# DReg-NeRF: Deep Registration for Neural Radiance Fields - Supplementary Material

Yu Chen    Gim Hee Lee
Department of Computer Science, National University of Singapore
chenyu@comp.nus.edu.sg    gimhee.lee@nus.edu.sg

## 1. NeRF Network Architecture

We present the network architecture of our used NeRF network in Fig. 1. The resolution level is 16. The number of hash table entries in each level is $2^{19}$, where the feature dimension of each hash table entry is 2. The coarsest level is 16. We use NeRFAcc [2] to train NeRF models, where only a single resolution occupancy grid is used to skip empty space instead of multi-resolution occupancy grids as in the original InstantNGP implementation.

## 2. More Qualitative Results

We present more qualitative results in Fig. 2. We further visualize the rendered RGB images and depth images in Fig. 3, Fig. 4 and Fig. 5. In the left part of each figure, we visualize the rendered RGB images and depth images, where the top row shows results from ground truth transformation, the middle row shows results from the predicted transformation, and the bottom row shows results without applying the transformation. The right part of each figure presents camera poses and occupancy grids before registration on the top row, and the camera poses and occupancy grids after registration on the bottom row.

## References

[1] Matt Deitke, Dustin Schwenk, Jordi Salvador, Luca Weihs, Oscar Michel, Eli VanderBilt, Ludwig Schmidt, Kiana Ehsani, Aniruddha Kembhavi, and Ali Farhadi. Objaverse: A universe of annotated 3d objects. *CoRR*, abs/2212.08051, 2022.

[2] Ruilong Li, Matthew Tancik, and Angjoo Kanazawa. Nerfacc: A general nerf acceleration toolbox. *CoRR*, abs/2210.04847, 2022.

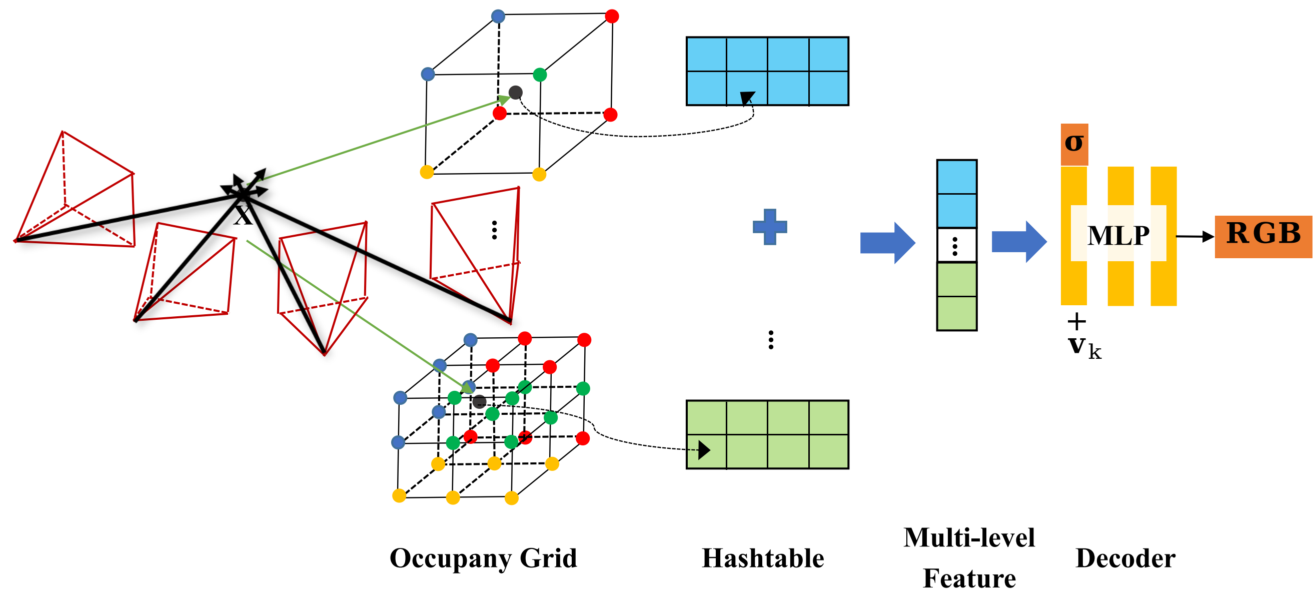


Figure 1: **Network architecture of our used NeRF**. A single-resolution occupancy grid is used to skip empty space. The dimension of each hidden layer is 64. The view direction is concatenated with the feature embedding after the first hidden layer without applying positional encodings.

