Supplementary Material

Fangyin Wei[1]   Elena Sizikova[2]   Avneesh Sud[3]   Szymon Rusinkiewicz[1]   Thomas Funkhouser[1,3]
[1] Princeton University, NJ, USA   [2] New York University, NY, USA   [3] Google Research, CA, USA

In this Supplementary Material, we provide additional results for the experiments on piecewise anisotropic scaling and non-rigid deformation on the chair, airplane, and human body datasets (described in the main manuscript). Sec. 1 and 2 demonstrate qualitative semantic parameter editing results on rigid and non-rigid shape categories, respectively, while Sec. 3 shows results on out-of-distribution examples. Sec. 4 provides additional comparisons to Neural Cages [4], a recent technique for source-to-target matching with a learned deformation, and DualSDF [2], a recent deformation method with learned re-synthesis.

1. Piecewise Anisotropic Scaling Results

A common editing strategy for objects with well-defined semantic parts is anisotropic scaling of different or independent parts. We show anisotropic editing results produced by the proposed method for chairs in Fig. 1 and Fig. 2, and for airplanes in Fig. 3. All shown examples are testing examples, i.e. held out during the training stage. For each class, we show an input shape (yellow) and the estimated synthetic template (blue). Note that the latter matches the realistic shape in all structural components relevant to editing, such as the seat of the chair or wing length of the airplane. Therefore, the proposed model correctly infers semantic parameters of realistic shapes when trained with only synthetic parameter annotations. In gray, we show the deformation result of one semantic parameter per column, with the first row decreasing and the second row increasing that parameter. As hoped, modifying one semantic parameter does not affect other parameters, i.e. the shape of other parts. For example, modifying the depth of a chair seat does not change the shape of its legs. At the same time, parameters correctly update groups of similar parts. For example, modifying the "leg height" parameter for a chair updates all four legs, while modifying "wing length" updates both airplane wings. Note that details are well-preserved after deformation (*e.g.*, patterns on the chair back and seat, curved chair legs, engines and landing gear of the airplane). The proposed method allows for substantial parameter change (*e.g.*, height of chair back and legs, width of chair seat, length of airplane fuselage), in contrast to methods that train on shape databases supervised with ground truth deformed

shapes that might lack such extreme deformations.

One advantage of the proposed method is to support continuous deformation. We show such results for the chair in main manuscript Fig. 4. Please find this in chair.html contained in SupplementaryMaterial.zip.

2. Non-Rigid Deformation Results

In Fig. 4 and Fig. 5, we show results for editing semantic pose parameters of human bodies on DFAUST [1] and Buff [5] datasets, respectively. Both source inputs are characters held out during the training stage. Our system is able to apply the deformations defined by a simple, parameterized skinned model (SMPL [3]) to realistic human input shapes, accommodating both large (hip movement) and subtle (ankle movement) motions. Note that in the main manuscript, we demonstrate only one degree of freedom per joint because of space constraints. In Fig. 4, we provide full deformation results for all three degrees of freedom of each joint presented in the main manuscript. We can see that the proposed method accepts and recognizes the parameters of shapes in various poses. Our model also correctly decouples the effects of all parameters, even those affecting the same joint. In Fig. 5, we deform an arbitrary set of shape and pose parameters for the same input shape (yellow). Note that this input shape is a scanned clothed male and is noisy, not watertight, and with many isolated faces around the body.

We also provide continuous deformation results for shapes in main manuscript Fig. 5. Please find this in humanbody.html contained in SupplementaryMaterial.zip.

3. Out-of-Distribution Shapes

To evaluate the generalization ability of the proposed method, we test it on shapes falling outside of the shape distribution of the training dataset. Fig. 6 and Fig. 7 show examples from all three classes, illustrating cases in which the test shapes are topologically different from the training examples or are missing some of the components present in the synthetic shapes. For example, in Fig. 6, the first chair (top row) and the toilet in the second row have only one leg each, the chair in the bottom row has arms and pillows. The first airplane has horizontal stabilizers on the top, the third airplane has three vertical stabilizers, the second

and last airplanes have a delta wing, the third and fourth airplanes have straight wings and propellers on the wings, while in Fig. 7 the human has long hair and an initial pose with the hands almost close together. Note that each joint has up to three degrees of freedom and can be rotated in two directions for each degree of freedom, therefore there are many deformation results under the same parameter name in Fig. 7. In all cases, the proposed method produces semantically meaningful results and preserves input shape details.

4. Comparison to Prior Work

To compare against Neural Cages [4], which was designed for the task of source-to-target deformation rather than direct manipulation of semantic parameters, we modify the proposed method to use a deformation determined by synthetic estimates from both the source and target shapes. Fig. 8 and Fig. 9 show examples using four different source chairs and 22 target chairs. Fig. 10 shows all three human body shapes evaluated in and with templates provided by the original Neural Cages work. All three human body examples are not seen during the training of the proposed method. Both methods are able to match the target shape globally. However, note the spurious deformation in the legs in the Neural Cages results as observed in the main manuscript, for example, the 6th, 10th, and last deformed shapes for the first source chair, as well as the fact that in most examples in Fig. 8 and Fig. 9, the Neural Cages result matches the seat thickness of the source, not the target. In contrast, the proposed method does not introduce unnecessary deformations, yet is able to deform the source locally, not just globally, to match the seat thickness. In the human body results in Fig. 10, Neural Cages can introduce distortions in regions with large deformation, such as the knees in example (1a), (1c), (2b) and (3d) or disproportionately scaled hands and feet in examples (1d) (2a) (3a) (3b) and (3c). We observe that, compared to deformation methods based on coarse geometric handles, our results support more granular manipulation even in extreme poses.

We also compare the proposed method with DualSDF [2] on the task of part manipulation. We consider examples common to DualSDF's training set and our testing set. To generate results for DualSDF, we first obtain the embedding of an input shape and then manually adjust the radius of a primitive on the corresponding part along the corresponding dimension. Fig. 11 shows the results for chairs and airplanes, the two classes shared by DualSDF and our method. The DualSDF reconstruction, which requires re-synthesis, loses details of the original shape *e.g.*, the poles on the back and the joints on the legs for the chair, the landing gear and flap track fairings on the back of the wings for the airplane. In addition, manipulating a single primitive in one part will cause deformation in other parts as well, since the latent embeddings from DualSDF are highly entangled. For ex-

ample, when manipulating primitives on the fuselage (the first two airplane examples), the wings are also changed by DualSDF, but correctly preserved by the proposed method.

