Learning a Predictable and Generative Vector Representation for Objects

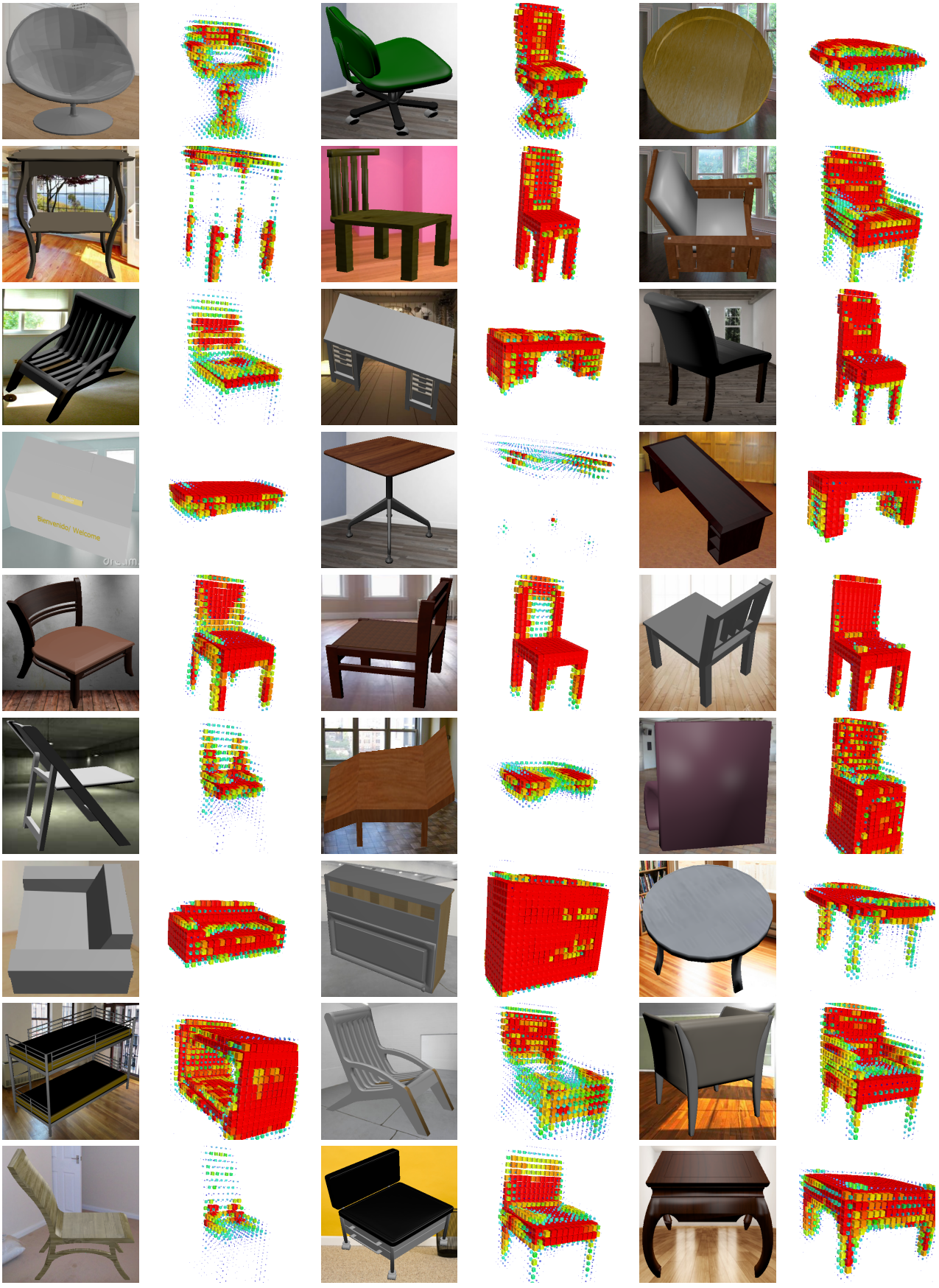
Supplementary Material

This is a abridged low-quality version. Full high-res version available on project webpage.

