5 mm	$\sim 2\ mm$	8.4×10^6	-	4.46×10^{12}
MLA Mesoscope [5]	2592 × 1944	$7\ \mu m$	$200\ \mu m$	$8 \times 7\ mm^2$	2.5 mm	$\sim 12\ mm$	2.16×10^7	2.5 hr. Computing cluster/ 8-core CPU, 256 GB RAM MATLAB	1.09×10^{14}
3D Miniscope [6]	648 × 486	2.76 $\sim 3.9\ \mu m$	$15\ \mu m$	0.9 × 0.7 mm^2	0.39 mm	~1.67 mm (through GRIN objective lens)	2.15×10^6	8-24 min. NVIDIA RTX 2080-Ti GPU, MATLAB	7.72×10^{12}
GEOM-Scope (this work)	5120 × 5120	$40\ \mu m$	$300\ \mu m$	~23 × 23 mm^2	~5 mm	18~30 mm	5.5×10^6	30 sec. Desktop/CPU, <~50 MB RAM; 18 sec. Workstation/ 16-core CPU, <~400 MB RAM MATLAB	1.44×10^{14}

Table S1. Comparison of performance metrics between different lensless imagers.

Note:

1. The depth of field is specified/estimated as the range where the specified resolutions can be maintained. The actual depth of field can be larger at a tradeoff of degraded resolution.
2. The maximum resolvable points are either reported value or derived value as the ratio of imaged volume and 3D resolution.
3. The maximum pixel-voxel pair counts are calculated value from the product of effective sensor pixel number and reconstruction voxel number (either reported value or derived value as the ratio of imaged volume and 3D resolution).

Figure S1

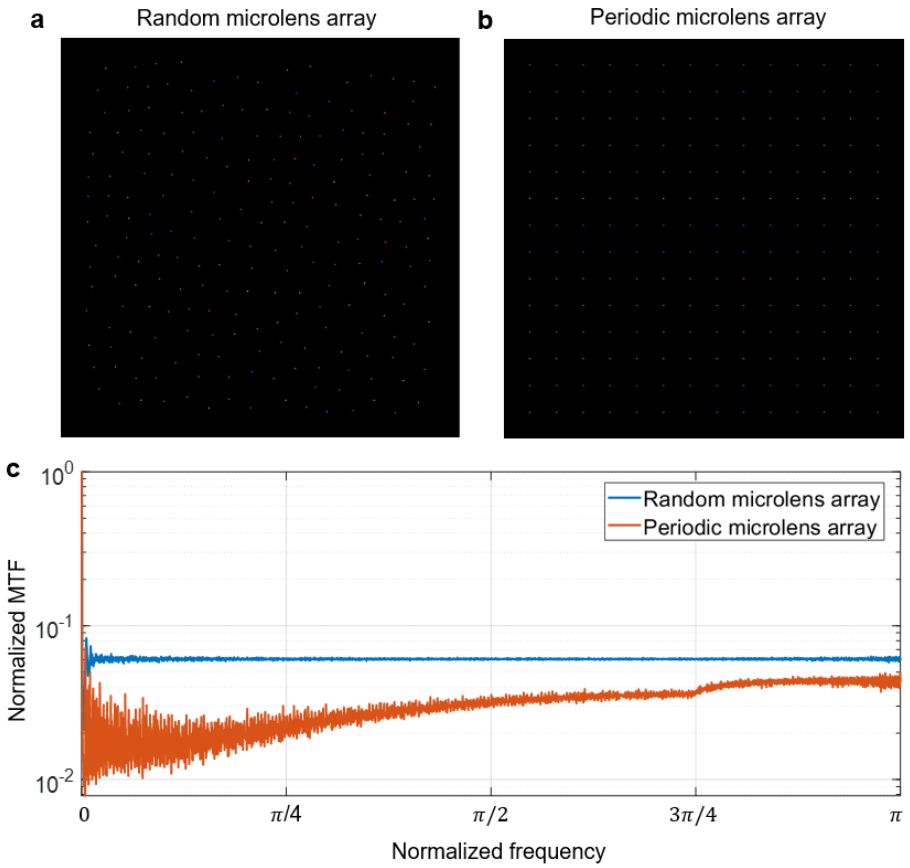

Figure S1. Modulation transfer function (MTF) of random microlens array and periodic microlens array. (a) Centroid positions of individual lens units of the random microlens array with a total lens unit of 213. (b) Centroid positions of individual lens units of the periodic microlens array with a total lens unit of 196. (c) Normalized MTF of the random microlens array and periodic microlens array.

Figure S2

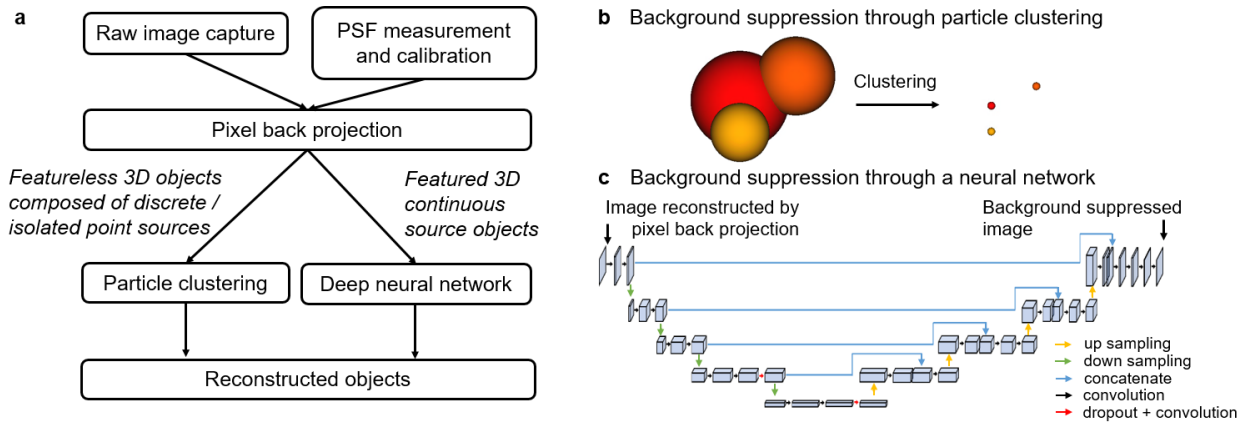

Figure S2. Hybrid reconstruction algorithm in GEOMScope. (a) Flow chart of hybrid reconstruction algorithm in GEOMScope. (b) Particle clustering algorithm, suitable for sparse featureless point objects. (c) Convolutional neural network, suitable for less sparse 3D objects with continuous features. Both the input and output of the neural network are 2D images at single depth.