References

[1] F. Bogo, J. Romero, G. Pons-Moll, and M. J. Black. Dynamic FAUST: Registering human bodies in motion. In *CVPR*, 2017. 1, 6

[2] Z. Hao, H. Averbuch-Elor, N. Snavely, and S. Belongie. DualSDF: Semantic shape manipulation using a two-level representation. *arXiv preprint arXiv:2004.02869*, 2020. 1, 2, 13

[3] M. Loper, N. Mahmood, J. Romero, G. Pons-Moll, and M. J. Black. SMPL: A skinned multi-person linear model. *ACM Trans. Graphics (Proc. SIGGRAPH Asia)*, 2015. 1

[4] W. Yifan, N. Aigerman, V. Kim, S. Chaudhuri, and O. Sorkine-Hornung. Neural cages for detail-preserving 3D deformations. *arXiv preprint arXiv:1912.06395*, 2019. 1, 2, 10, 11, 12

[5] C. Zhang, S. Pujades, M. J. Black, and G. Pons-Moll. Detailed, accurate, human shape estimation from clothed 3D scan sequences. In *CVPR*, 2017. 1, 7
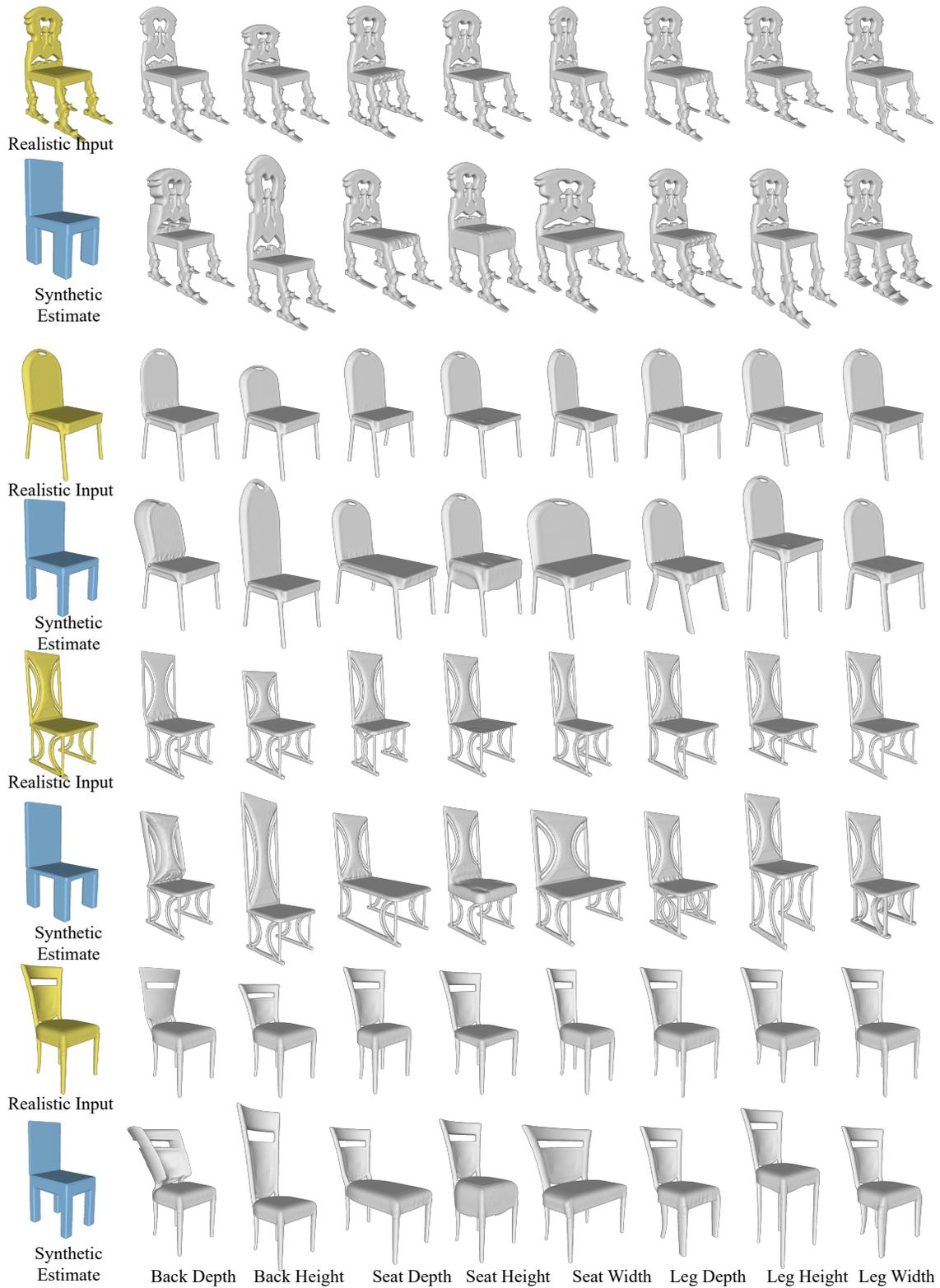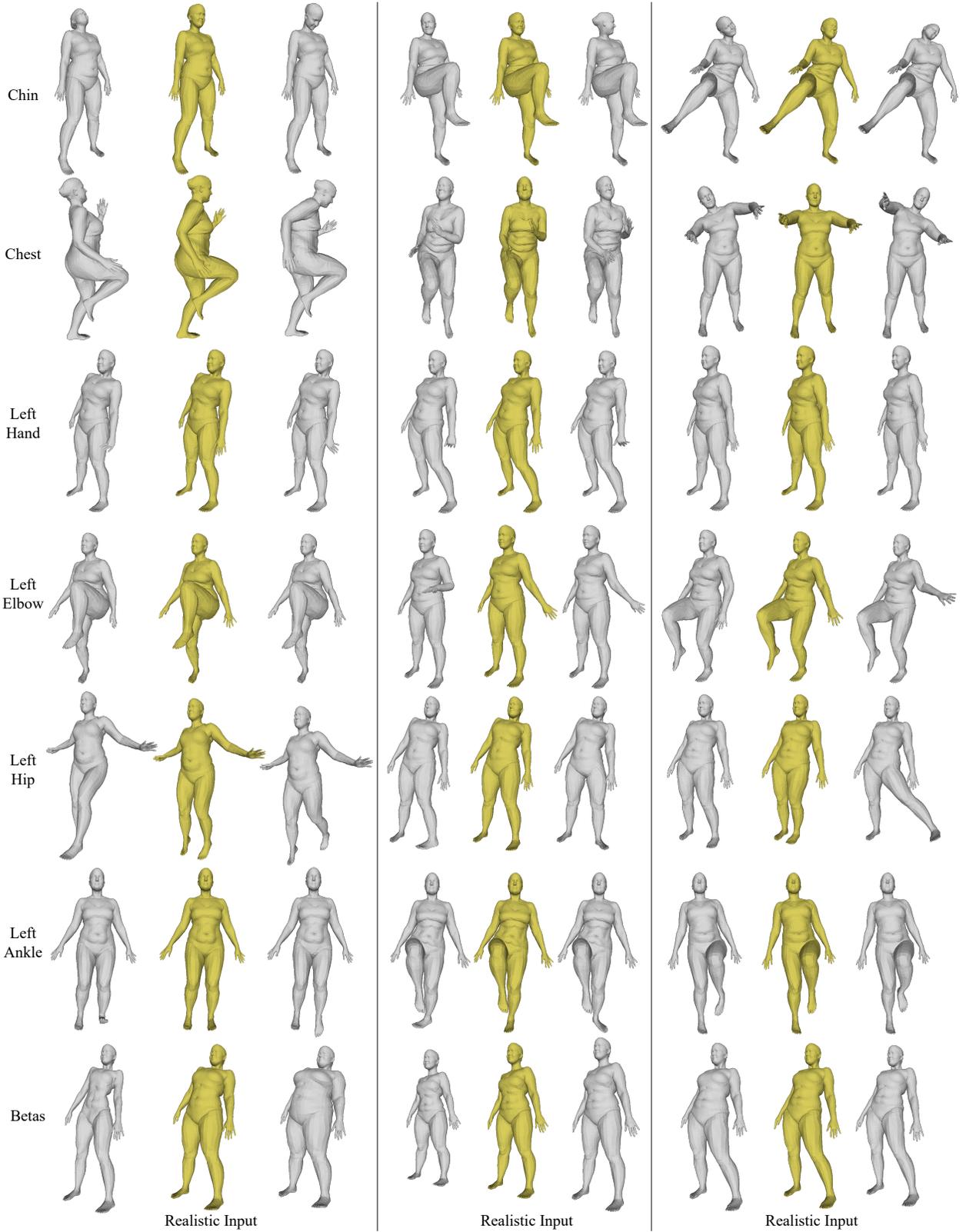
Figure 1: Editing results for piecewise anisotropic scaling on chairs. Realistic input shapes (yellow) are fit to synthetic templates (blue), then edited by decreasing (top row) and increasing (bottom row) each shape parameter. The proposed method preserves details and independently controls each semantic parameter.
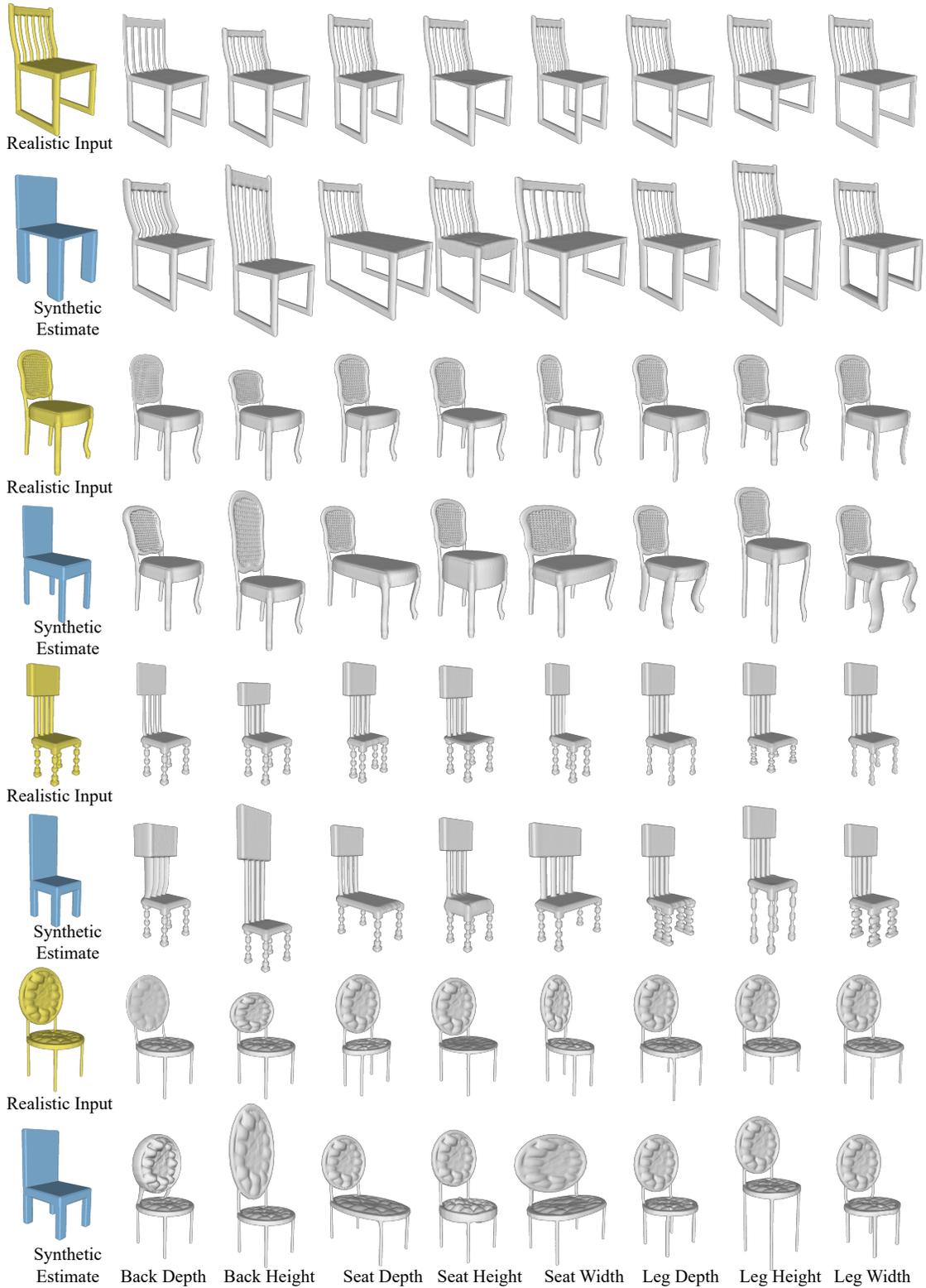