Contents

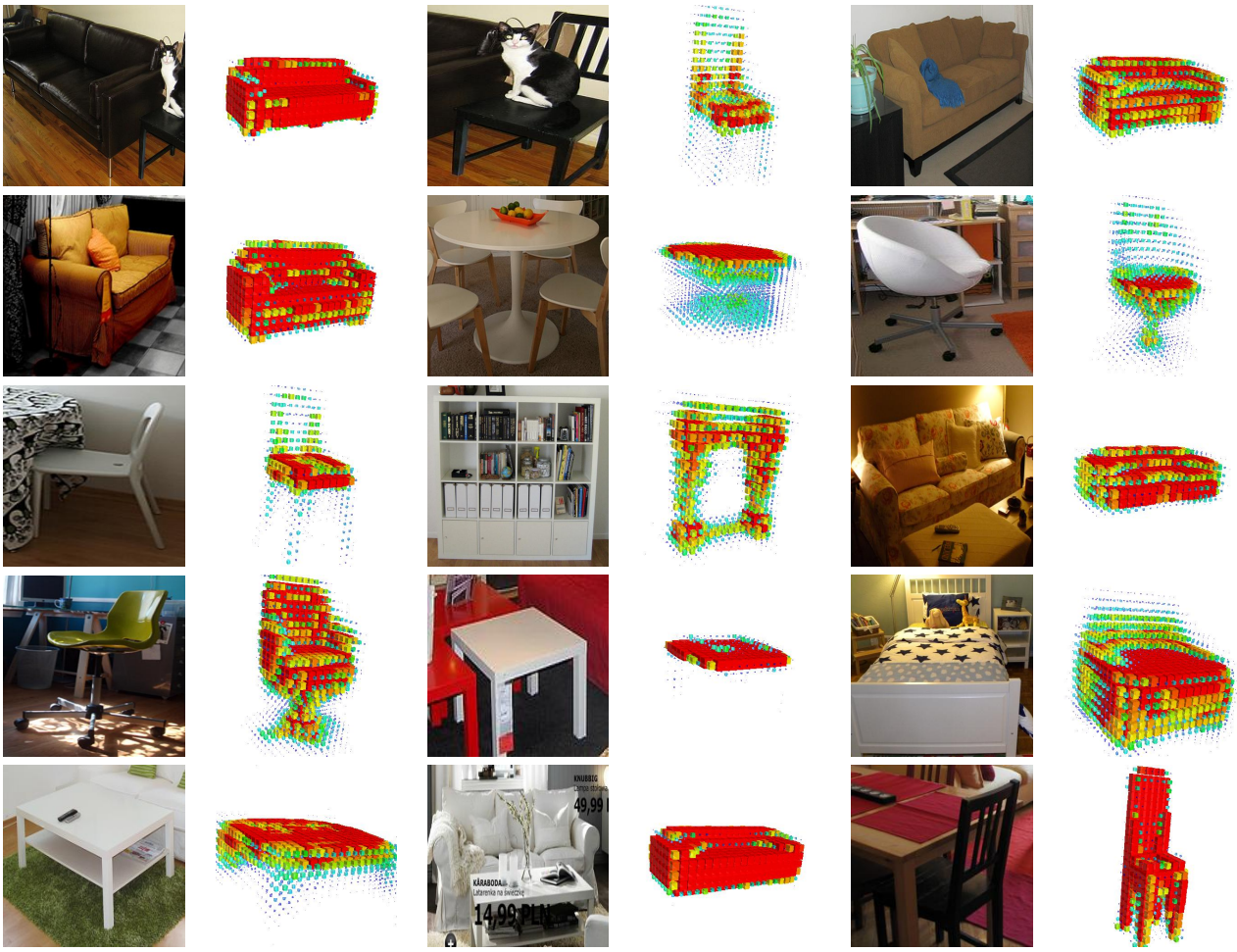
1 Reconstruction Results on Synthetic Test Data

Some **randomly picked images** and corresponding reconstructions from the 23975 renderings of 4058 test models.



