Appendix

7. Additional Analysis

Parameter Study In Tab. 7, we provide an additional standard metric, end-point-error (EPE), for the same parameter study experiments in the main paper. For most benchmarks and experiments, the best model by px accuracy is also the best by EPE, demonstrating the robustness of our results.

8. Additional Generation Details

8.1. Indoor Floating Objects Scene Type

For each scene, we first generate the indoor scene using our modified Infinigen Indoors settings. Given this indoor scene, we place the random objects in the scene, respecting collisions with existing scene objects. We then place the 20 camera rigs using the default Infinigen Indoors placement algorithm, which attempts to maximize the standard deviation of depth values. Finally, we apply augmentations as described in Section 3. We randomize the materials of all floating and background objects, replacing an object's default material with a randomly chosen material with probability 0.5. To account for the large variance in object sizes, we normalize the floating object sizes. When removing fully transparent materials from the scene, we also take particular care to handle external windows - replacing the glass material for these with an opaque surface would ruin lighting for the rest of the scene, so we instead automatically delete the geometry to eliminate it from the ground truth.

8.2. Dense Floating Objects Scene Type

For each scene, we first spawn 200 objects into a scene at the origin, and use Blender's key-frame animation system to assign each object different locations, orientations, etc. at every single keyframe. Objects were placed at a distance sampled from Uniform[5,50] m away from the camera and the camera baselines were sampled from Uniform[0.1, 0.4] m. Similar to indoor scenes, we place point lights at random distances and normalize floating object sizes. We rescale object size based on distance to prevent far-away objects from being completely occluded by nearby objects and apply random warping to objects. We also perform additional lighting augmentations, randomizing the sky background's strength and color.

8.3. Nature Scene Type

We generate nature scenes with the Infinigen Nature system, using default configurations unless otherwise specified in this section and 3. By default, the Infinigen system provides configurations to render videos from a camera trajectory placed within a scene. However, these video sequences have high amounts of overlap between consecutive frames, so we render videos from camera trajectories at 6 frames per second to reduce similarity between consecutive frames. We render 50 frames from each video and sample the baseline of the stereo rig from Uniform[0.075, 0.5].

8.4. Ground Truth

For all frames, we provide camera intrinsics and extrinsics, left and right depth maps, and occlusion maps. We convert depth maps to disparity using the camera intrinsics and extrinsics. We compute occlusion maps by checking left-to-right and right-to-left re-projection consistency, with a 1 percent difference threshold. We also provide object segmentation maps, where each pixel has a value associated with an object's index value. Because procedural objects are never reused, we provide object names associated with each index, which enables the creation of flexible semantic segmentation maps. Similarly, materials are also procedural and never reused, so we provide material segmentation maps with material names. All frames are rendered at 1280 x 720 resolution.

8.5. Dataset post-processing

Certain types of undesirable scenes are hard to prevent during dataset generation, so we provide tools to remove them as a post-processing step. We use a simple mean of pixel intensities to remove frames that are extremely dark as a result of lighting augmentation. We filter frames with depth values lower than 12.5 cm to eliminate scenarios where the camera partially clips into an object. We optionally apply a more aggressive distance filter of 5 m for the nature split to filter out extremely high disparity samples, and release both the filtered and unfiltered nature split.

8.6. Atmospheric effects

We disable rain, dust, and snow particles from generation because they produce depth values that are almost invisible in rendered images. Similarly, we disable the coral reef scene type because the background objects are insufficiently visible.

8.7. List of Assets Used

ArmChair	Balloon
BarChair	BasketBase
Bathtub	BathroomSink
Bed	BedFrame
BeverageFridge	Blanket
BookColumn	Book
BookStack	Bottle
Bowl	Can
CeilingClassicLamp	CellShelf
Chair	Chopsticks
Cup	CurvedStaircase
DeskLamp	Dishwasher
FloorLamp	FoodBag
FoodBox	Fork
GlassPanelDoor	Hardware
Jar	KitchenCabinet
KitchenIsland	KitchenSpace
Knife	Lamp
LargePlantContainer	LargeShelf
Lid	LiteDoor
LouverDoor	Mattress
Microwave	Monitor
NatureShelfTrinkets	OfficeChair
Oven	Pallet
Pants	PanelDoor
Pillow	PlantContainer
Plate	Pot
Rug	Shirt
SimpleBookcase	SimpleDesk
Sink	SingleCabinet
Sofa	SpiralStaircase
Spoon	StraightStaircase
TableCocktail	TableDining
Tap	Towel
TriangleShelf	TV
TVStand	Vase
WallArt	Wineglass
UShapedStaircase	

