Supplementary Material
DeepVoxels++: Enhancing the Fidelity of Novel View Synthesis from 3D Voxel Embeddings

Tong He[1], John Collomosse[2,3], Hailin Jin[2], Stefano Soatto[1]

[1]UCLA. [2]Adobe Research. [3]CVSSP, University of Surrey, UK.

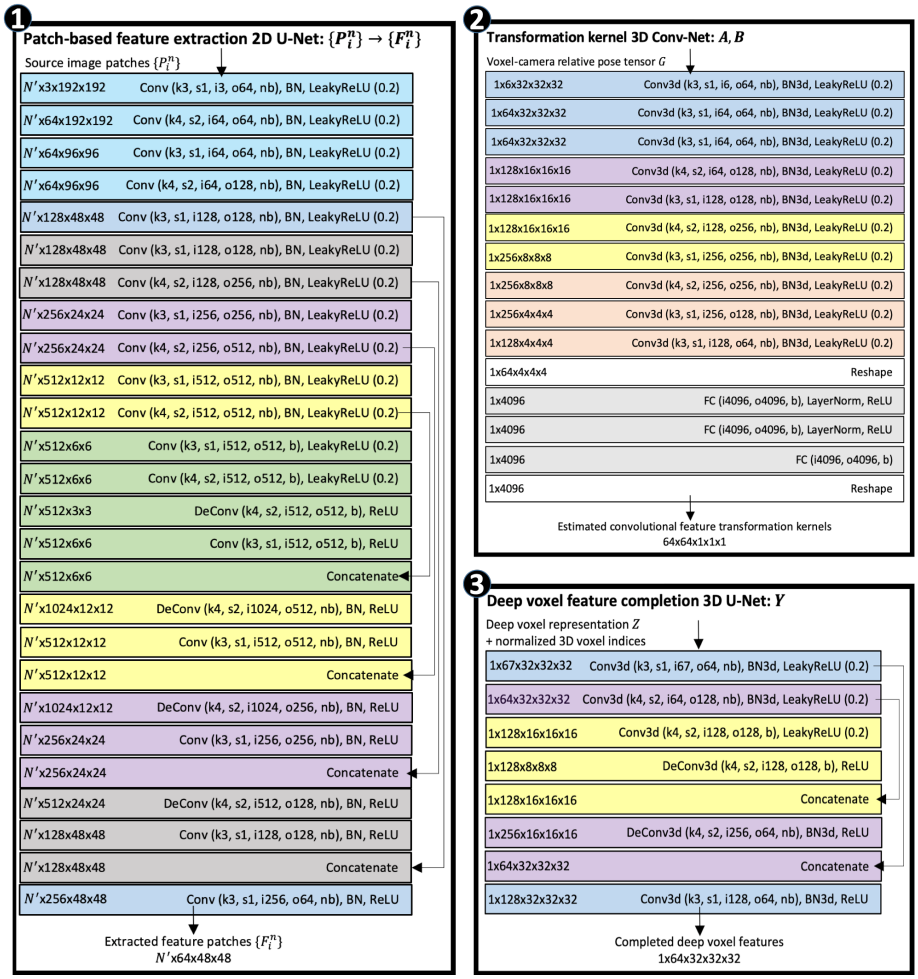

Fig. 1. 1) Patch-based feature extraction 2D U-Net; 2) Feature transformation kernel estimation 3D convolutional network; 3) Deep voxel feature completion 3D U-Net. Conv(k, s, i, o, (n)b) are (kernel size, stride, input, output, (no) bias), and BN means batch normalization. Please zoom in for details.