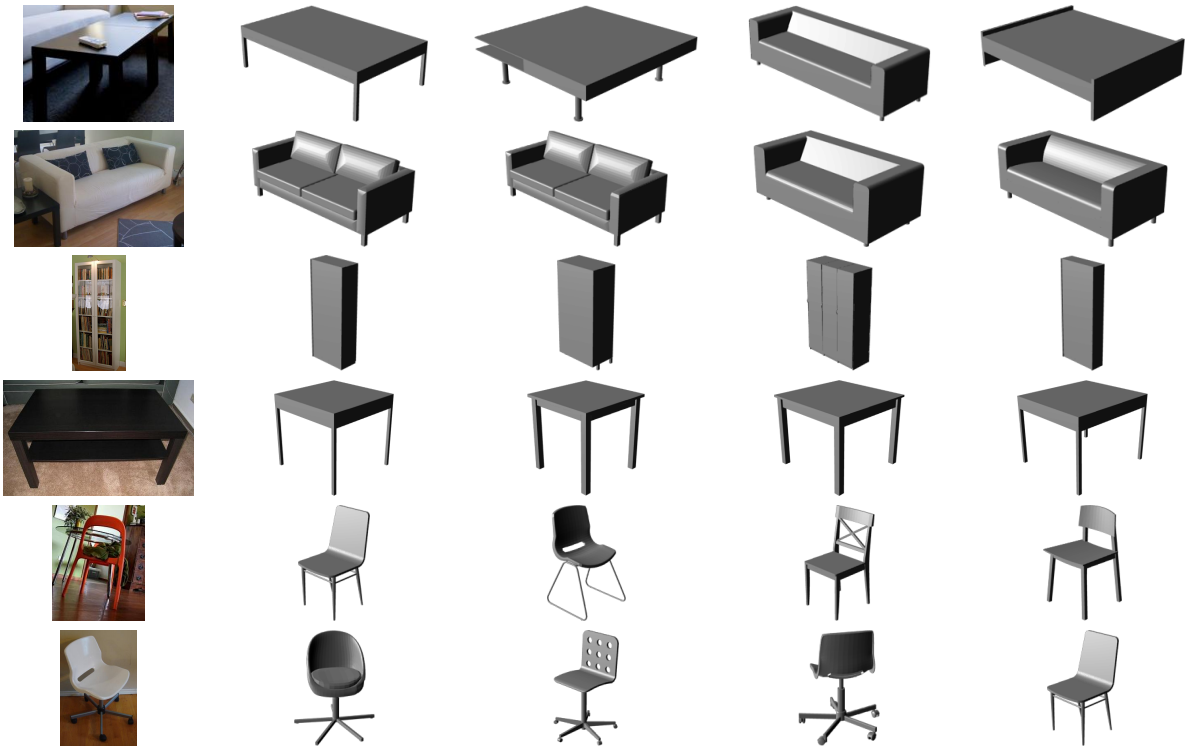
2 Reconstruction Results on Natural Images from IKEA Dataset

Select natural images and reconstructions on IKEA dataset.



3 Nearest Neighbor on Natural Images from IKEA Dataset

Select natural images and 3D model nearest neighbors in IKEA dataset.