Figure S3

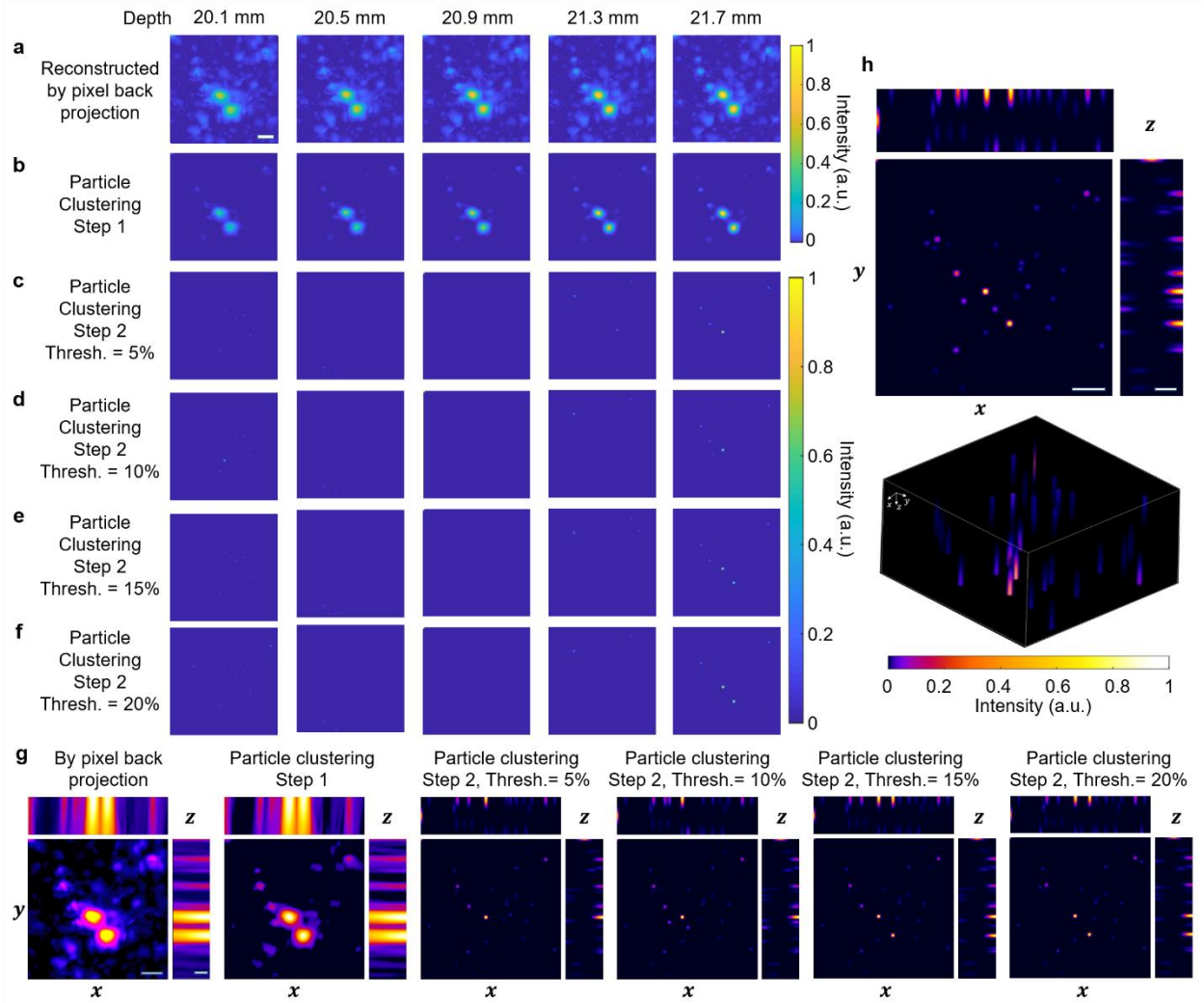

Figure S3. Demonstration of particle clustering algorithm. The data is part of the experimental data from Figure 5 in the main text. (a) Fluorescent particles reconstruction through pixel back projection algorithm. (b) Results after step 1 of particle clustering: the reconstructed objects are separated and clustered into isolated particle groups. (c-f) Results after step 2 of particle clustering: each clustered particle group is subdivided into smaller particle clusters. The particles are blurred intentionally for clearer visualization. Different thresholds are used in (c)-(f). (g) Maximum intensity projection view of the 3D volume in (a-f) at xy, xz and yz plane. (h) Maximum intensity projection view at xy, xz and yz plane and 3D view of the final reconstruction results. Scale bar: 1 mm in (a-h). We use different colorbars in (a-b), (c-f) and (g-h) for better visualization effect.

Figure S4

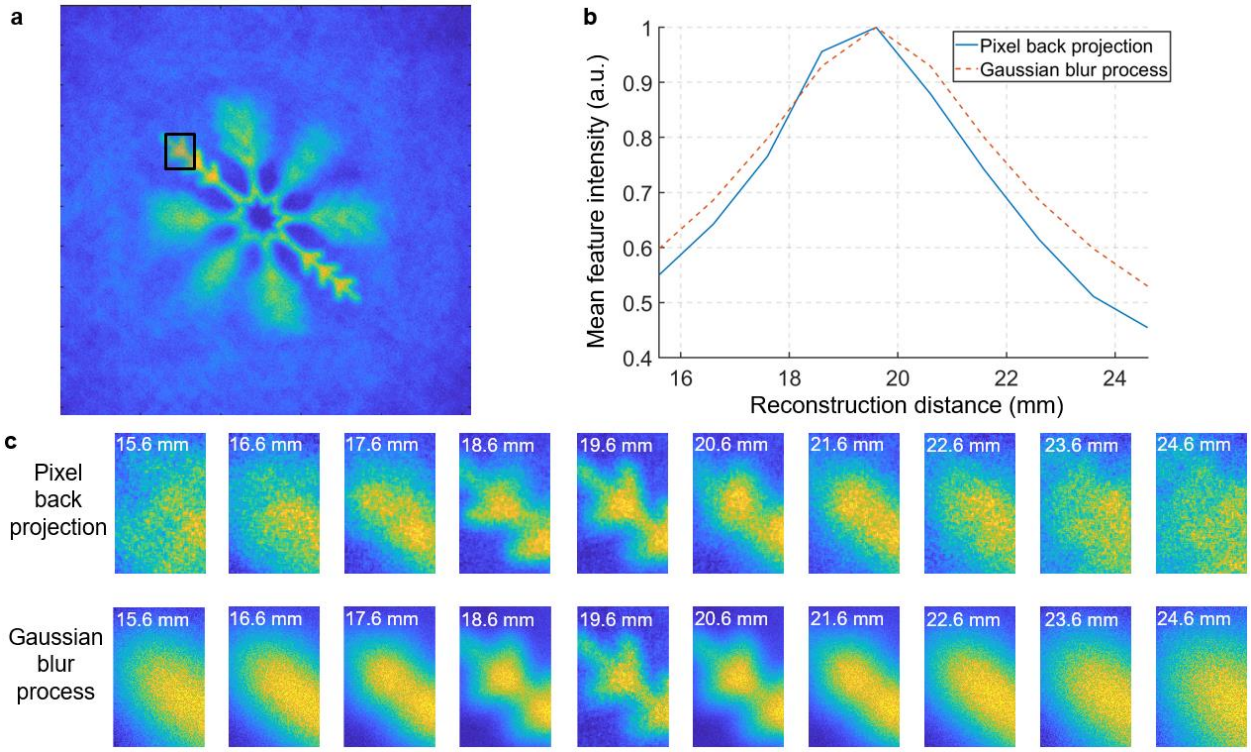

Figure S4. Validation of Gaussian blur process to model the defocus light in the artificially generated training data. (a) Reconstructed 3D snowflake object at a certain depth, with feature of interest being selected in the box. (b) Mean intensity of the feature of interest reconstructed at 10 different depths by pixel back projection, and that from those generated by the Gaussian blur process of the in-focus reconstruction. (c) Intensity profile of the defocused light reconstructed by pixel back projection (top row) and that generated by the Gaussian blur process (bottom row).