Experiment	Method	Middlebury 2014				Middlebury 2021		ETH3D		KITTI				Booster (Q)	
		H		F		–		–		2012		2015		–	
		px	EPE	px	EPE	px	EPE	px	EPE	px	EPE	px	EPE	px	EPE
Floating Object Density	No floating Objects	12.52	3.91	18.44	12.13	16.31	3.29	4.47	0.36	4.42	0.93	6.19	1.20	16.40	4.14
	0 to 10 floating objects	7.78	2.19	11.42	5.28	11.30	1.66	**3.62**	0.31	4.44	0.91	6.09	1.15	12.21	2.64
	10 to 30 floating objects	**6.60**	**1.77**	**9.19**	**4.37**	**10.28**	**1.55**	3.92	**0.30**	**4.05**	**0.86**	**5.11**	**1.05**	**10.60**	**2.20**
Background Objects	No background objects	8.35	2.15	10.42	5.58	12.01	1.68	4.39	0.35	4.20	0.92	6.28	1.21	12.72	3.34
	With background objects	**6.60**	**1.77**	**9.19**	**4.37**	**10.28**	**1.55**	**3.92**	**0.30**	**4.05**	**0.86**	**5.11**	**1.05**	**10.60**	**2.20**
Floating Object Type	Floating chairs	**5.29**	1.50	9.24	5.84	9.81	1.41	3.64	0.29	4.90	1.01	7.02	1.22	11.22	2.71
	Floating shelves	7.13	1.89	10.28	5.05	10.24	1.43	3.51	0.31	4.32	0.94	6.06	1.16	11.63	2.31
	Floating bushes	6.04	**1.48**	9.21	3.83	**9.54**	**1.40**	**3.13**	**0.28**	**3.86**	0.86	5.42	1.11	12.19	2.72
	All generators used	6.60	1.77	**9.19**	**4.37**	10.28	1.55	3.92	0.30	4.05	**0.86**	**5.11**	**1.05**	**10.60**	**2.20**
Object Material	No materials	9.02	2.32	11.42	6.37	10.65	1.56	3.48	0.32	4.34	0.93	6.07	1.18	14.07	3.67
	One diffuse material	7.21	1.82	10.09	5.28	**9.65**	**1.43**	**2.77**	**0.29**	**3.76**	**0.84**	5.41	1.09	12.73	3.62
	Only metal and glass	8.37	2.06	11.28	6.47	11.85	1.73	4.95	0.34	4.06	0.86	**4.97**	**1.04**	**9.80**	**1.60**
	All materials used	**6.60**	**1.77**	**9.19**	**4.37**	10.28	1.55	3.92	0.30	4.05	0.86	5.11	1.05	10.60	2.20
Stereo Baseline	Uniform[0.04, 0.1]	9.60	4.03	32.47	33.02	22.18	14.70	**2.89**	**0.27**	5.13	1.05	6.64	1.22	17.03	5.22
	Uniform[0.2, 0.3]	7.01	2.17	9.75	5.39	10.50	1.72	14.05	6.00	**3.94**	**0.86**	5.37	1.10	**8.96**	**1.76**
	Uniform[0.04, 0.4]	**6.60**	**1.77**	**9.19**	**4.37**	**10.28**	**1.55**	3.92	0.30	4.05	**0.86**	**5.11**	**1.05**	10.60	2.20
Lighting	Realistic Lighting	6.91	**1.74**	9.62	**4.36**	**10.06**	**1.46**	**3.81**	0.31	**3.95**	0.88	5.45	1.10	**10.45**	**2.09**
	Augmented Lighting	**6.60**	1.77	**9.19**	4.37	10.28	1.55	3.92	**0.30**	4.05	**0.86**	**5.11**	**1.05**	10.60	2.20

Table 7. **Expanded metrics for parameter study**. We compute EPE on all valid pixels.