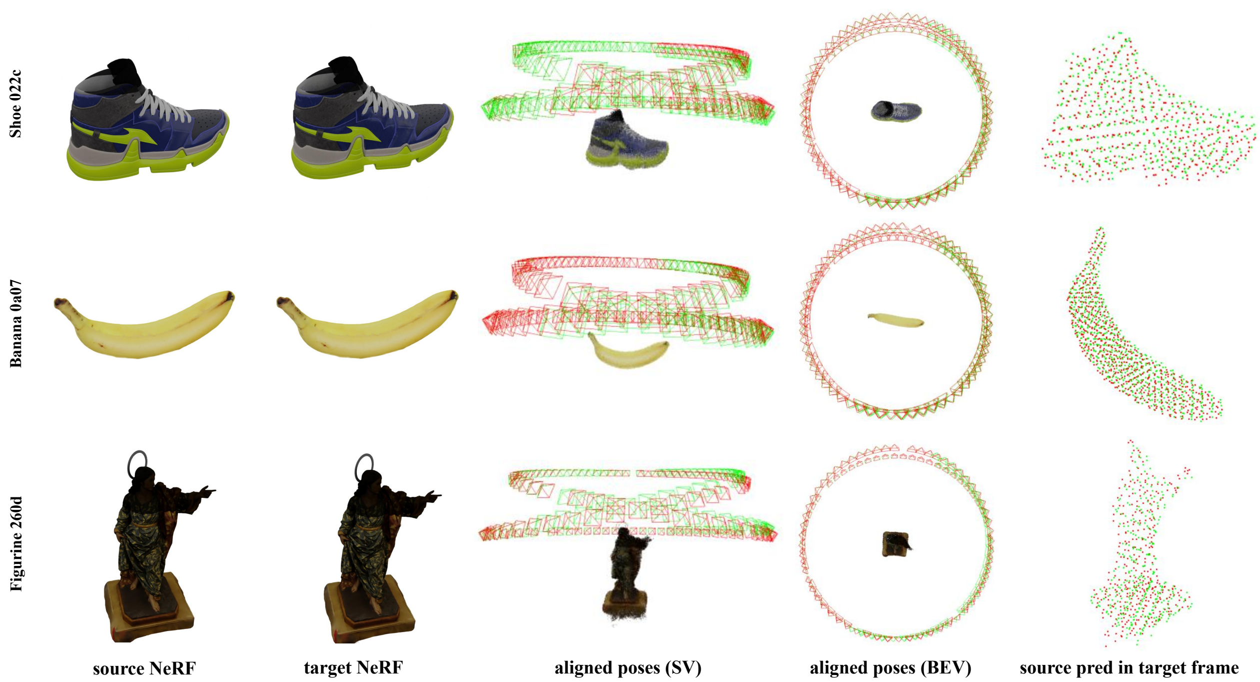


Figure 2: **The qualitative results on Objaverse [1] dataset after NeRF registration**. From left to right are respectively the rendered images by the source NeRF model, the rendered images by the target NeRF model, the side view (SV) of the aligned camera poses, the birds-eye-view (BEV) of the aligned camera poses, the concatenated predictions by transforming the source prediction to the target NeRF's coordinate frame. red and green, respectively, denote the results from source NeRF and target NeRF.