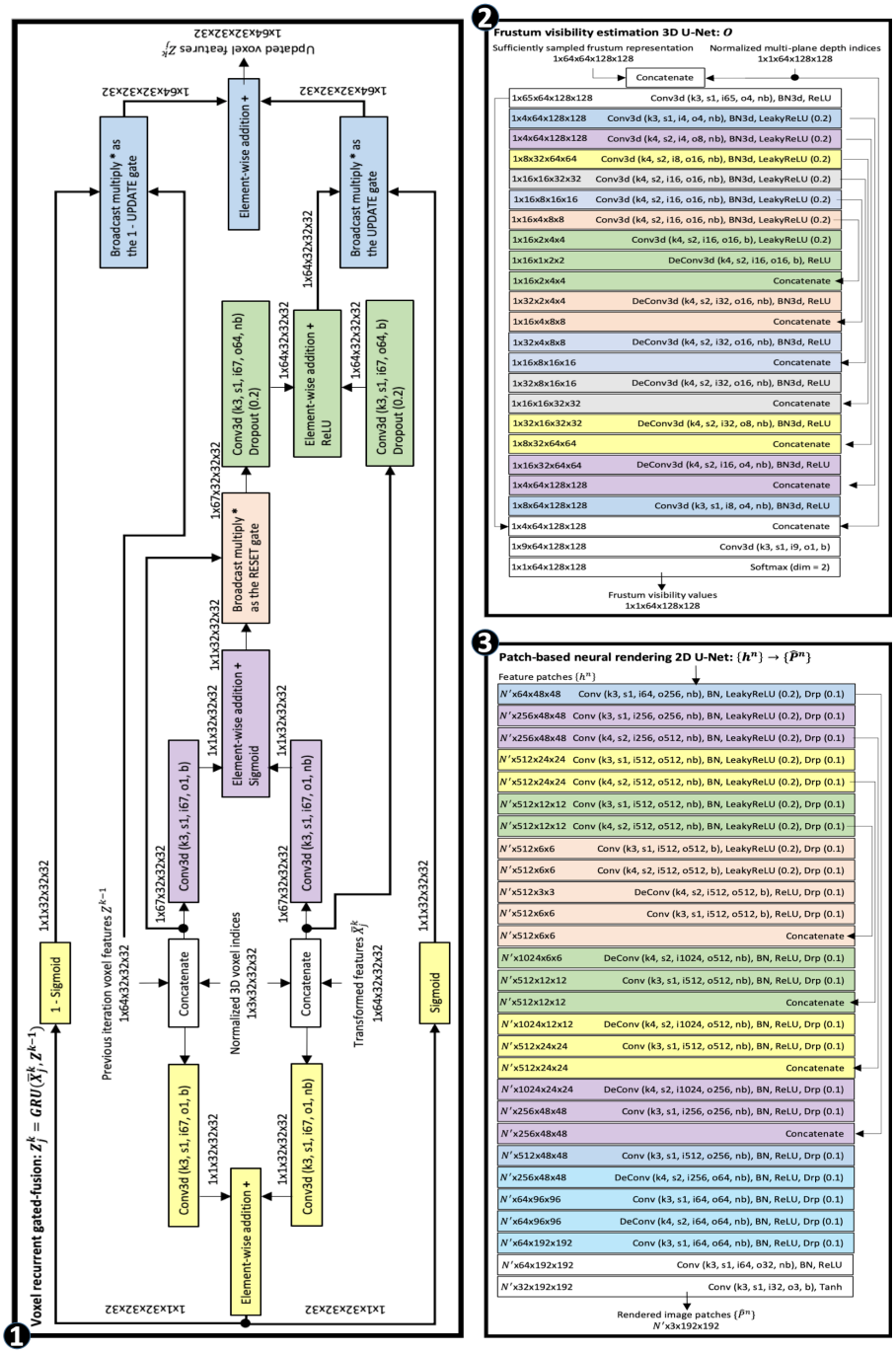

Fig. 2. 1) Voxel feature recurrent-concurrent aggregation; 2) Frustum visibility estimation 3D U-Net; 3) Patch-based neural rendering 2D U-Net. Please zoom in for details.