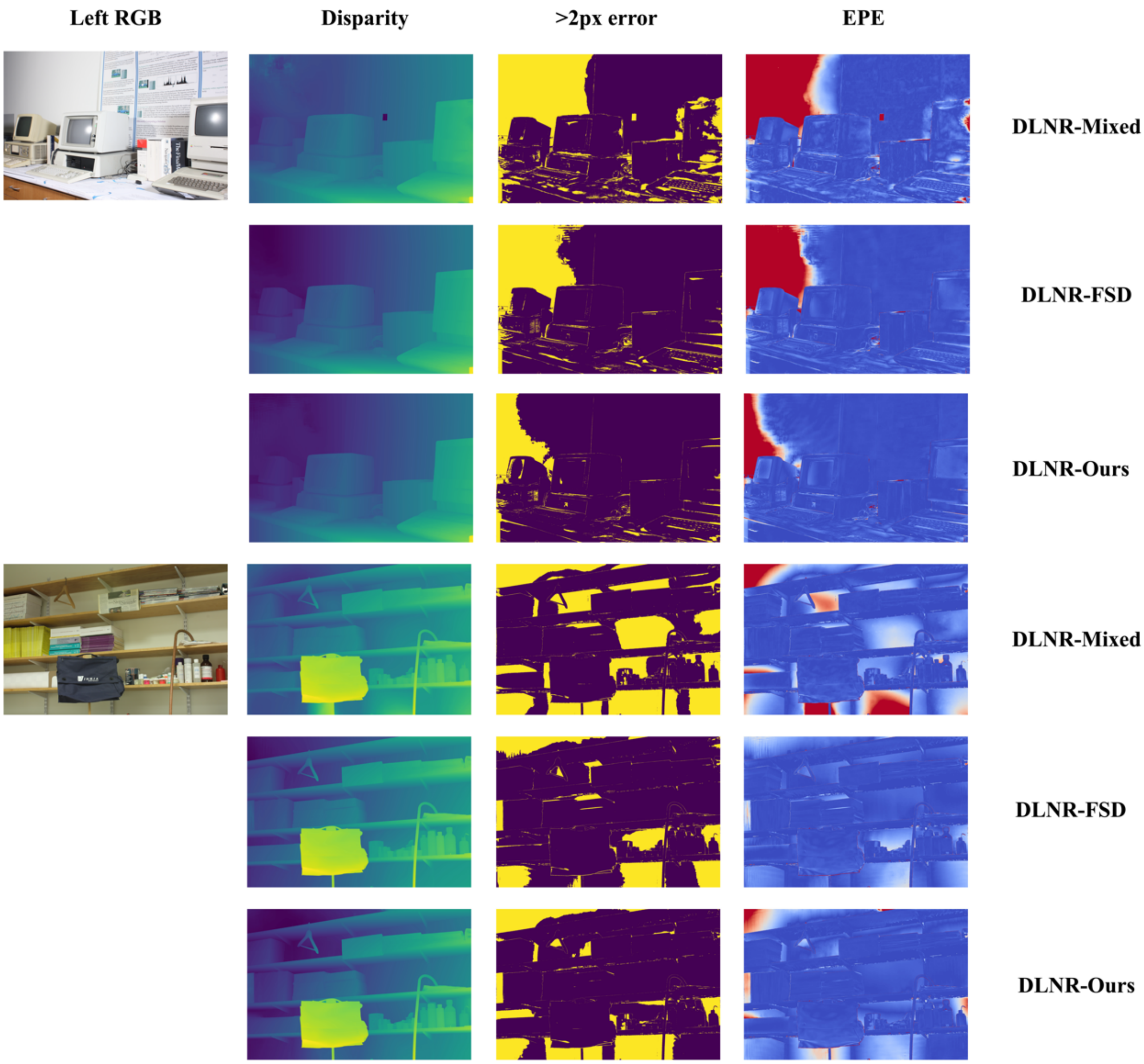

Figure 6. Qualitative results comparing DLNR-WMGStereo-150k, DLNR-FSD, and DLNR-Mixed on the 2014 Middlebury Eval set. Training on our dataset results in more accurate capture of walls and textureless regions than DLNR-Mixed and comparable performance with DLNR-FSD.