Figure S5

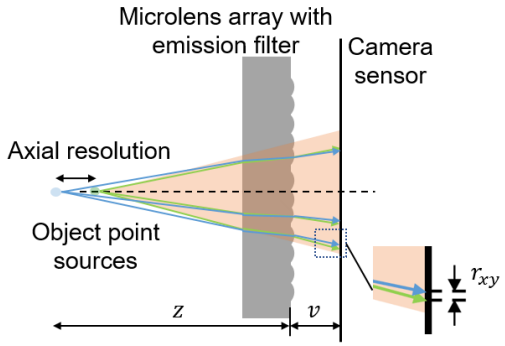

Figure S5. Evaluation of axial resolution. The axial resolution is evaluated as the minimum axial distance of two object point sources when their imaged spots are separated by a resolvable distance (i.e. lateral resolution r_{xy}) on the camera sensor.

Figure S6

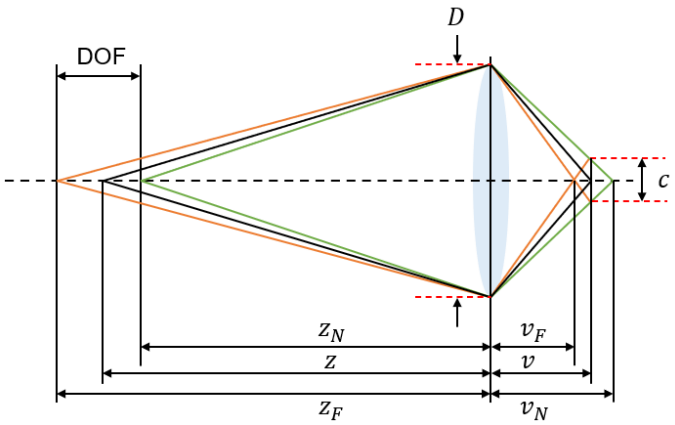

Figure S6. Evaluation of depth of field (DOF). The depth of field is evaluated as the axial range in the object space when the object points on the two edges have an image size matching the defined confusion circle.

Figure S7

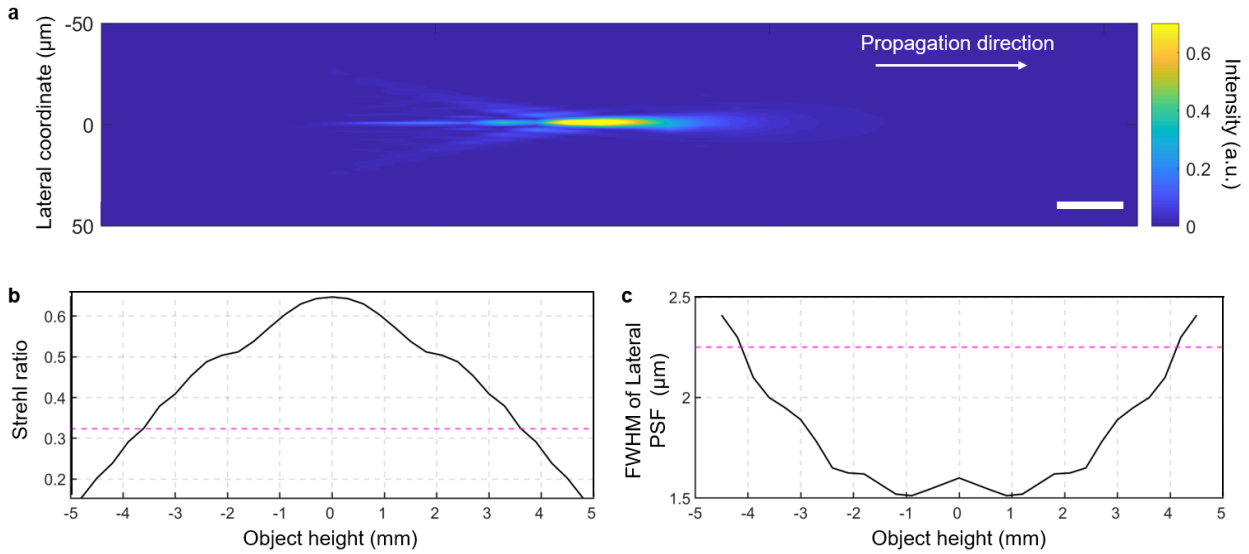

Figure S7. Performance of the single lens unit in our microlens array design. (a) Cross section profile of the focal spot of a single lens unit, simulated by physical optics propagation. Scale bar: 100 μm. (b) Strehl ratio from ray tracing simulation. The dashed line indicates half of the peak Strehl ratio. (c) FWHM of the lateral PSF from ray tracing simulation. The dashed line indicates half of the sensor pixel size.