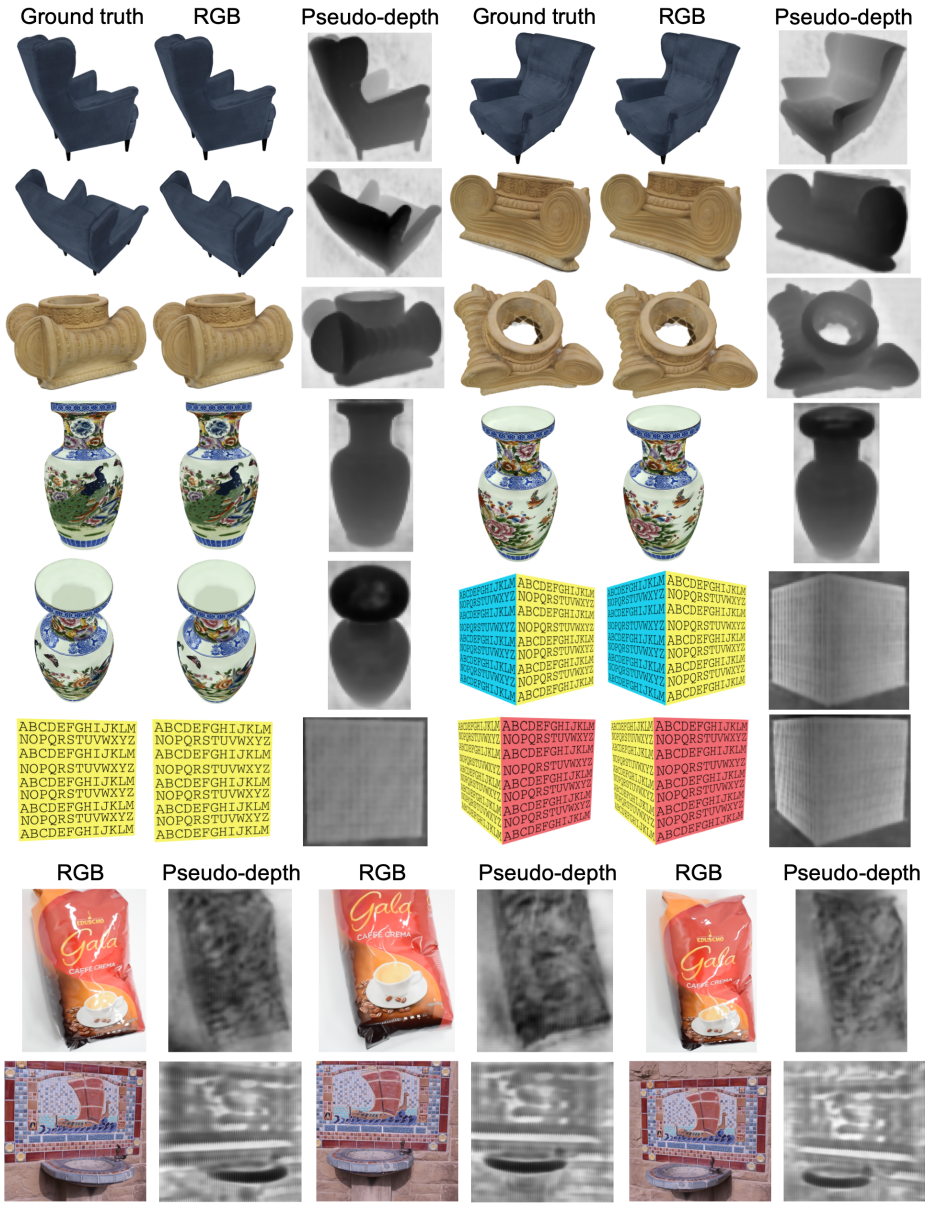

Fig. 3. Novel-view synthesis results (RGB) and the corresponding pseudo-depth maps of objects with large viewpoint changes. Although our method DeepVoxels++ is designed for objects with diffuse reflectance, we show some preliminary results on specularity modeling. Our proposed method of learning voxel feature transformation kernels potentially can also model other view-dependent effects (*e.g.* specularity) besides image-plane perspective transformations of diffuse surfaces. But the pseudo-depth maps (*e.g.* coffee bag) are noisy due to specularity. As future work, it is worth conducting further evaluations under various lighting situations on objects with specular reflectance.