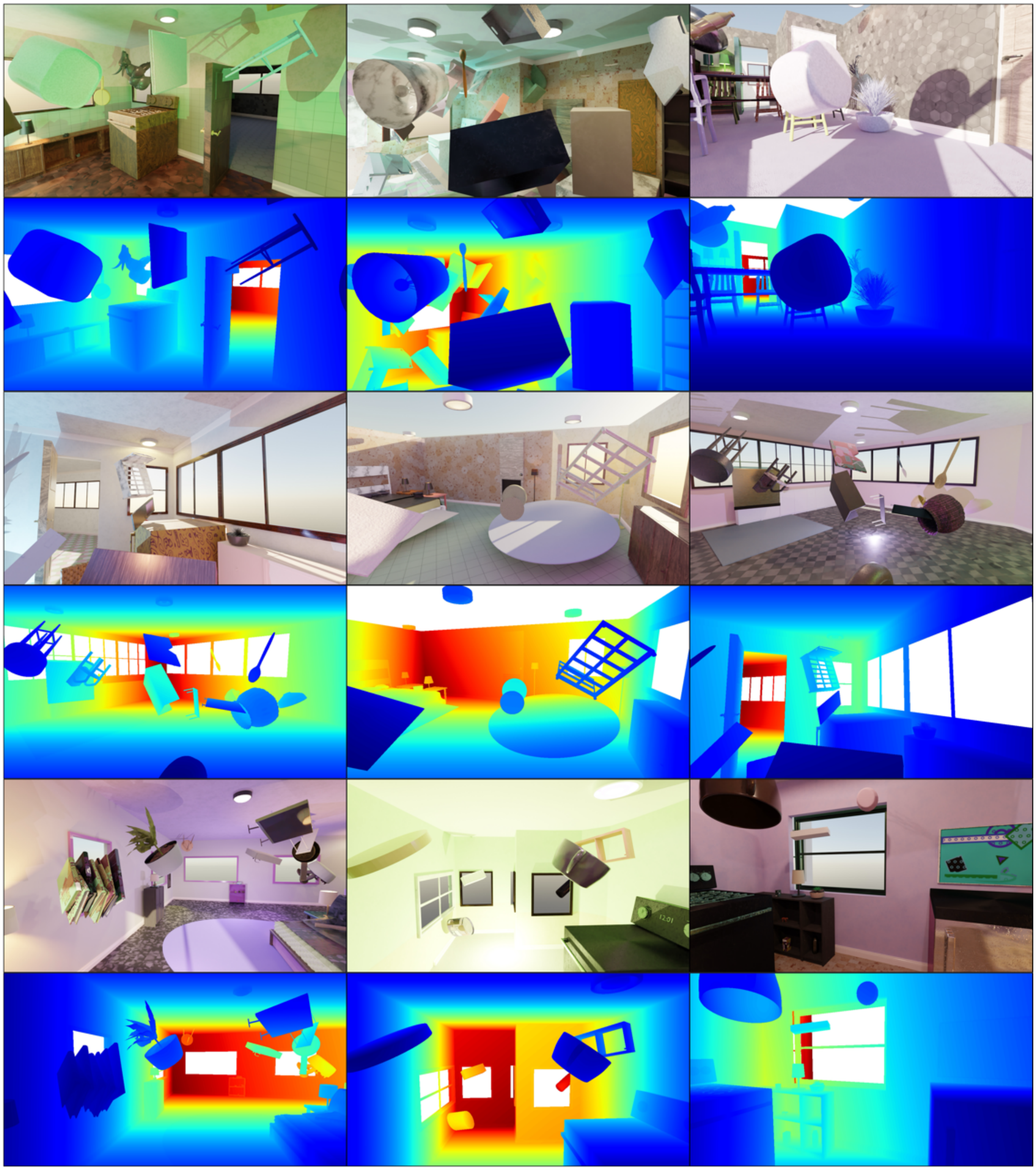

Figure 7. Non-cherrypicked, random samples from our floating indoor dataset type. We remove the glass portion of the window mesh entirely, instead of replacing the material, in order to maintain good lighting, as specified in Section 3. This modification is a configurable option, and new data can be generated with the window meshes in place.