Figure 2: Additional editing results with piecewise anisotropic scaling on chairs. Realistic input shapes (yellow) are fit to synthetic templates (blue), then edited by decreasing (top row) and increasing (bottom row) each shape parameter. The proposed method preserves details and independently controls each semantic parameter.
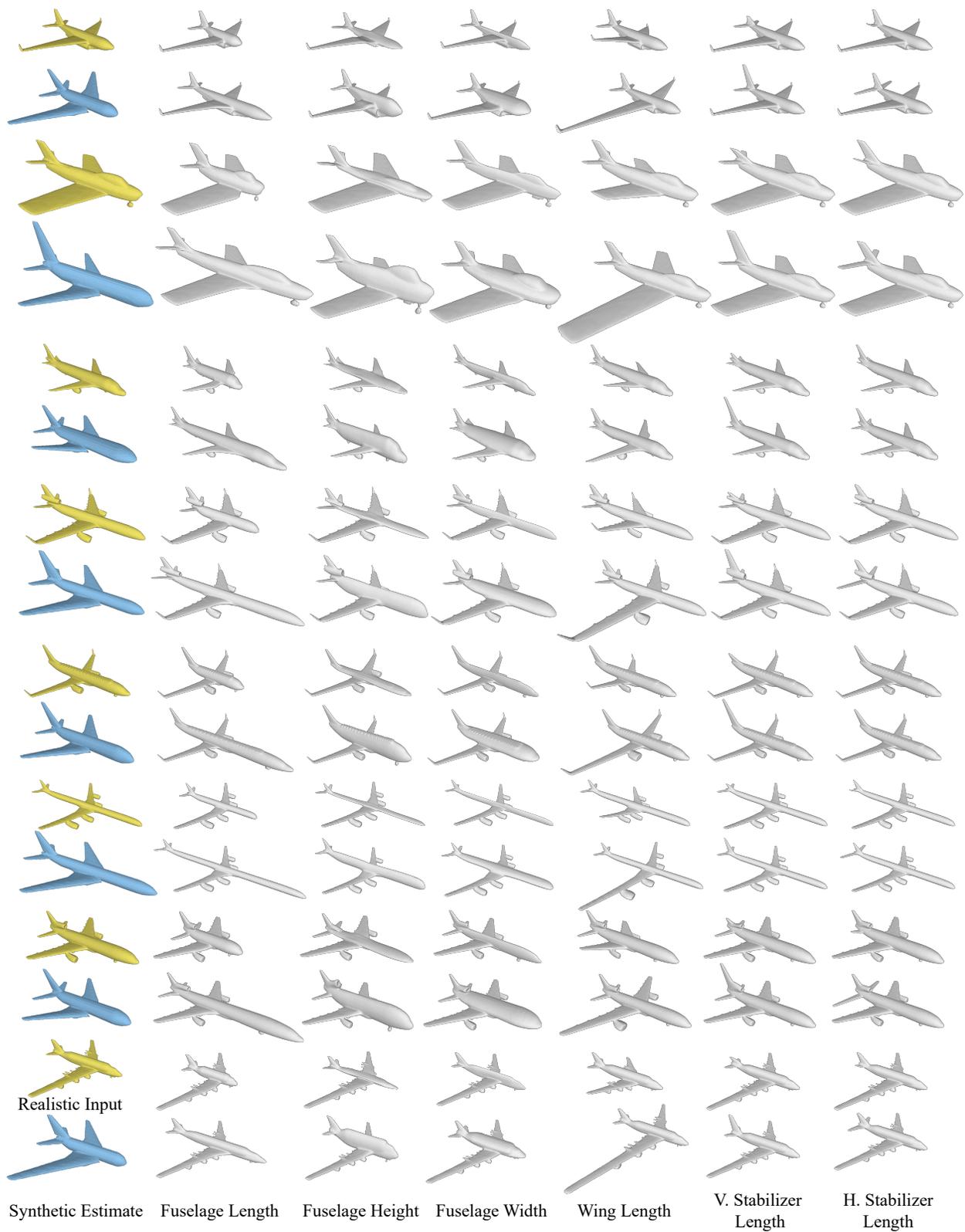
Figure 3: Editing results with piecewise anisotropic scaling on airplanes. Realistic input shapes (yellow) are fit to synthetic templates (blue), then edited by decreasing (top row) and increasing (bottom row) each shape parameter in turn. The proposed method preserves details and independently controls each semantic parameter.