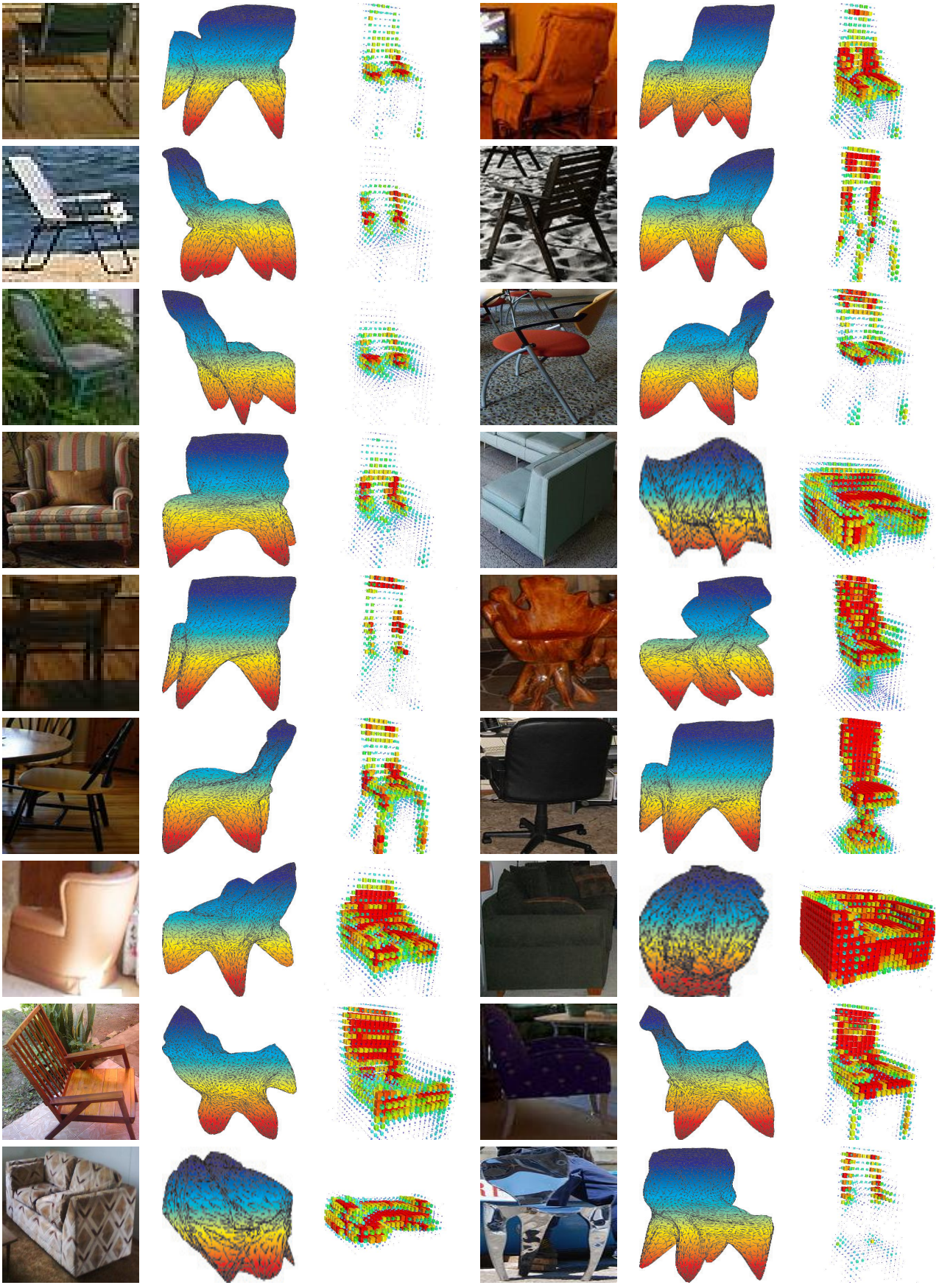
4 Comparison with Kar *et al.* [1] (3D Prediction)