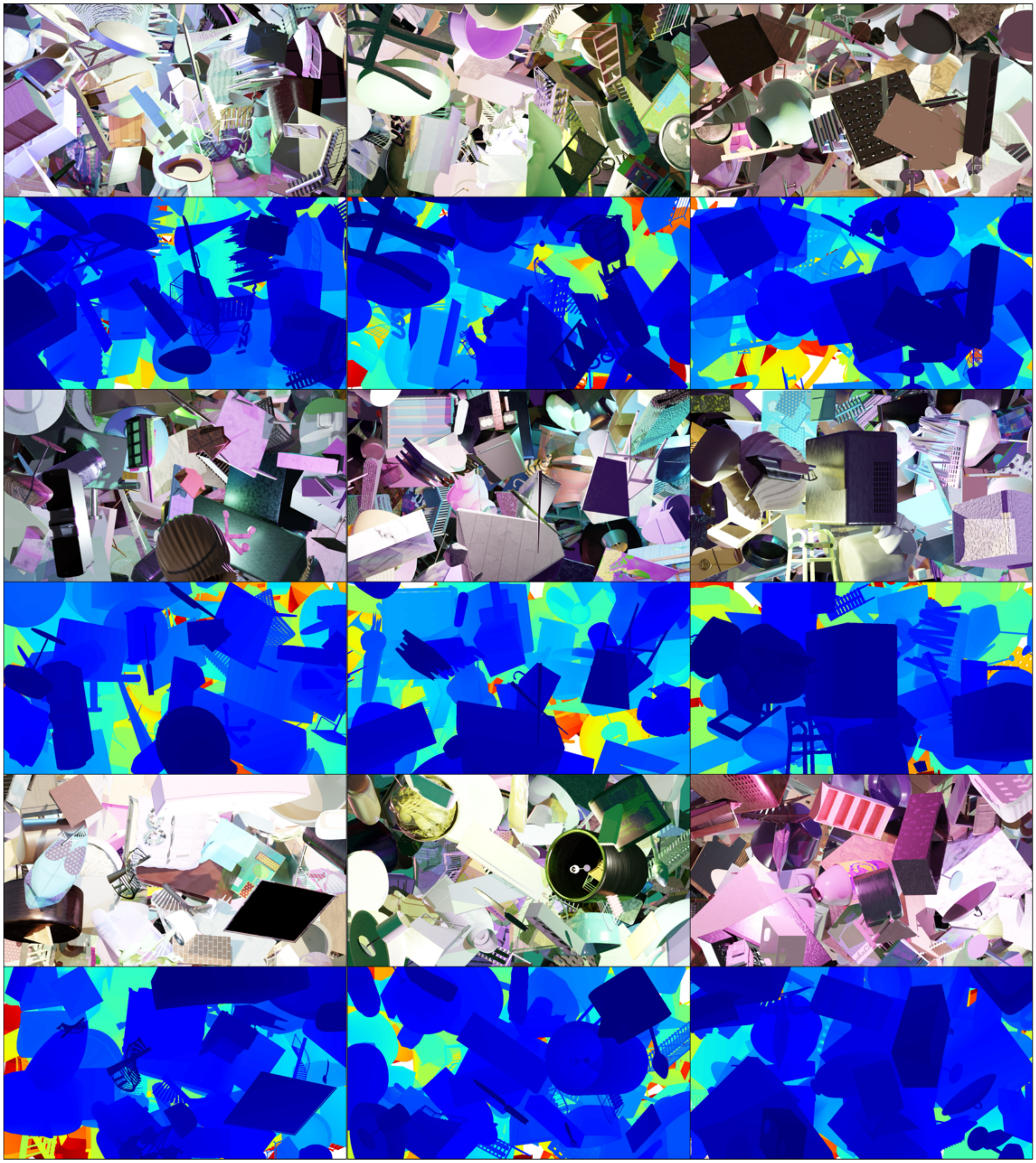

Figure 8. Non-cherrypicked, random samples from our dense floating dataset type.