| Synthetic Estimate | Fuselage Length | Fuselage Height | Fuselage Width | Wing Length | V. Stabilizer Length | H. Stabilizer Length |

Figure 4: Editing results for deformation on human bodies from the DFAUST [1] dataset. In each set of three shapes, an input human body (yellow) is deformed by changing one semantic parameter in both directions. The first six rows show results of deforming three degrees of freedom of all joints presented in the main manuscript Fig. 5, with one joint in each row. The last row shows deformation results of the three shape parameters (betas).
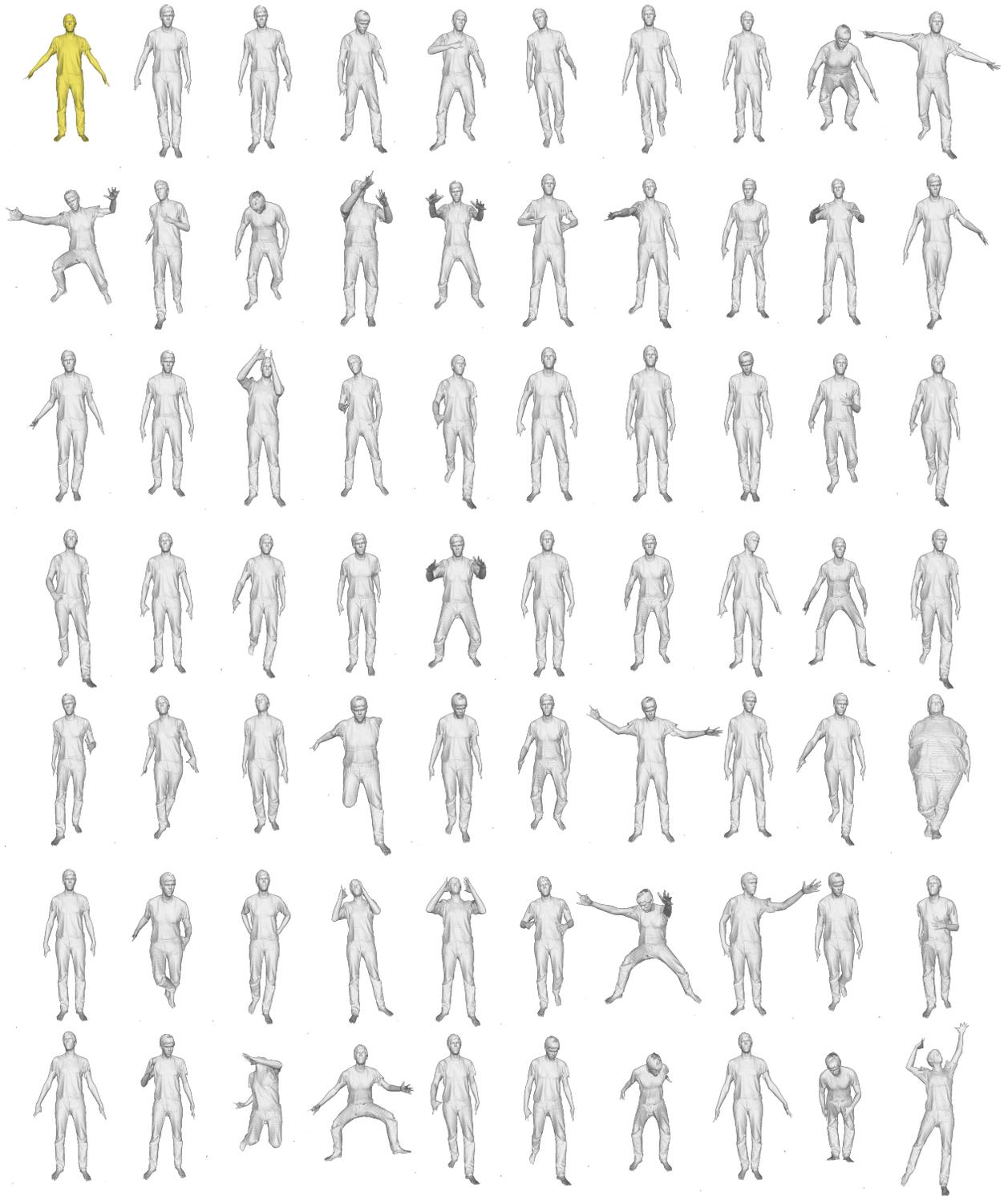
Figure 5: Editing results for deformation of human bodies from the Buff [5] dataset. Given an input testing shape, we are able to arbitrarily manipulate any set of semantic parameters and output the deformed shapes.
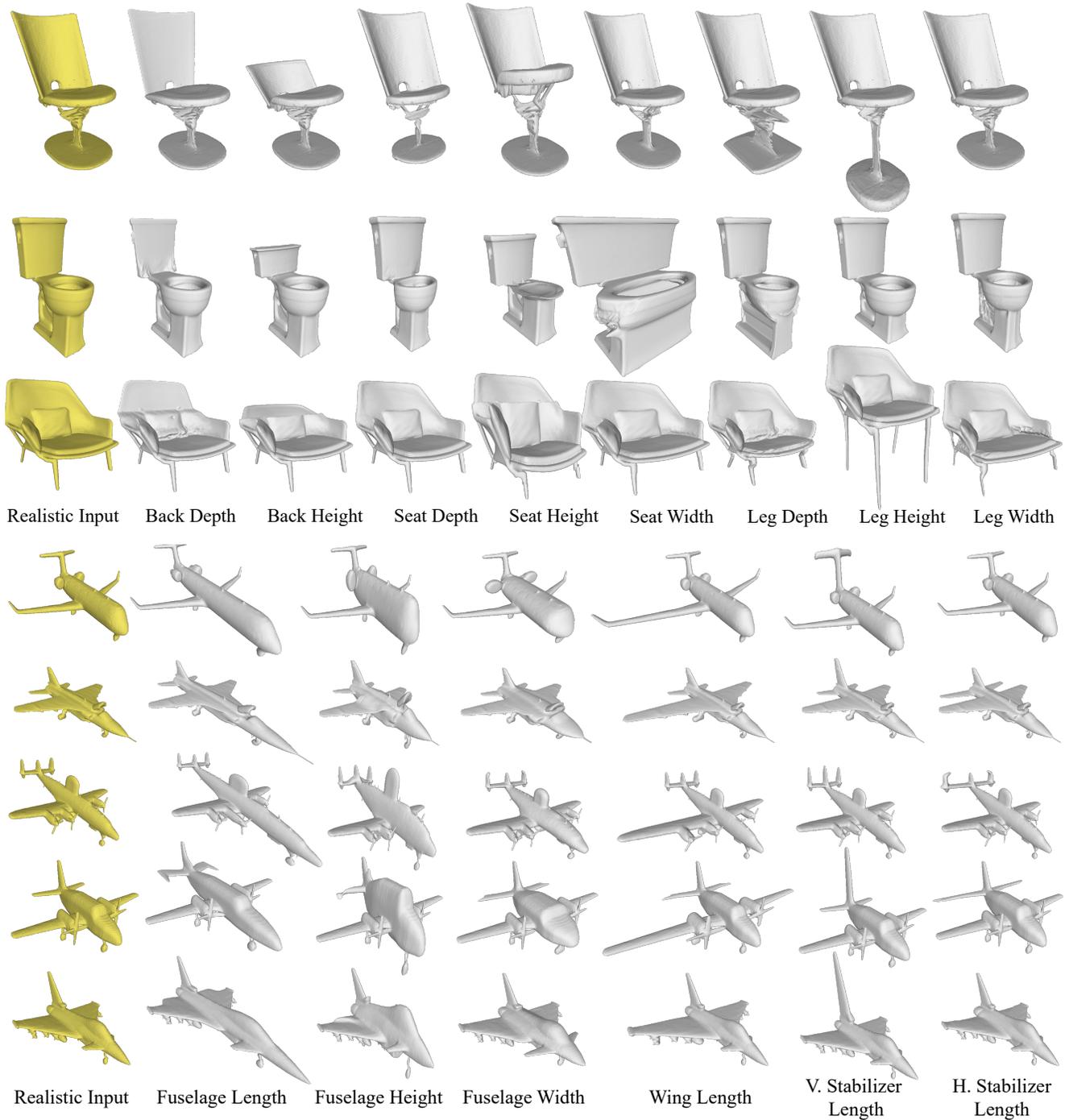