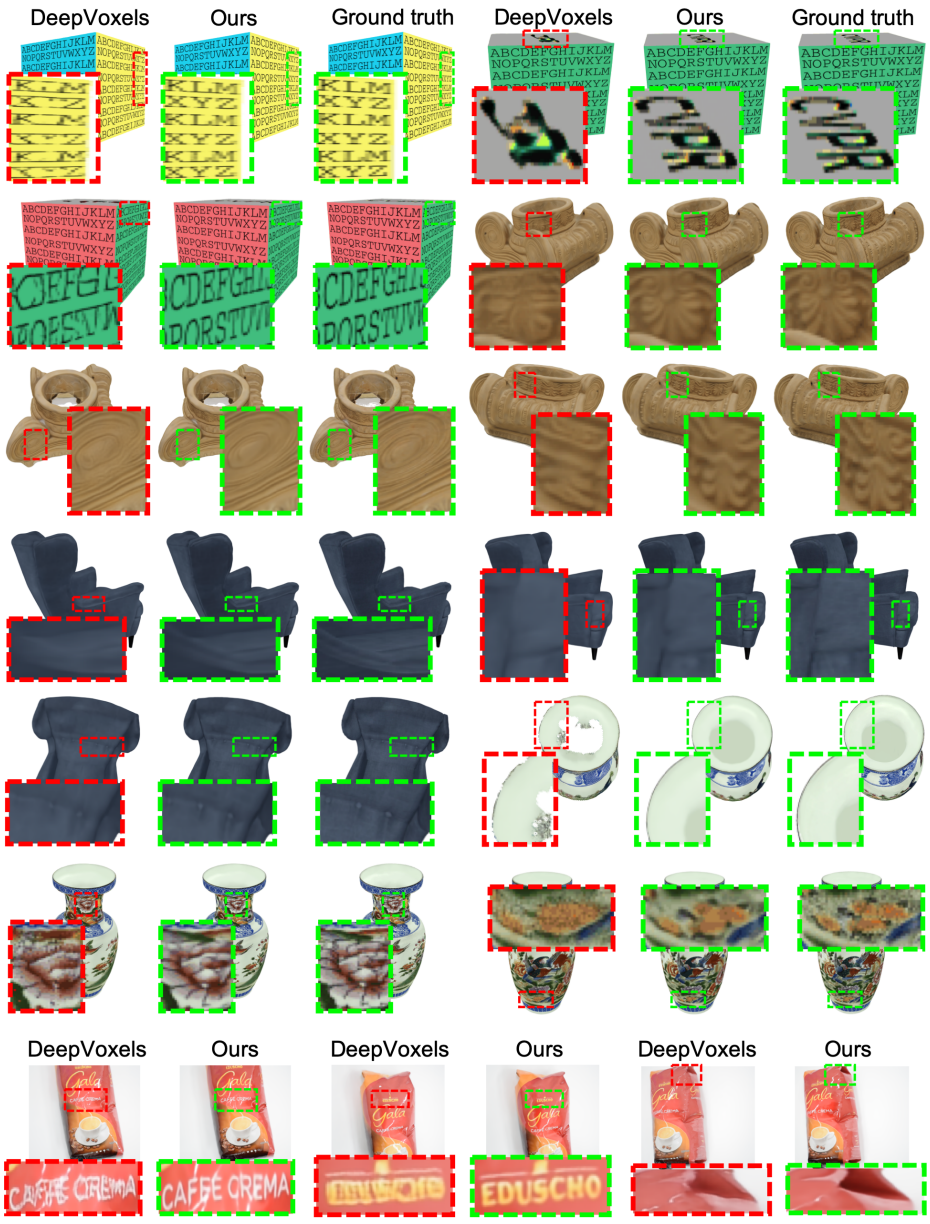

Fig. 4. Visual comparisons between the prior best voxel-based method DeepVoxels and our method, DeepVoxels++. In contrast with DeepVoxels, our results have fewer rendering artifacts like aliasing and holes of the vase, *and* preserve more texture/shape details such as letters on the cube and fine structures of the Greek pedestal. As discussed in ablation studies of the main paper, the enhanced novel view synthesis fidelity is attributed to a series of technical improvements as well as an implementation trick of 3D voxel embeddings sufficient sampling.