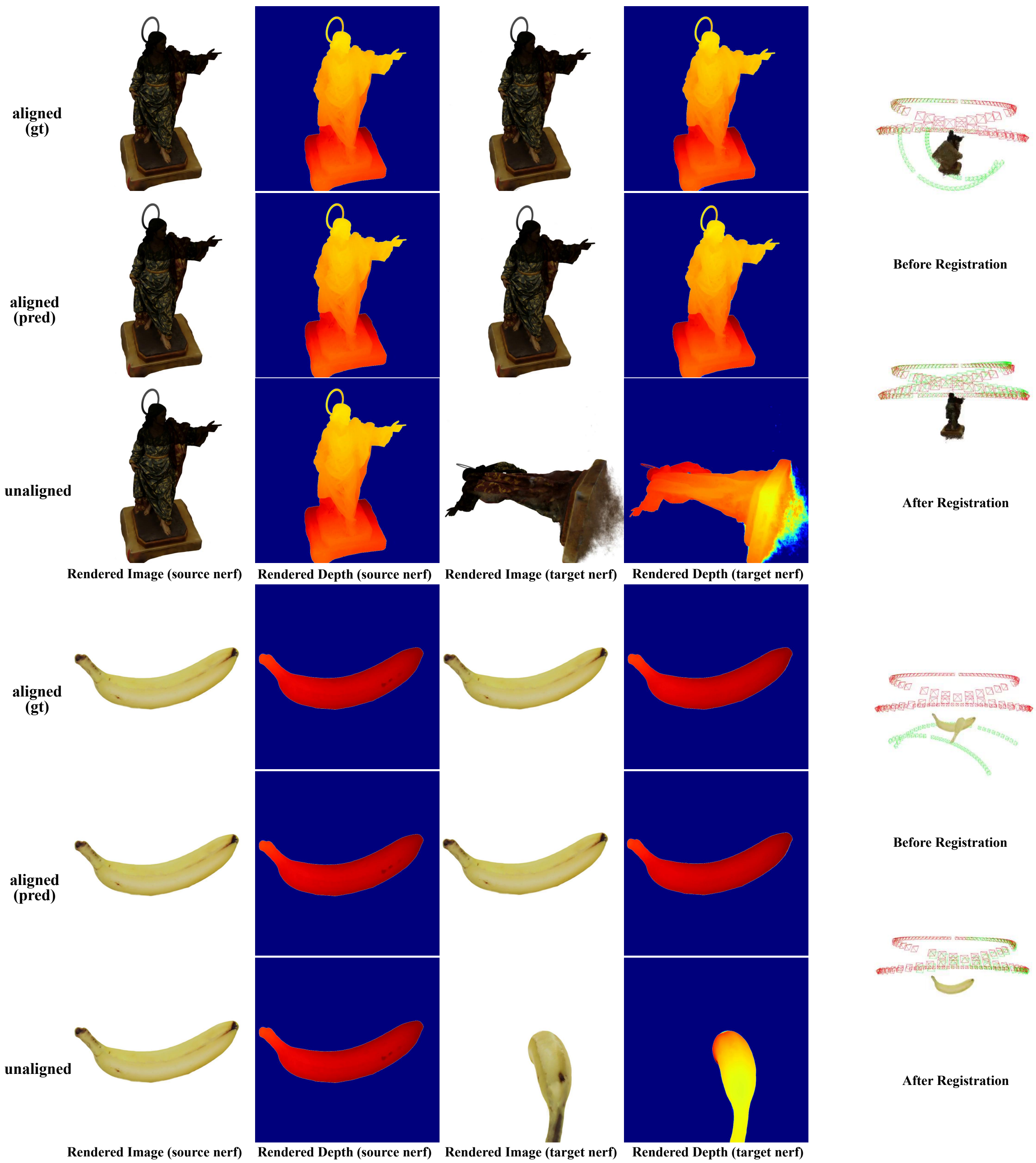


Figure 3: **View synthesis comparison.** Left: (1) Top row: results from ground truth transformation; (2) Middle row: results from the predicted transformation; (3) Bottom row: results without applying transformation. Right: (1) Camera poses and occupancy grids before registration; (2) Camera poses and occupancy grids after registration.