Figure 6: Editing results for out-of-distribution shapes on chairs and airplanes. Our method produces semantically meaningful results, and correctly preserves detail, even if the input shape has missing parts or different topology relative to the synthetic template.
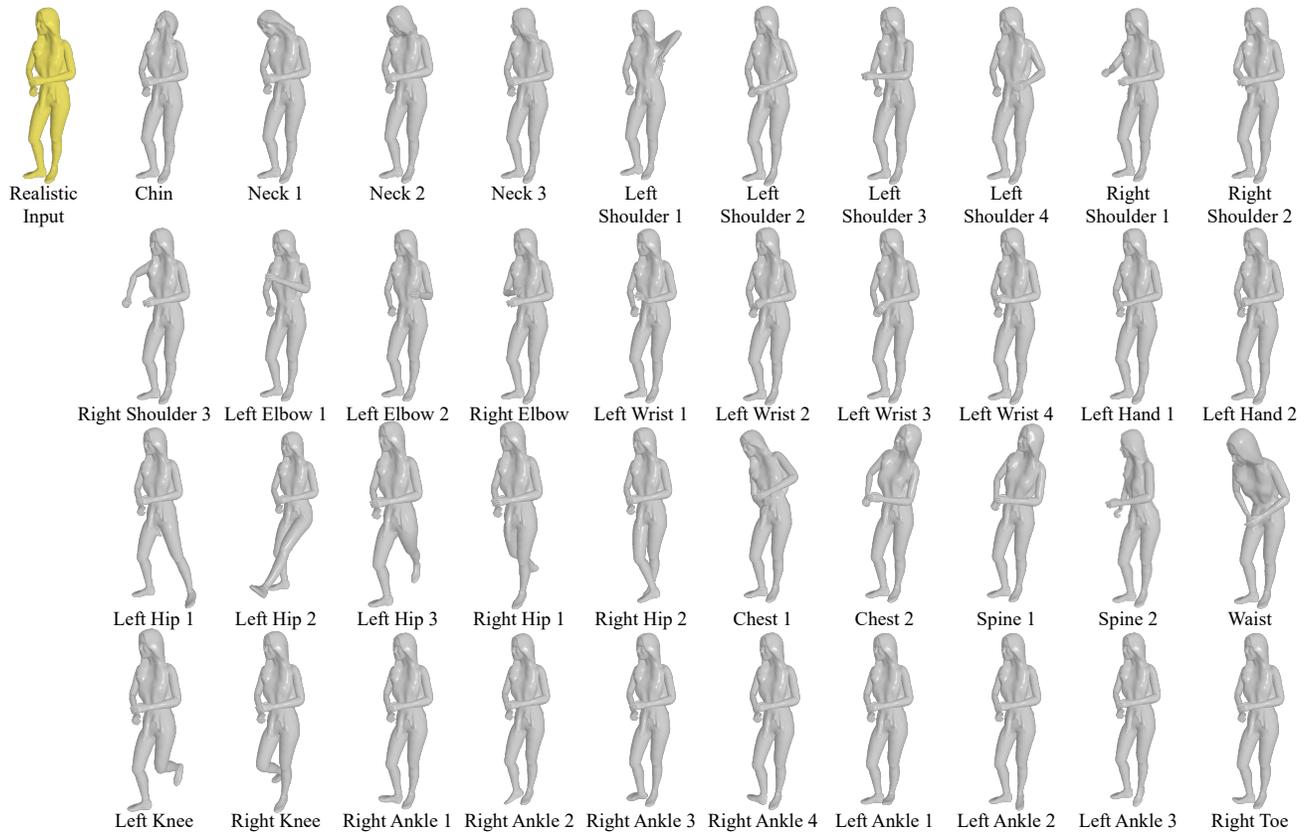
Figure 7: Editing results for out-of-distribution shapes on human bodies. Our method produces semantically meaningful results, and correctly preserves detail, even if the input shape has missing parts or topology different from the synthetic template.
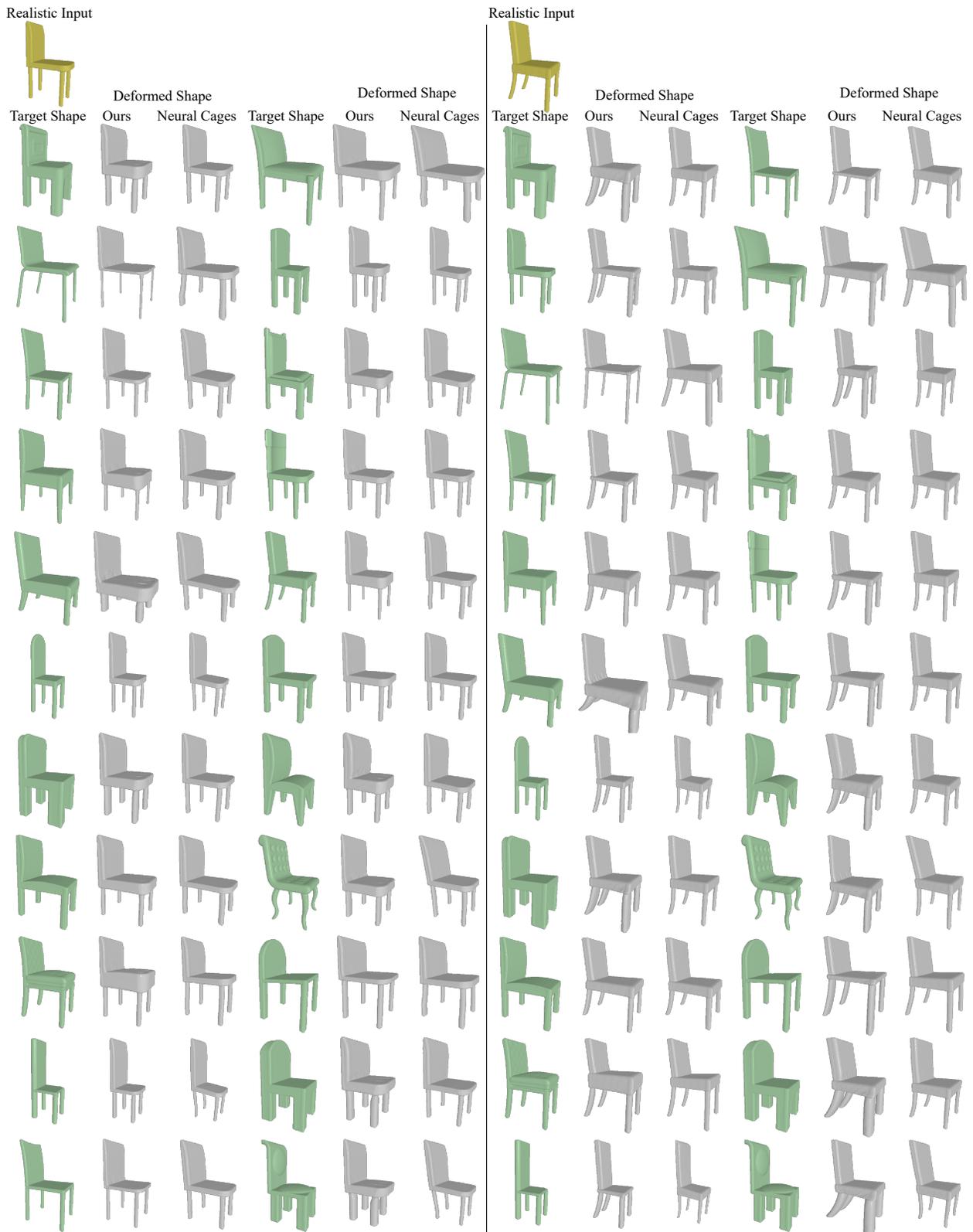
Figure 8: Comparison to Neural Cages [4] (chairs). Each source shape (yellow) is deformed to match 22 target shapes (green), using both the proposed method and Neural Cages. Both methods are able to globally match the target shape, but Neural Cages often exhibits distortion in regions of large