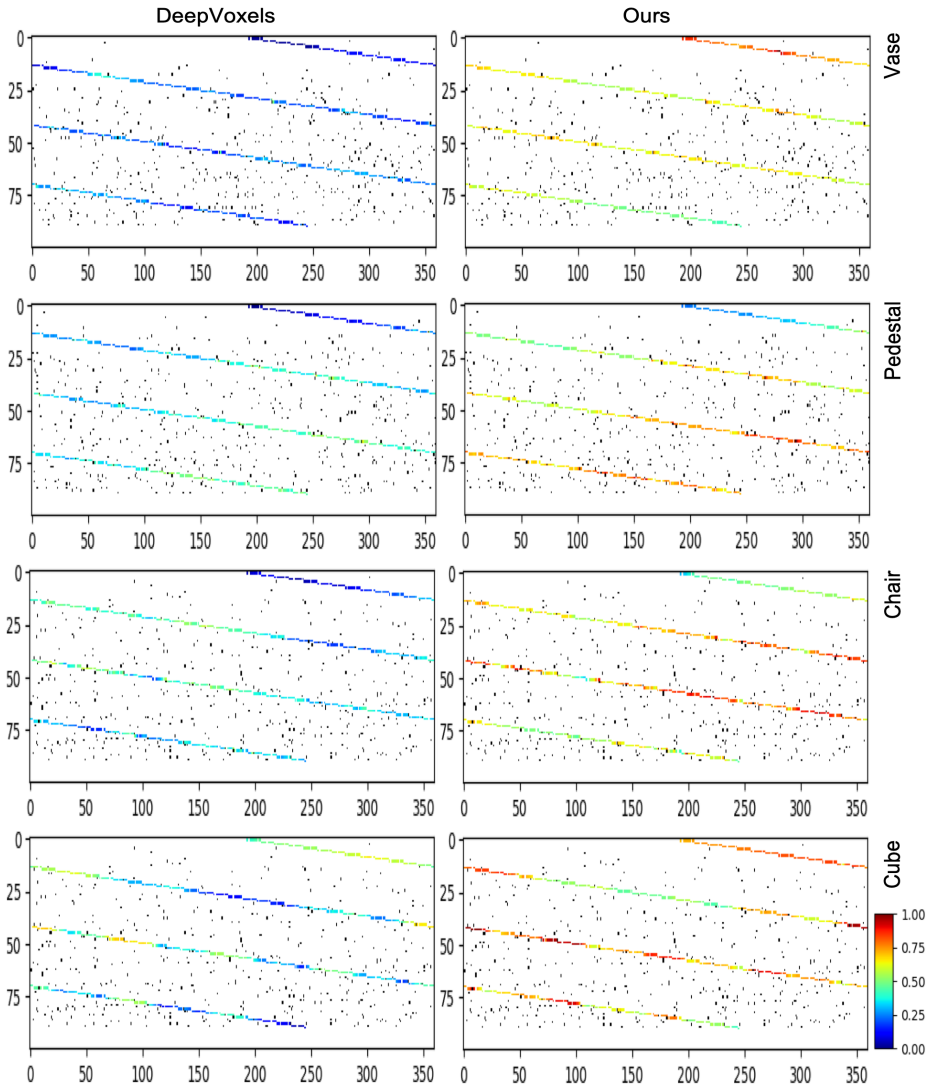

Fig. 5. Normalized azimuth-elevation PSNR maps. Each object is normalized independently by the largest PSNR value of DeepVoxels' and ours novel-view synthesis results. Horizontal: $[0°, 360°]$ azimuth. Vertical: $[0°, 100°]$ elevation. Black dots are the training poses. Colored spiral lines are the test pose trajectories. Red color means large normalized PSNR value and blue means small. These plots prove that our improvement over DeepVoxels is due to consistently improved rendering quality across 1000 dense test views of the objects, not caused by over-fitting at certain viewpoints that are close to the training data. They also showcase smooth viewpoint interpolation paths between training views (*i.e.* black dots).