Figure S8

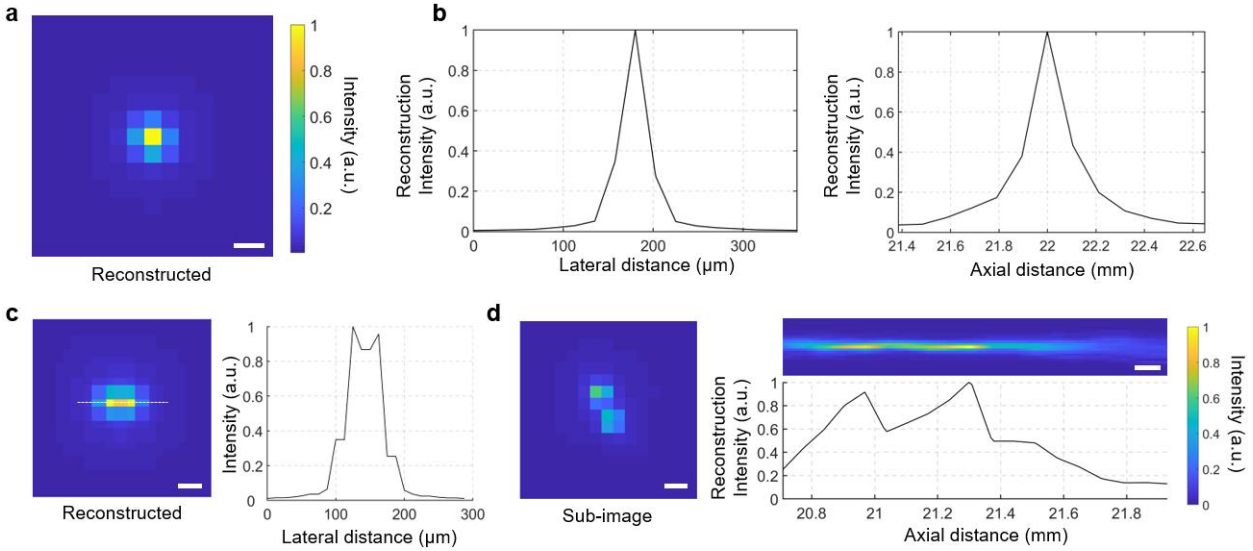

Figure S8. Simulation on the 3D resolving power of GEOMScope. (a) The lateral spatial profile of the reconstruction of a single point object. Scale bar: 40 μm. (b) Line profile of the reconstructed single point object along the lateral (left) and axial (right) direction. (c) Simulated two-point lateral resolution: two object point sources laterally separated by 40 μm at z=20 mm can be resolved. Left, the spatial profile of the reconstructed object; right, line profile along the lateral direction of the reconstructed object. Scale bar: 40 μm. (d) Simulated two-point axial resolution: two object point sources axially separated by 300 μm at z=20 mm can be resolved. Left, the sub image from one lens unit, showing the two axially separated point sources are separated laterally in the sub image. Right, the reconstructed objects with its line profile along the axial direction. Scale bar: left, 10 μm; right, 100 μm.

Figure S9

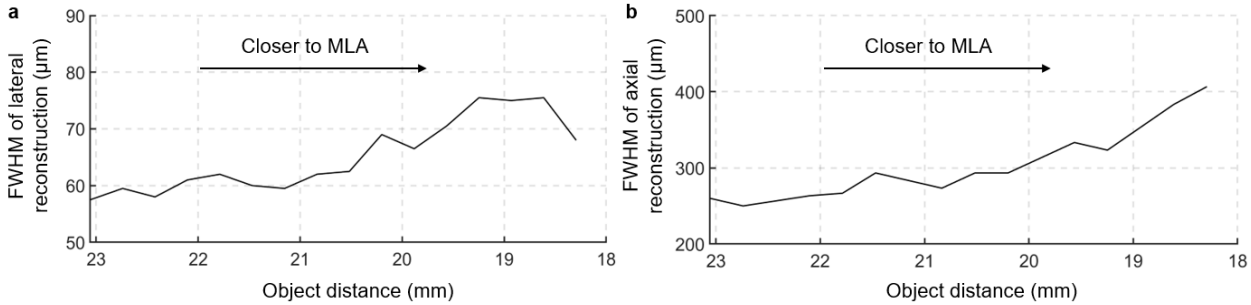

Figure S9. Experimental measurement of the reconstruction profile of a single object point versus object depth. (A) FWHM of the point reconstruction profile along lateral direction versus object depth. (B) FWHM of the point reconstruction profile along axial direction versus object depth. MLA, microlens array.

Figure S10

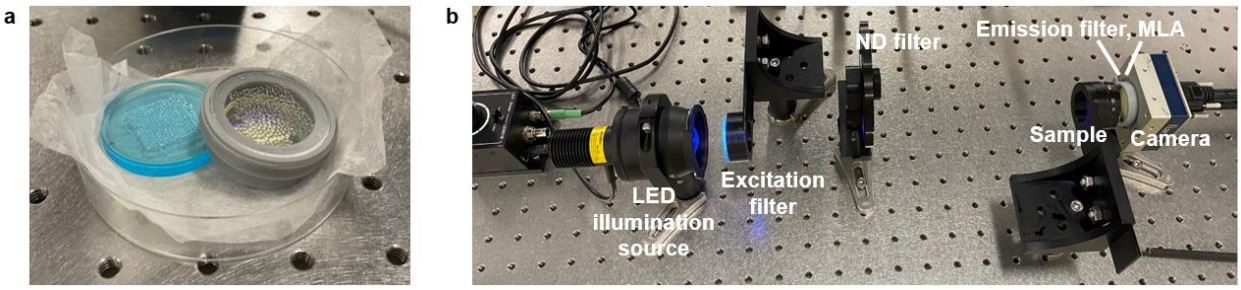

Figure S10. (a) Microlens array mold (left) and the fabricated microlens array in a 3D printed mount to be attached to a camera (right). (b) Experimental setup. ND filter, neutral density filter; MLA, microlens array.