deformation and cannot match semantic parameters if local deformation is required. In contrast, our method support more accurate and granular manipulation.
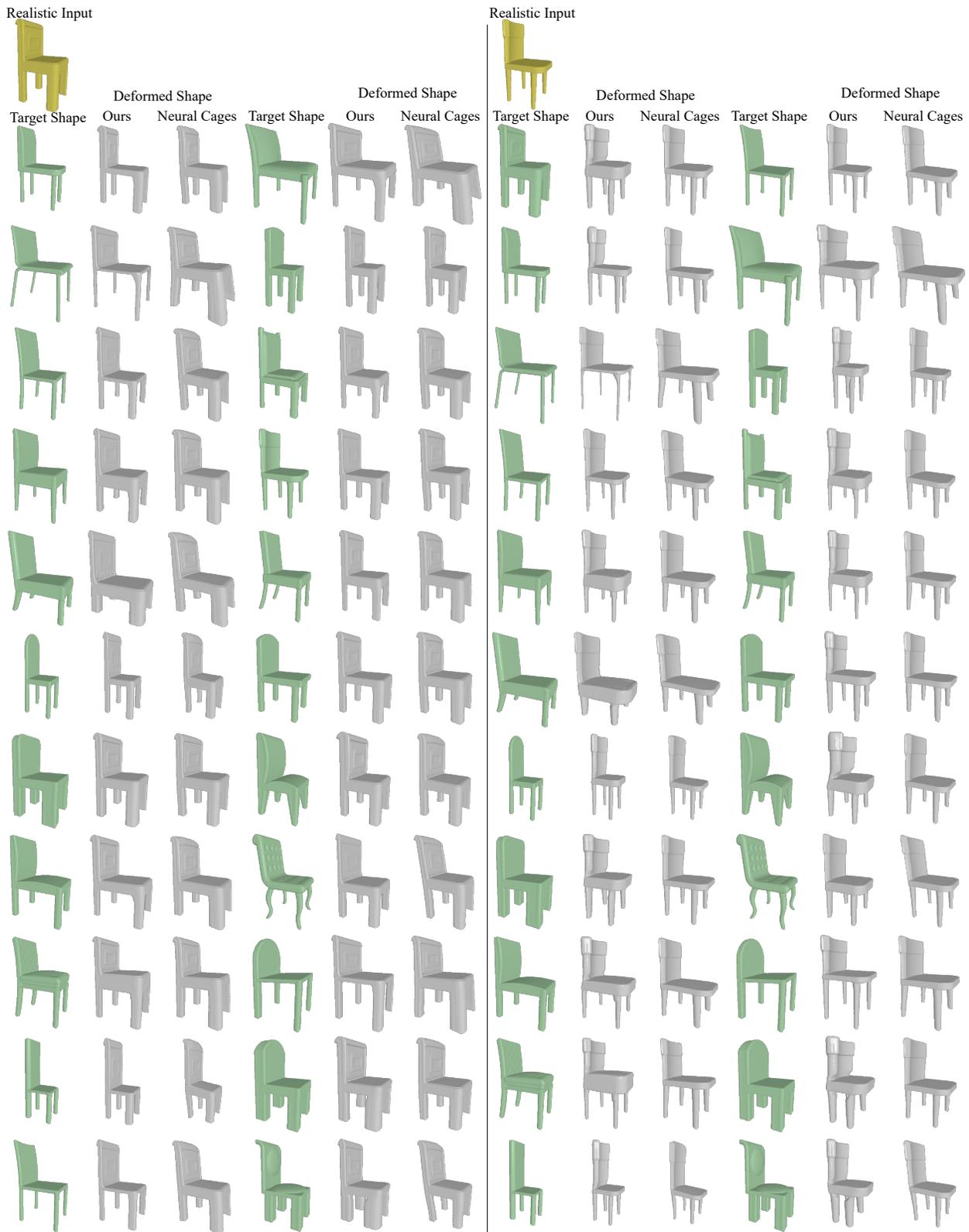
Figure 9: Additional comparison to Neural Cages [4] on chairs. Each source shape (yellow) is deformed to match 22 target shapes (green), using both the proposed method and Neural Cages. Both methods are able to globally match the target shape, but Neural Cages often exhibits distortion in regions of large deformation and cannot match semantic parameters if local deformation is required. In contrast, our method supports more accurate and granular manipulation.
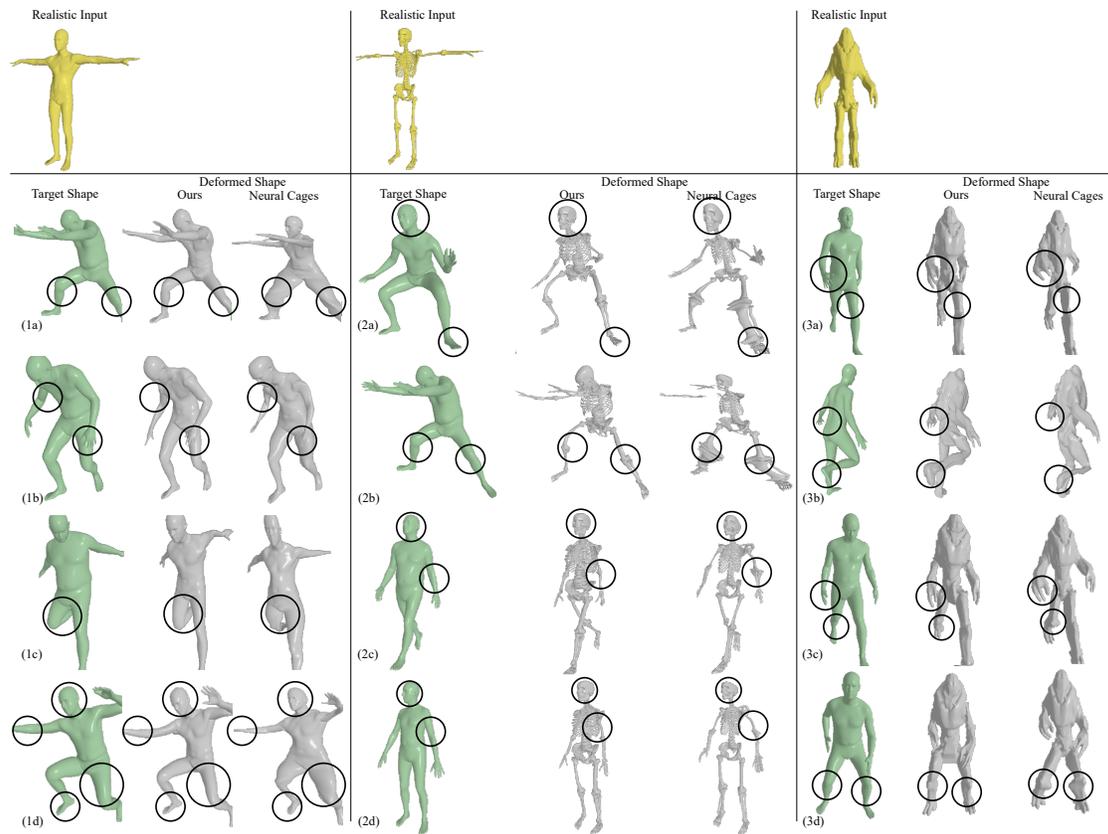
Figure 10: Additional Comparison to Neural Cages [4] on human bodies. Each source shape (yellow) is deformed to match four target shapes (green), using both the proposed method and Neural