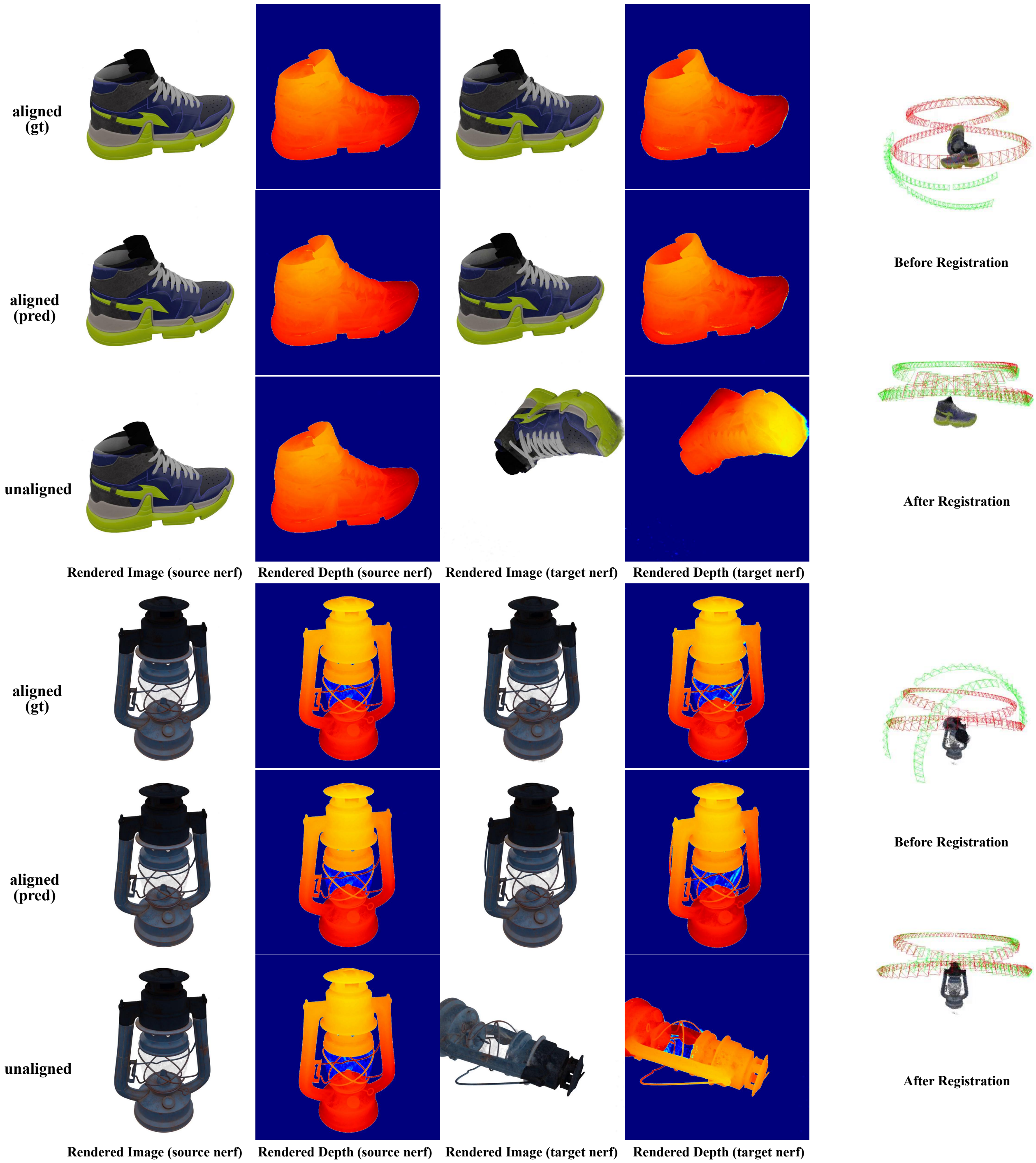


Figure 4: **View synthesis comparison on object.** Left: (1) Top row: results from ground truth transformation; (2) Middle row: results from the predicted transformation; (3) Bottom row: results without applying transformation. Right: (1) Camera poses and occupancy grids before registration; (2) Camera poses and occupancy grids after registration.