Figure S11

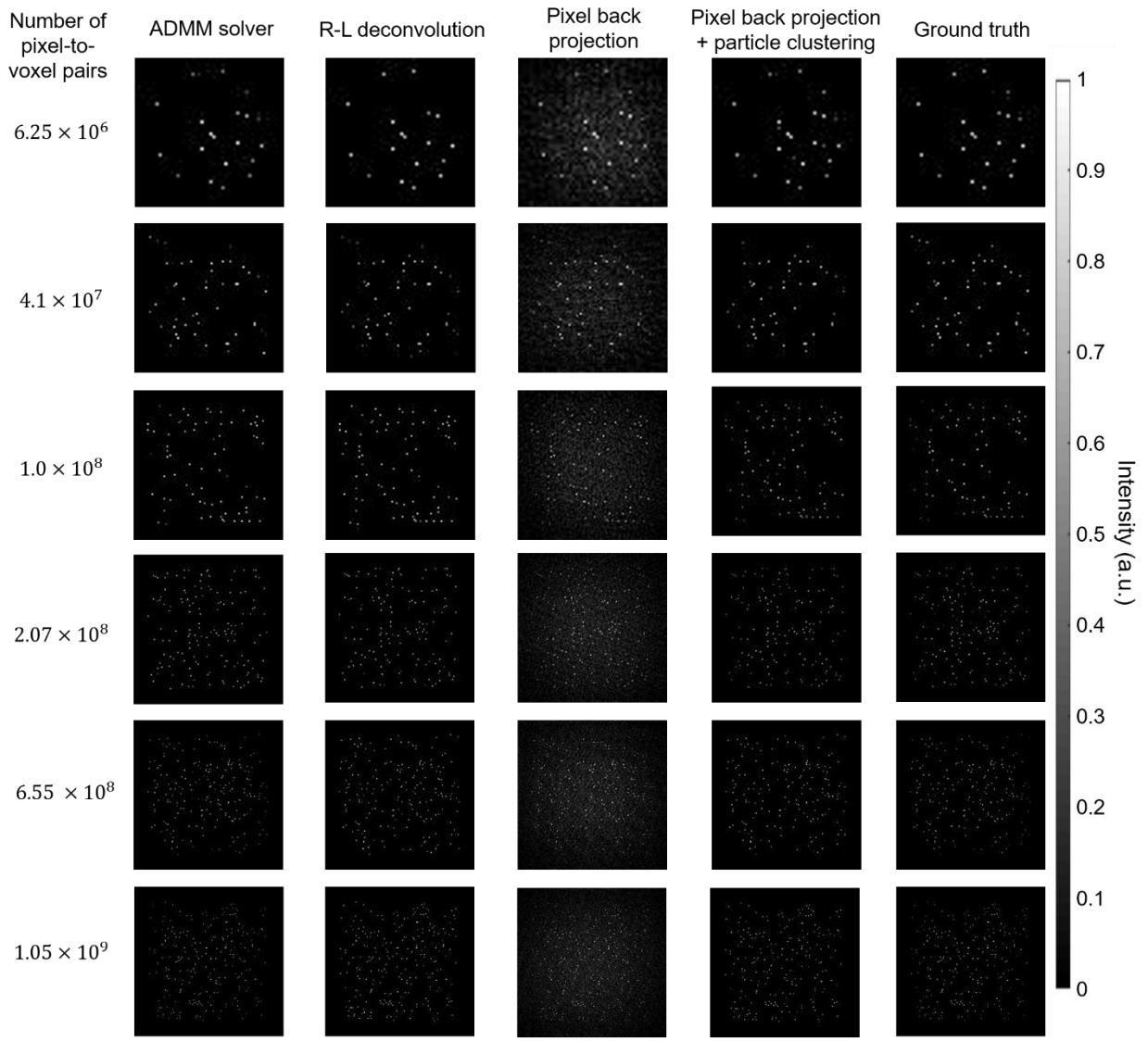

Figure S11. Comparison of the reconstruction results (for featureless discrete / isolated point objects) using ADMM solver, Richard-Lucy (R-L) deconvolution, and pixel back projection with and without particle clustering, for different number of pixel-voxel pairs.

Figure S12

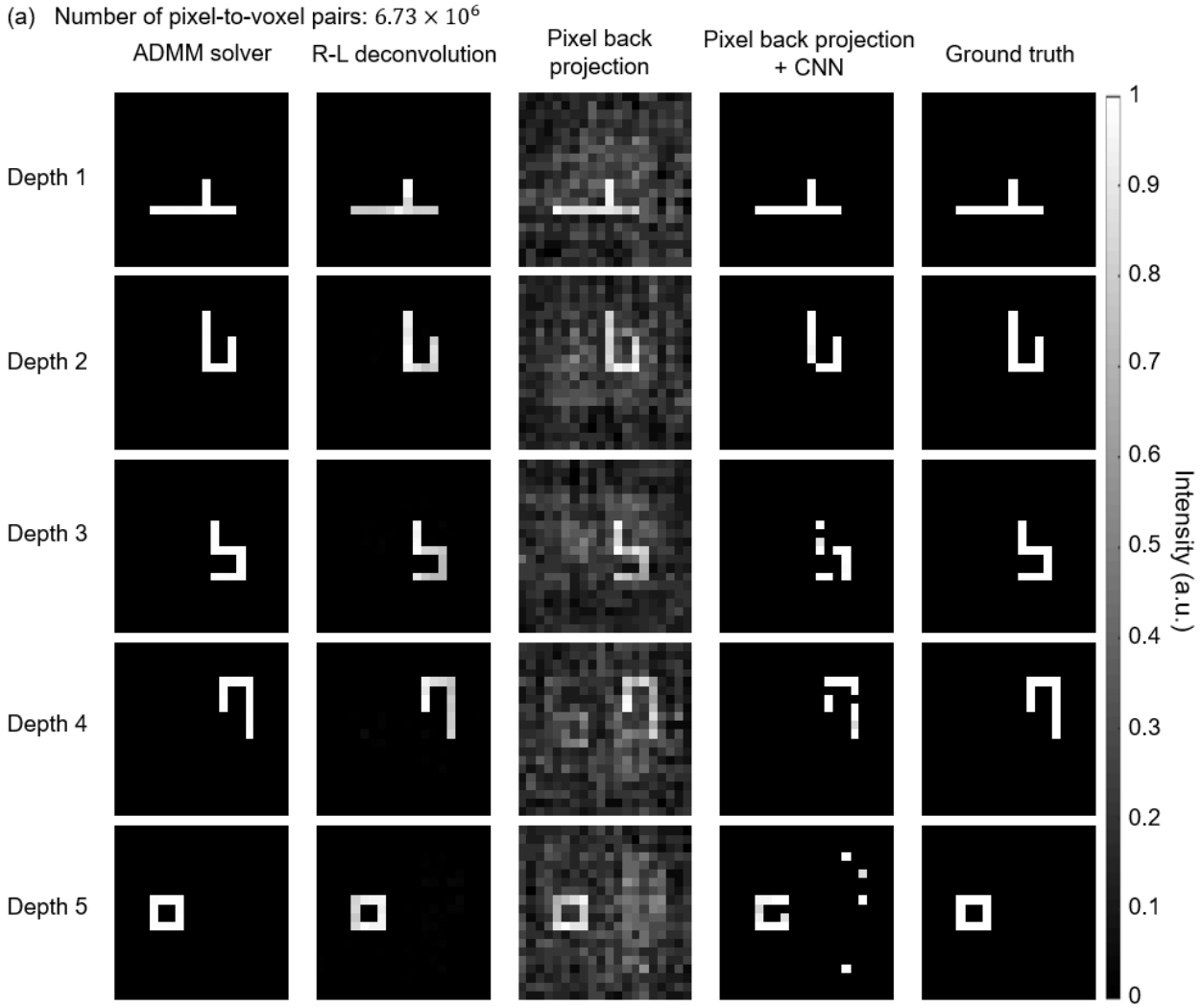

Figure S12. (a) Comparison of the reconstruction results (for 3D objects with continuous features) using ADMM solver, Richard-Lucy (R-L) deconvolution, and pixel back projection with and without convolutional neural network, for pixel-voxel pairs of 6.73×10^6.