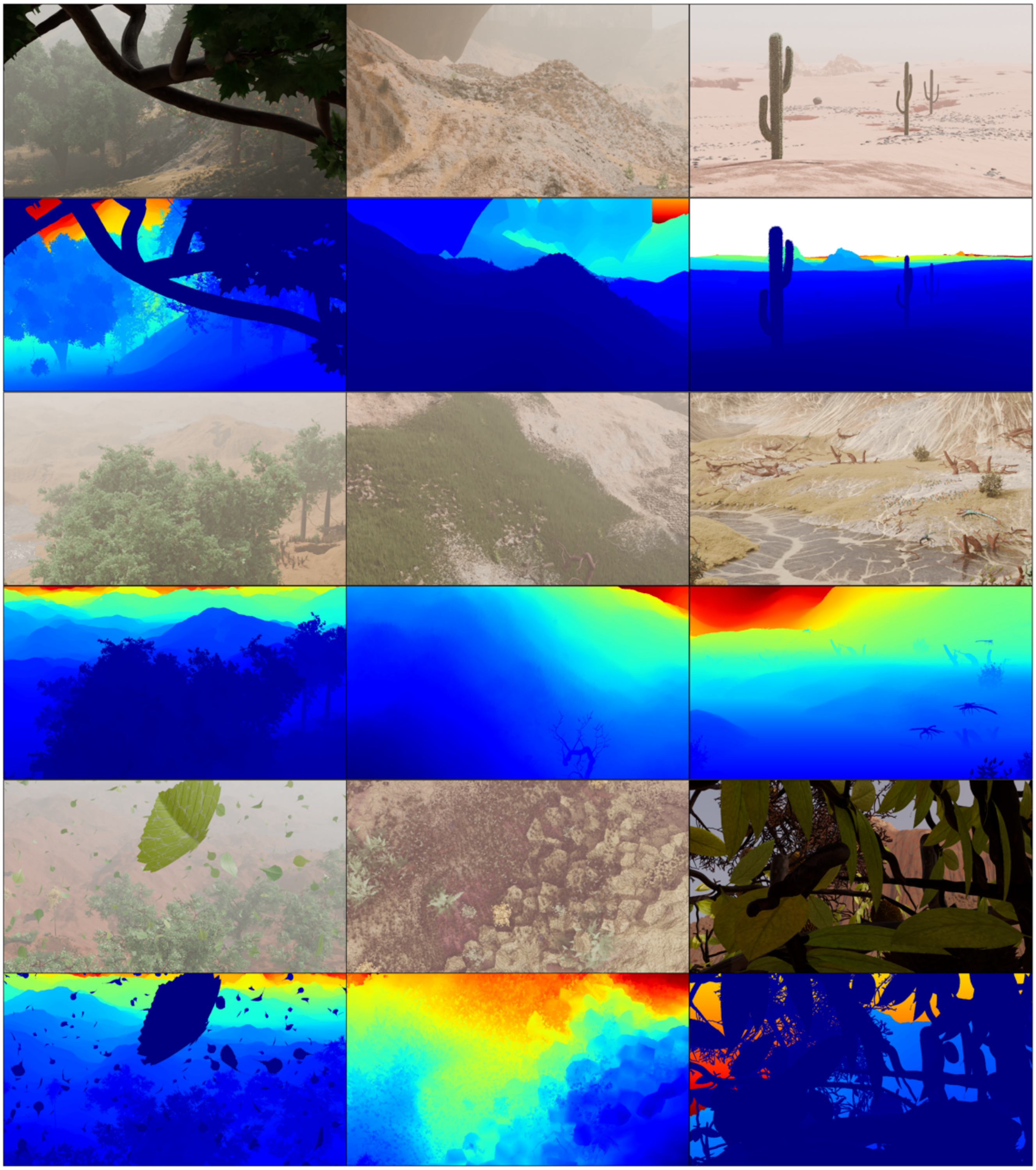

Figure 9. Non-cherrypicked, random samples from our nature dataset type.