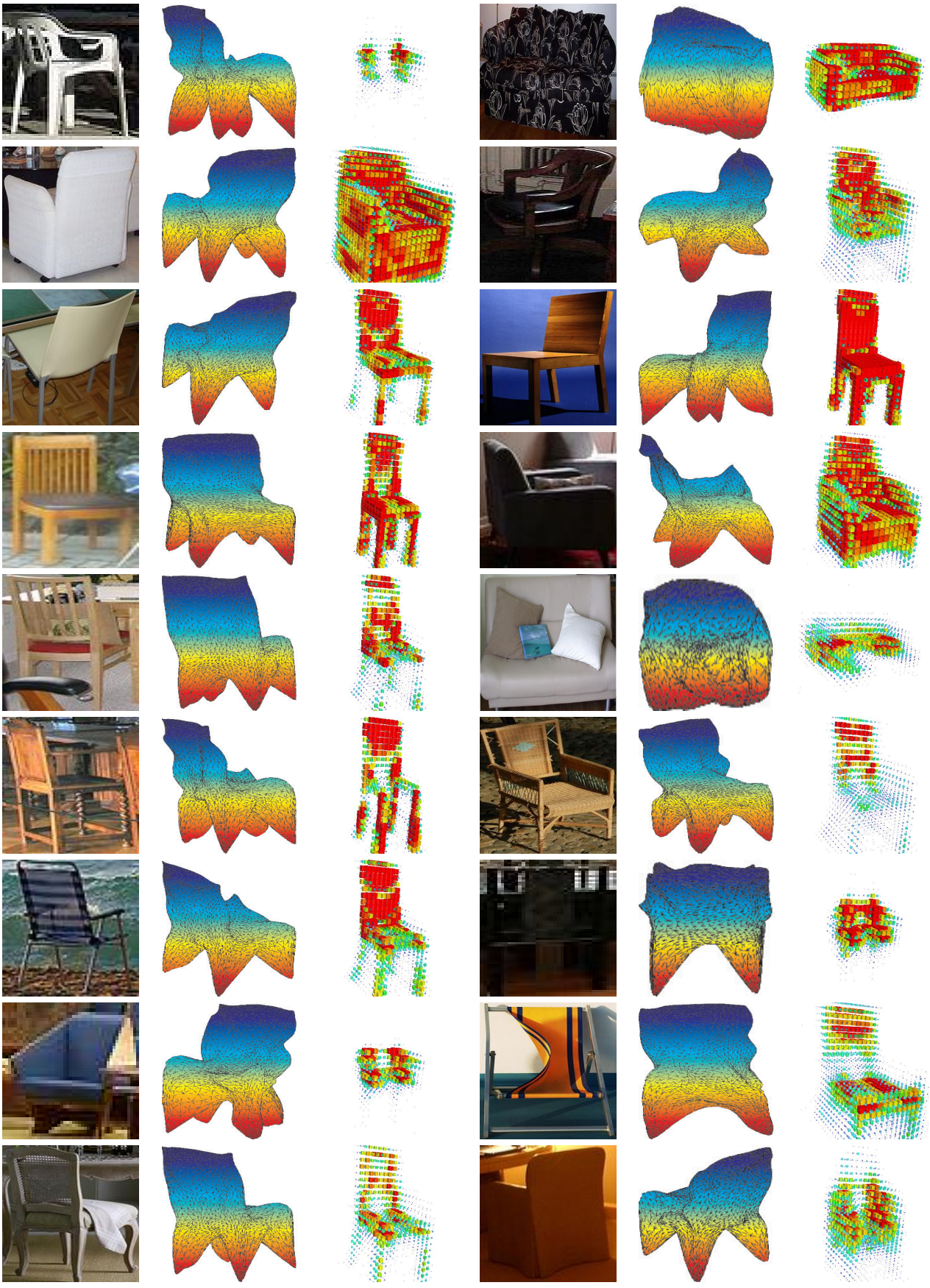
4.1 Quantitative Evaluation

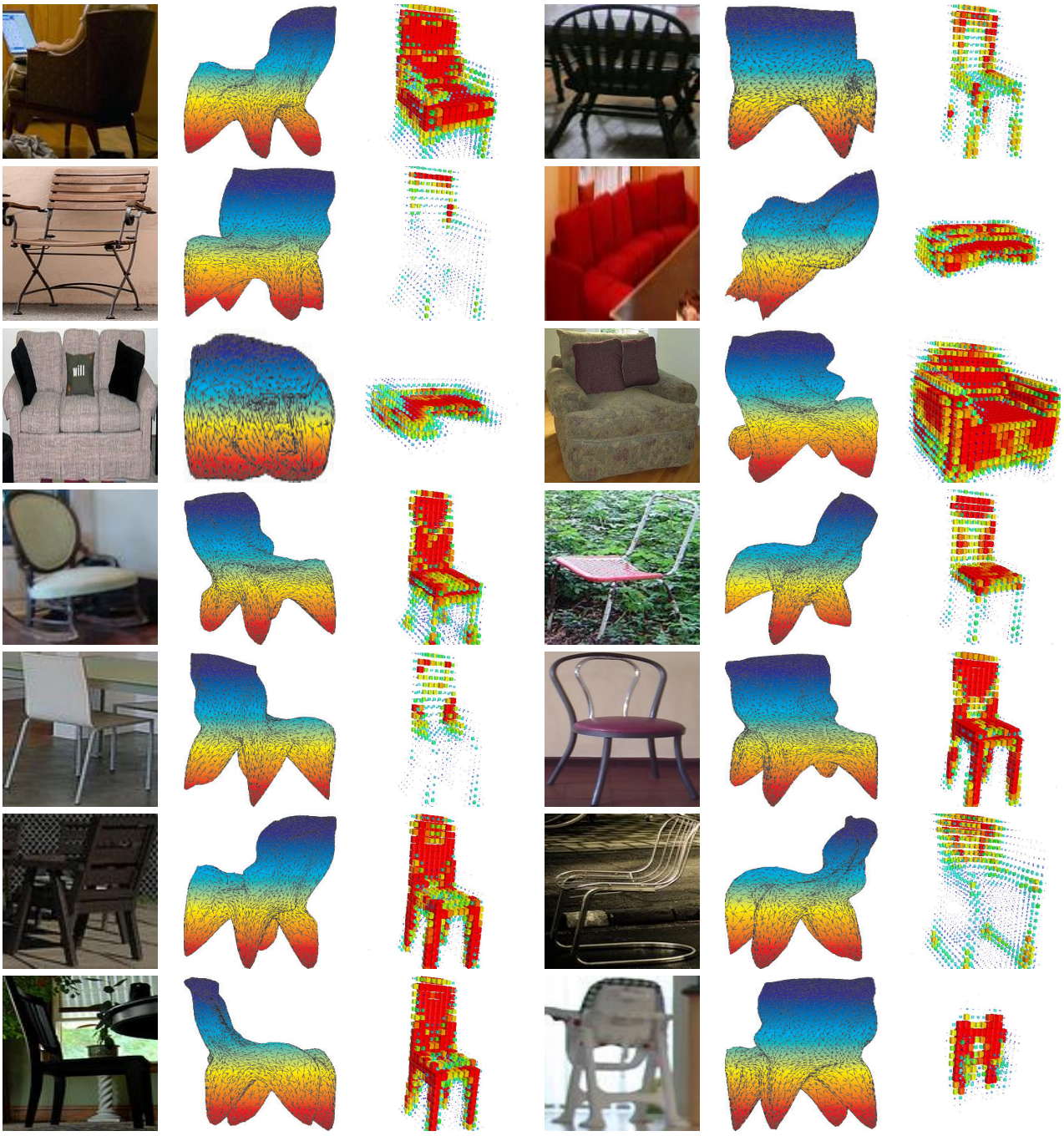
To compare, we first convert the prediction and ground truth obtained from the provided code by Kar *et al.* into an OBJ, by writing out the points and faces. Before computing the OBJ, we align the output points from their method with the ground truth as done in 'evaluation/evalMeshes.m' script provided by authors. We then voxelize the output to make it comparable with our approach.

4.2 Qualitative Results

Reconstructions on **randomly picked** Chair/Sofa images from PASCAL 3D+ using Kar *et al.* and our method. Complete qualitative results (on all 254 chair/sofa images) will be available on the project website.







5 Comparison with Li *et al.* [2] (Image based 3D Model Nearest-neighbor search)

We now report a comparison with [2] on their 315 image, 105 model labeled evaluation set. [2]'s method is an approach that is specific to nearest-neighbor model retrieval and has a number of advantages over our approach. Their features are hand-crafted and extracted directly from the underlying 3D model, as opposed to learned automatically from a coarse 20^3 voxel grid. Additionally, their method is designed to discriminate between similar models whereas our method's objective is trained purely for reconstruction. These two factors give Li et al.'s method substantial advantages in picking up on fine-grained details that would distinguish two chairs, for instance. As an added benefit, their approach is also class-specific. Despite these disadvantages, our method obtains strong performance. At recall@10, we get 82% compared to $\approx 95\%$ for them (obtained from Figure 8 in their paper).

6 More embedding space analysis

6.1 More interpolation Results

Randomly picked results for interpolation between two randomly picked models.

References

[1] Abhishek Kar, Shubham Tulsiani, João Carreira, and Jitendra Malik. Category-specific object reconstruction from a single image. In *CVPR*, 2015.

[2] Y. Li, H. Su, C. R. Qi, N. Fish, D. Cohen-Or, and L. J. Guibas. Joint embeddings of shapes and images via cnn image purification. *ACM TOG*, 2015.