Figure S12

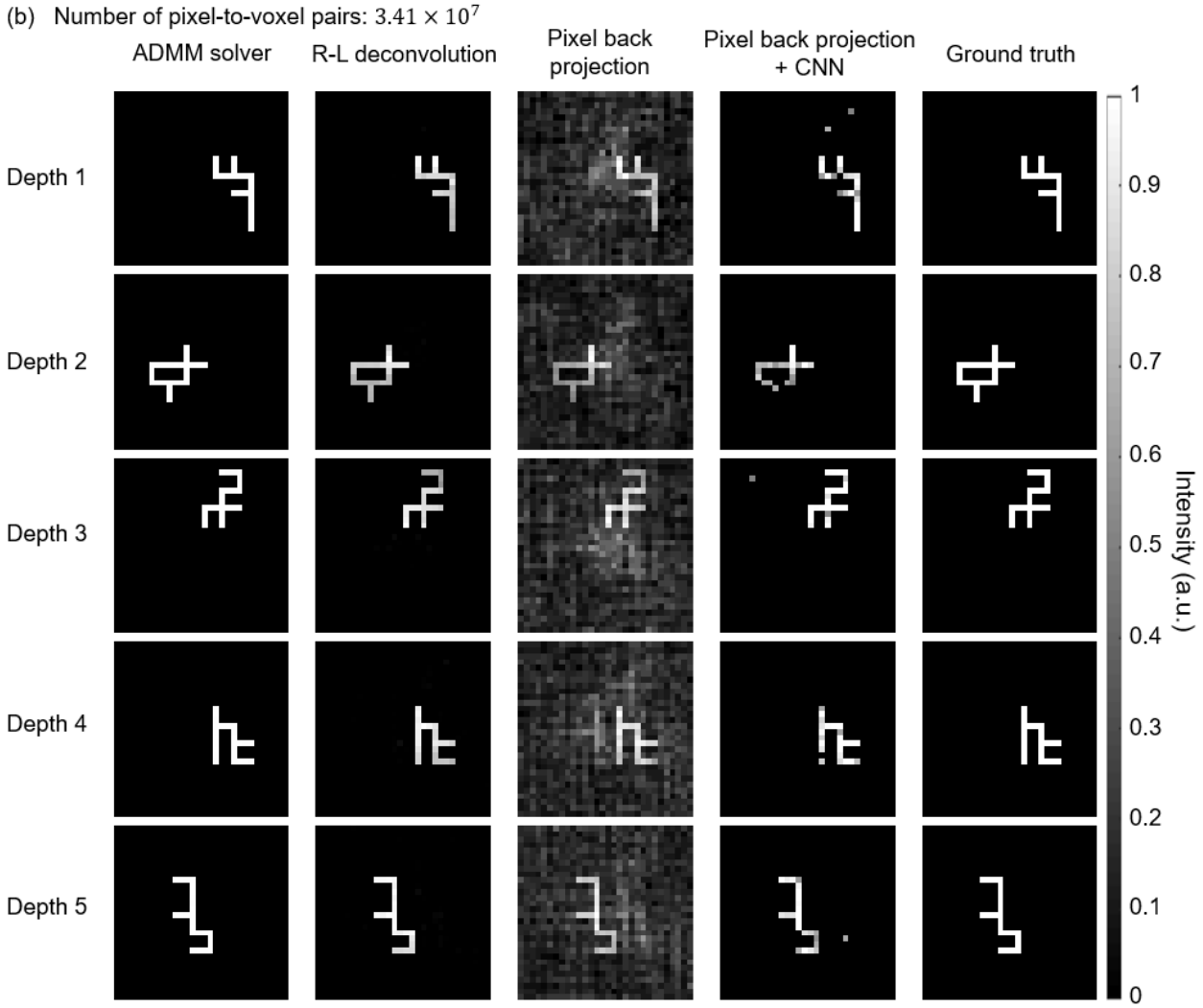

Figure S12. (b) Comparison of the reconstruction results (for 3D objects with continuous features) using ADMM solver, Richard-Lucy (R-L) deconvolution, and pixel back projection with and without convolutional neural network, for pixel-voxel pairs of 3.41×10^7.

Figure S12

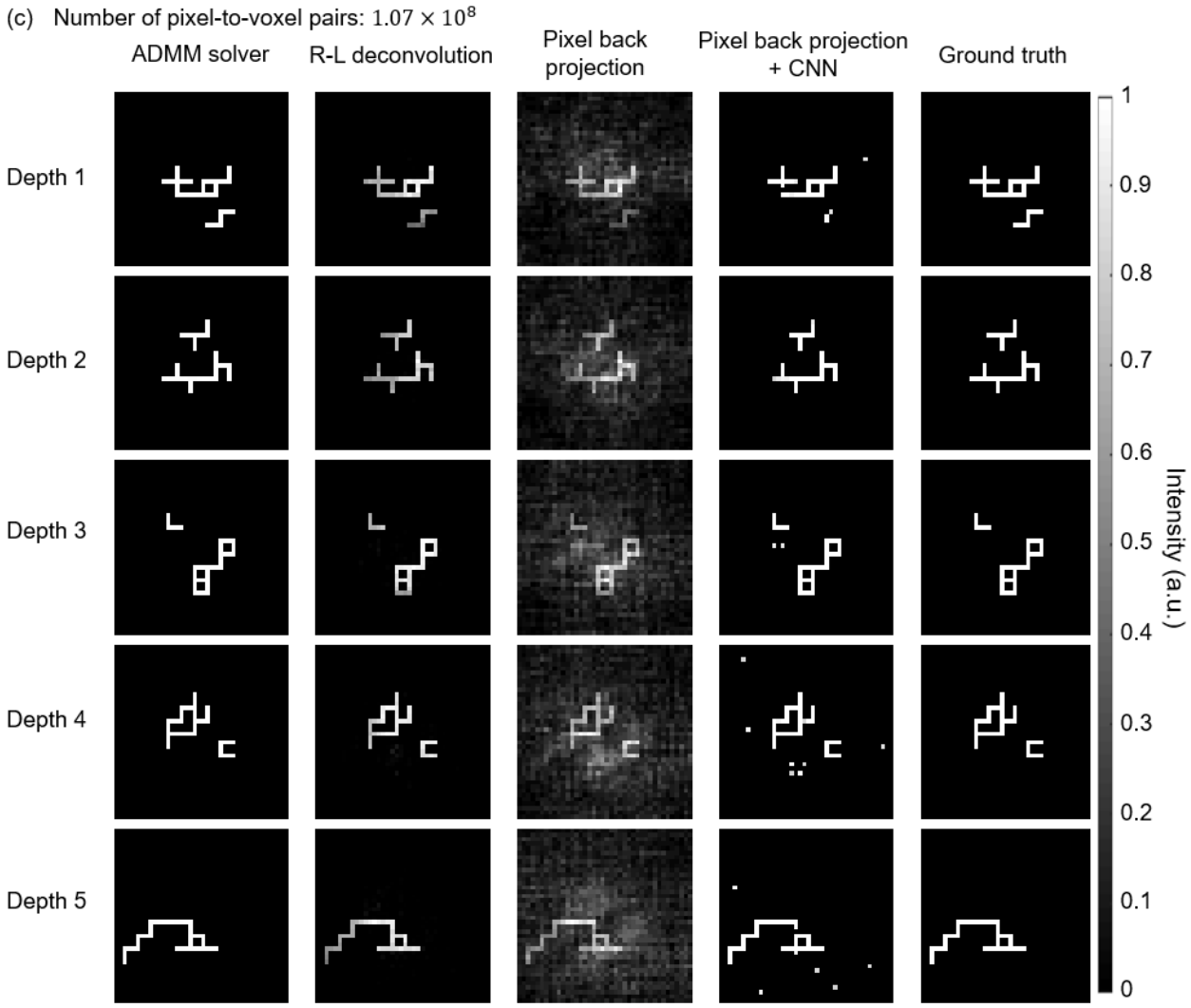

Figure S12. (c) Comparison of the reconstruction results (for 3D objects with continuous features) using ADMM solver, Richard-Lucy (R-L) deconvolution, and pixel back projection with and without convolutional neural network, for pixel-voxel pairs of 1.07×10^8.