Cages. Both methods are able to globally match the target shape, but Neural Cages often exhibits distortion in regions of large deformation and cannot match semantic parameters if local deformation is required. In contrast, our method support more accurate and granular manipulation, even for extreme poses.
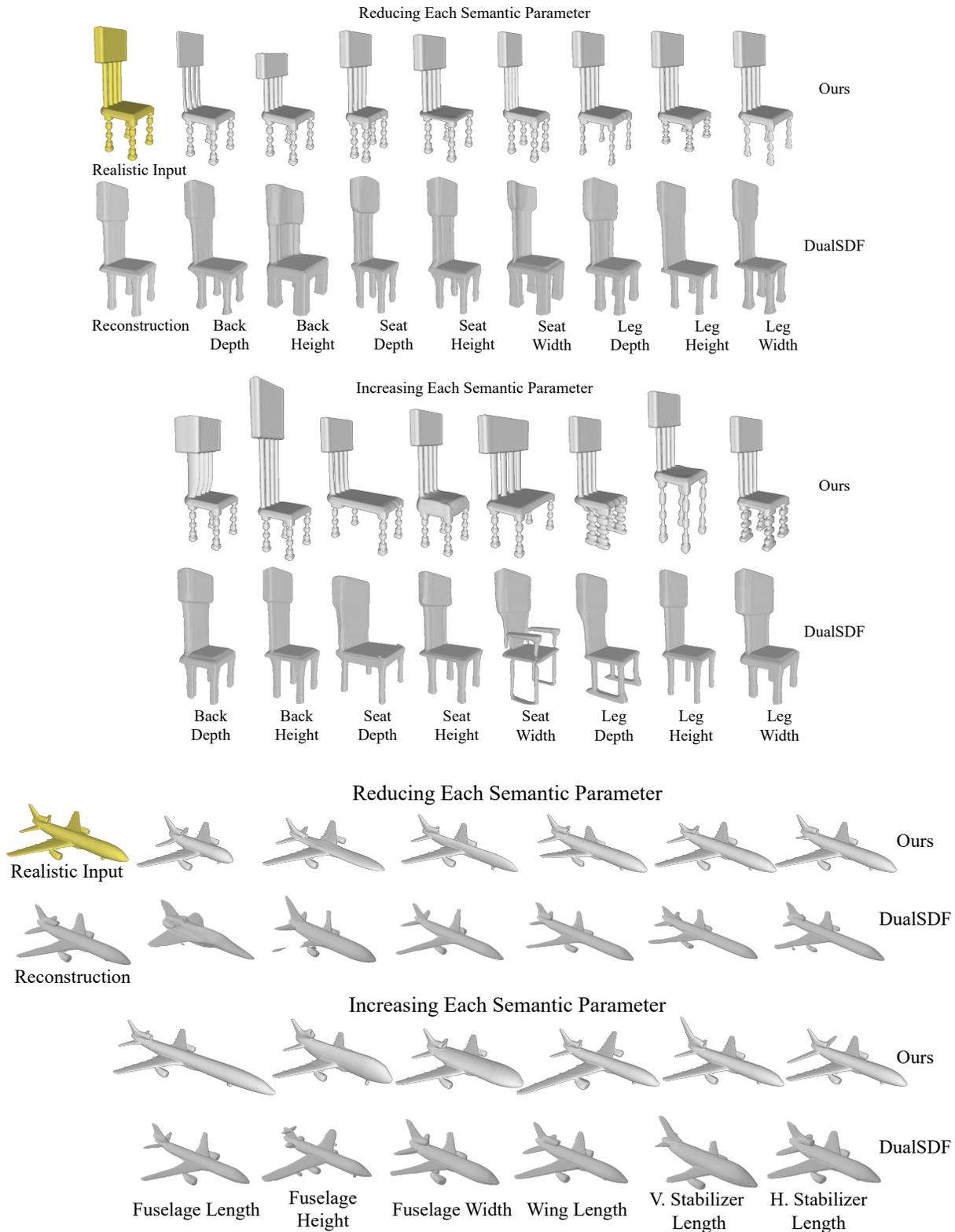
Figure 11: Comparison to DualSDF [2]. The input (yellow) is edited by changing a semantic parameter in our system, or adjusting the radius of a primitive in DualSDF. When changing a local parameter with DualSDF, the global shape is also affected, as seen in the arms that incorrectly show up in the result of "increasing seat width". Also, in contrast to the proposed method, the DualSDF results do not preserve details such as the poles on the chair back and the landing gear and flap track fairings on the back of the wings for the airplane.