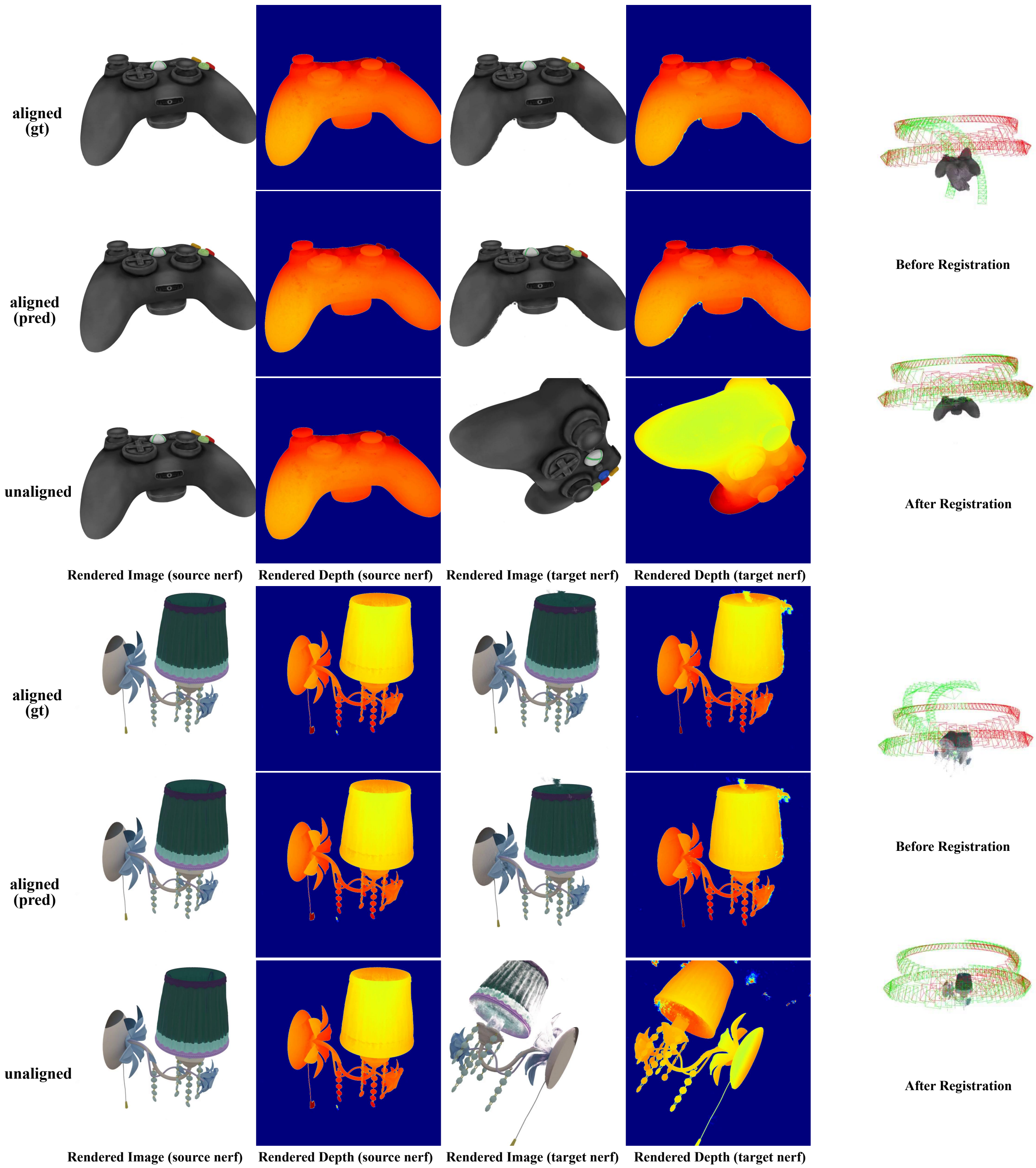


Figure 5: **View synthesis comparison.** Left: (1) Top row: results from ground truth transformation; (2) Middle row: results from the predicted transformation; (3) Bottom row: results without applying transformation. Right: (1) Camera poses and occupancy grids before registration; (2) Camera poses and occupancy grids after registration.