Figure S12

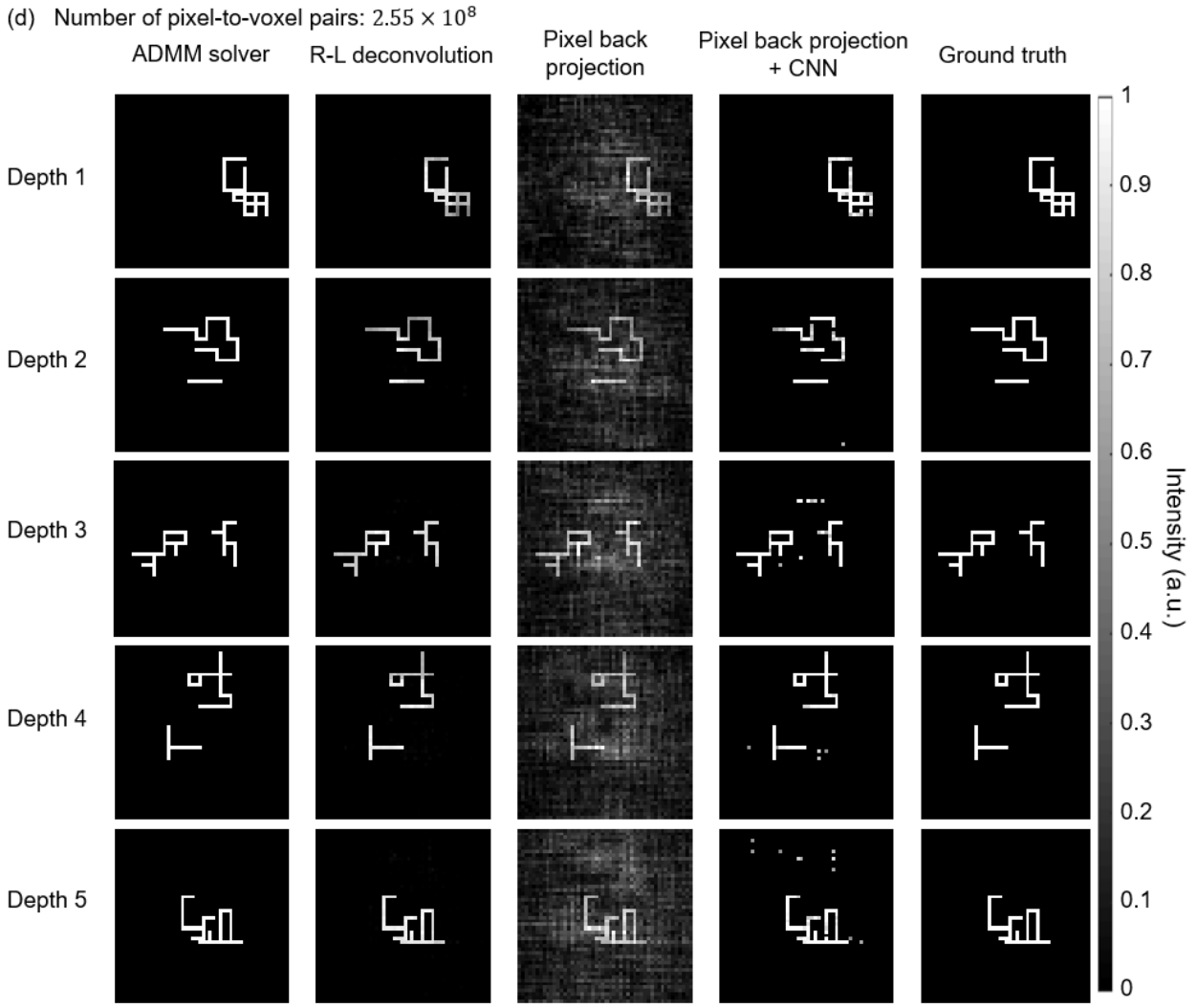

Figure S12. (d) Comparison of the reconstruction results (for 3D objects with continuous features) using ADMM solver, Richard-Lucy (R-L) deconvolution, and pixel back projection with and without convolutional neural network, for pixel-voxel pairs of 2.55×10^8.

Figure S12

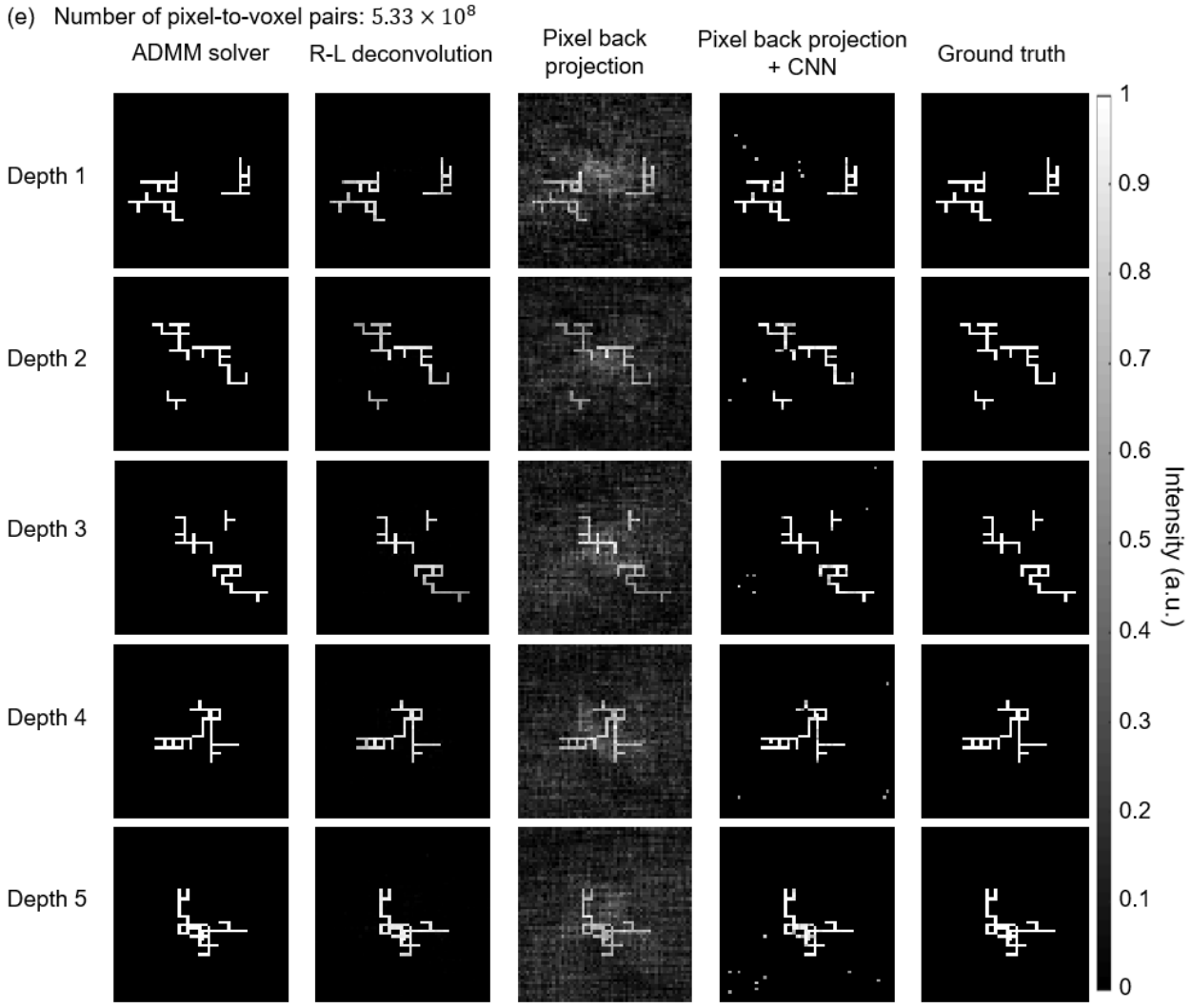

Figure S12. (e) Comparison of the reconstruction results (for 3D objects with continuous features) using ADMM solver, Richard-Lucy (R-L) deconvolution, and pixel back projection with and without convolutional neural network, for pixel-voxel pairs of 5.33×10^8.

Figure S12

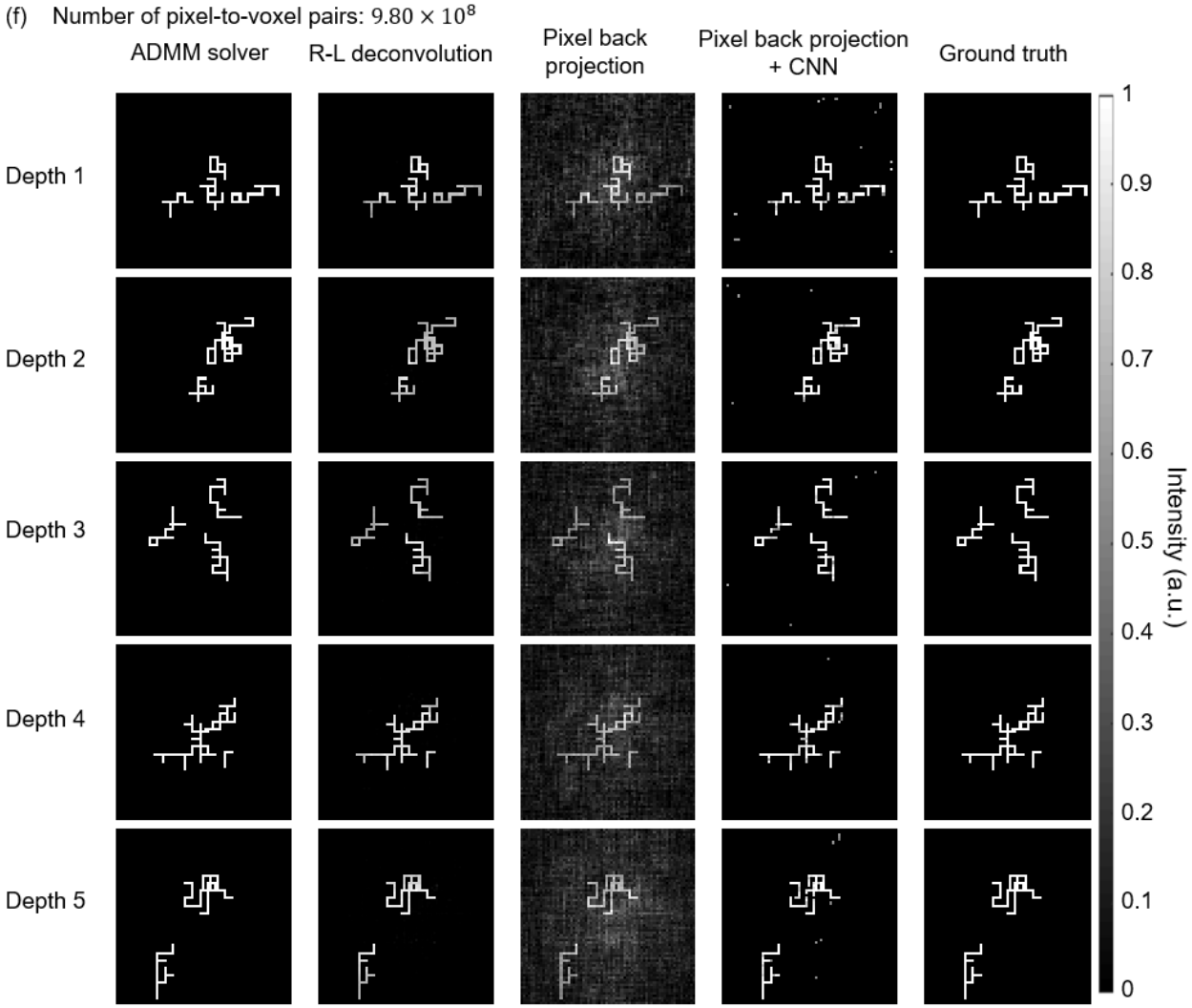

Figure S12. (f) Comparison of the reconstruction results (for 3D objects with continuous features) using ADMM solver, Richard-Lucy (R-L) deconvolution, and pixel back projection with and without convolutional neural network, for pixel-voxel pairs of 9.80×10^8.

References

1. H. M. Merklinger, *The INs and OUTs of FOCUS: An Alternative Way to Estimate Depth-of-Field and Sharpness in the Photographic Image* (1992).
2. J. K. Adams, V. Boominathan, B. W. Avants, D. G. Vercosa, F. Ye, R. G. Baraniuk, J. T. Robinson, and A. Veeraraghavan, "Single-frame 3D fluorescence microscopy with ultraminiature lensless FlatScope," Sci Adv **3**, e1701548 (2017).
3. N. Antipa, G. Kuo, R. Heckel, B. Mildenhall, E. Bostan, R. Ng, and L. Waller, "DiffuserCam: lensless single-exposure 3D imaging," Optica **5**, 1-9 (2018).
4. G. Kuo, F. L. Liu, I. Grossrubatscher, R. Ng, and L. Waller, "On-chip fluorescence microscopy with a random microlens diffuser," Opt Express **28**, 8384-8399 (2020).
5. Y. Xue, I. G. Davison, D. A. Boas, and L. Tian, "Single-shot 3D wide-field fluorescence imaging with a Computational Miniature Mesoscope," Sci Adv **6**, eabb7508 (2020).
6. K. Yanny, N. Antipa, W. Liberti, S. Dehaeck, K. Monakhova, F. L. Liu, K. L. Shen, R. Ng, and L. Waller, "Miniscope3D: optimized single-shot miniature 3D fluorescence microscopy," Light-Sci Appl **9**, 171 (2020).