

Deep Feature Deformation Weights

Supplementary Material

A. Ablations

Different Image Encoders. We show in Fig. 11 additional deformation results using DFD weights computed using other image features, *with no additional regularization or anchor points*. We find that deformation results are consistent and robust across all the image models we tested, though DINO and Diff3F give generally the best results.

Our weights also offer a unique perspective into image model interpretability. We visualize the DFD weights from different image models over the same shape in Fig. 12. We see that all 2D foundation models we tested converge to the same common global semantic understanding of shapes. We can also identify nuanced differences in image model behavior that coincides with prior reported observations. For example, CLIP-ViT [30] tends to focus much more on global understanding and less on local part relationships, whereas SAM2 [32] tends to better isolate local features. In practice, we find that DINO [27] and Diff3F [8] find the best balance between local and global shape understanding.

Barycentric Feature Distillation Ablation. We ablate on barycentric feature distillation on high resolution shapes from the Stanford 3D scanning repository. We take each shape, simplify them using QEM decimation, and distill features into our feature field using either vertex distillation (the method used by prior works) or barycentric feature distillation. Vertex distillation takes only the features at render pixels which contain a vertex, whereas barycentric distillation makes use of every pixel which contains a point on the 3D surface. To ensure a fair comparison, we optimize the vertex distillation feature field for an equivalent # FLOPs/pixel samples as our method. Fig. 13 shows that without the dense field sampling offered by barycentric feature distillation, features distilled from coarse shape renders are unable to interpolate well to shapes at their original resolution. The deformations produced by the vertex distillation are neither smooth nor visually-meaningful.

B. Rebinding Comparison to OptCtrlPoints

We quantitatively compare rebinding times against OptCtrlPoints [22], a recent method for efficient rebinding through an updated solve to the biharmonic coordinates optimization problem. We take 20 random shapes from our dataset, precompute the OptCtrlPoints factorization, and then randomly sample sets of 1, 10, and 100 control handles. We report results in Sec. A. Observe that for all control point set sizes, our method is still $\sim 1000\times$ faster than OptCtrlPoints, demonstrating that bind time optimization is still a significant bottleneck.



Figure 11. **Qualitative Results with Other Image Models.** We show deformation results using DFD weights computed using image features from other pretrained image models.

C. Additional Comparisons to Baselines

APAP-Bench 3D. We show in Fig. 14 a full comparison of the baselines against DFD for all the deformations shown in the APAP paper [44]. We also show in Fig. 15 the full

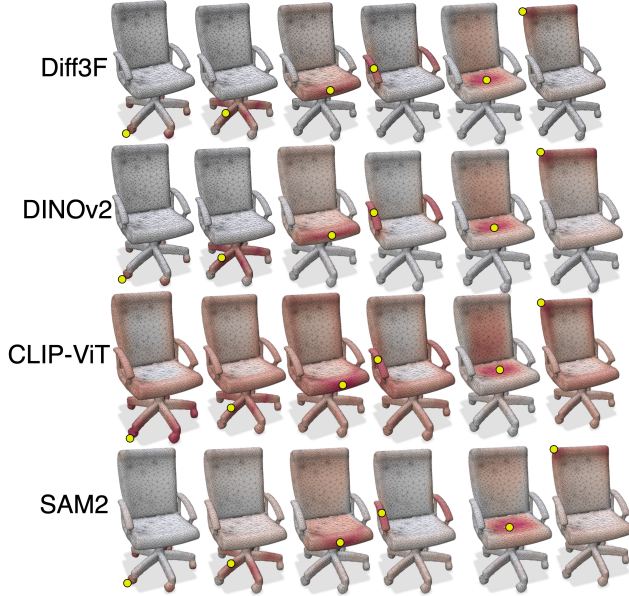


Figure 12. **Different Encoders.** By visualizing the DFD weights, we observe that pre-trained 2D foundation models contain approximately similar understanding of shape features and part relationships.

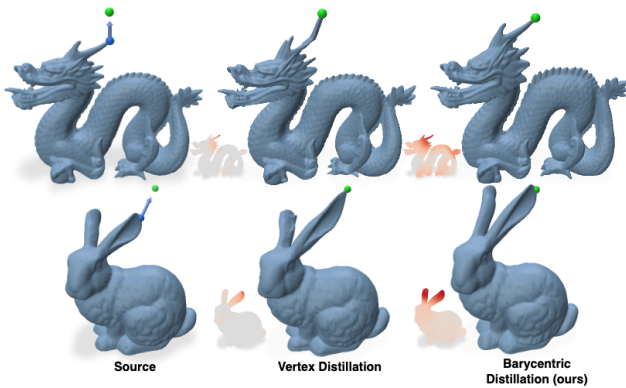


Figure 13. **Barycentric distillation ablation.** We distill DFD weights supervising only on vertex features on the decimated mesh and train for the same number of FLOPs as with barycentric distillation. The resulting deformations on the high resolution shapes are neither smooth nor visually-meaningful.

comparison without using the 0.01-ball sampling trick to increase the number of handle and fixed point constraints. As reported by APAP, all baselines other than APAP completely fail when dealing with a limited number of constraints. NeuralMLS and biharmonic coordinates degenerate, while ARAP produces global translations of the shape. APAP also experiences worse artifacts, whereas our method is completely stable with a single handle deformation.

DeepMetaHandles. We show a larger set of qualitative comparisons on shapes from the DeepMetaHandles dataset

Method / # Control Points	1	10	100
DFD (Ours)	0.004	0.012	0.034
Biharmonic (OCP)	4.6	11.35	21.8

Table 3. **Rebind Time (s).** All existing methods require solving an optimization problem for every new set of control points, which is expensive (Sec. 2.3). OptCtrlPoints (OCP) [22] is a recent method which aims to make the re-solve more efficient. To compare rebinding speeds, we take 20 random shapes from our dataset, precompute the OCP factorization, and randomly sample sets of 1, 10, and 100 control points 1000 times. Average time to rebinding is reported in seconds. OCP is still limited by the optimization solve, and is $>1000\times$ slower than our method.

in Fig. 16. As observed in the main paper, DFD weights produce deformations which are consistently smoother and more symmetry/part preserving than the baseline methods. DMH is the strongest baseline here because the model is both trained on this data and used to predict the handle deformations. Despite this, it is still inconsistent in predicting smooth and visual-aware deformations (e.g. rows 2-4, 8,9,11,13).

D. Topology Robustness

DFD weights are extremely robust to topological defects, thanks to the visual nature of the feature field supervision. We demonstrate this by showing pose changes on a problematic model in Fig. 17. The example shown has 3,804 boundary edges, 70 disconnected components, and 12 non-manifold edges, but our weights are able to still smoothly interpolate deformations and produce plausible pose changes.

E. High Resolution Distillation

We visualize distillation timings and deformations on very high resolution shapes from the Stanford 3D Scanning Repository in Fig. 18. Because these shapes are generated from scans, they contain topological defects, which are reported under each shape (#NM reports number of non-manifold elements and #H is the number of holes). We furthermore report the QEM decimation and rendering time (**R**), the feature field distillation time (**T**), and the pose/inference time (**I**). We emphasize that the feature field distillation itself is completely agnostic to mesh resolution, which is why **T** is largely constant across all shapes. Pose time **I** does scale with shape resolution, especially when memory limits require batching of the feedforward pass (Lucy model), but it is still very fast. The majority of the bottleneck is limited to **R**, which scales robustly with mesh resolution thanks to the efficiency of QEM relative to rendering. Even at the highest resolution, the entire distillation process end-to-end is under a minute.

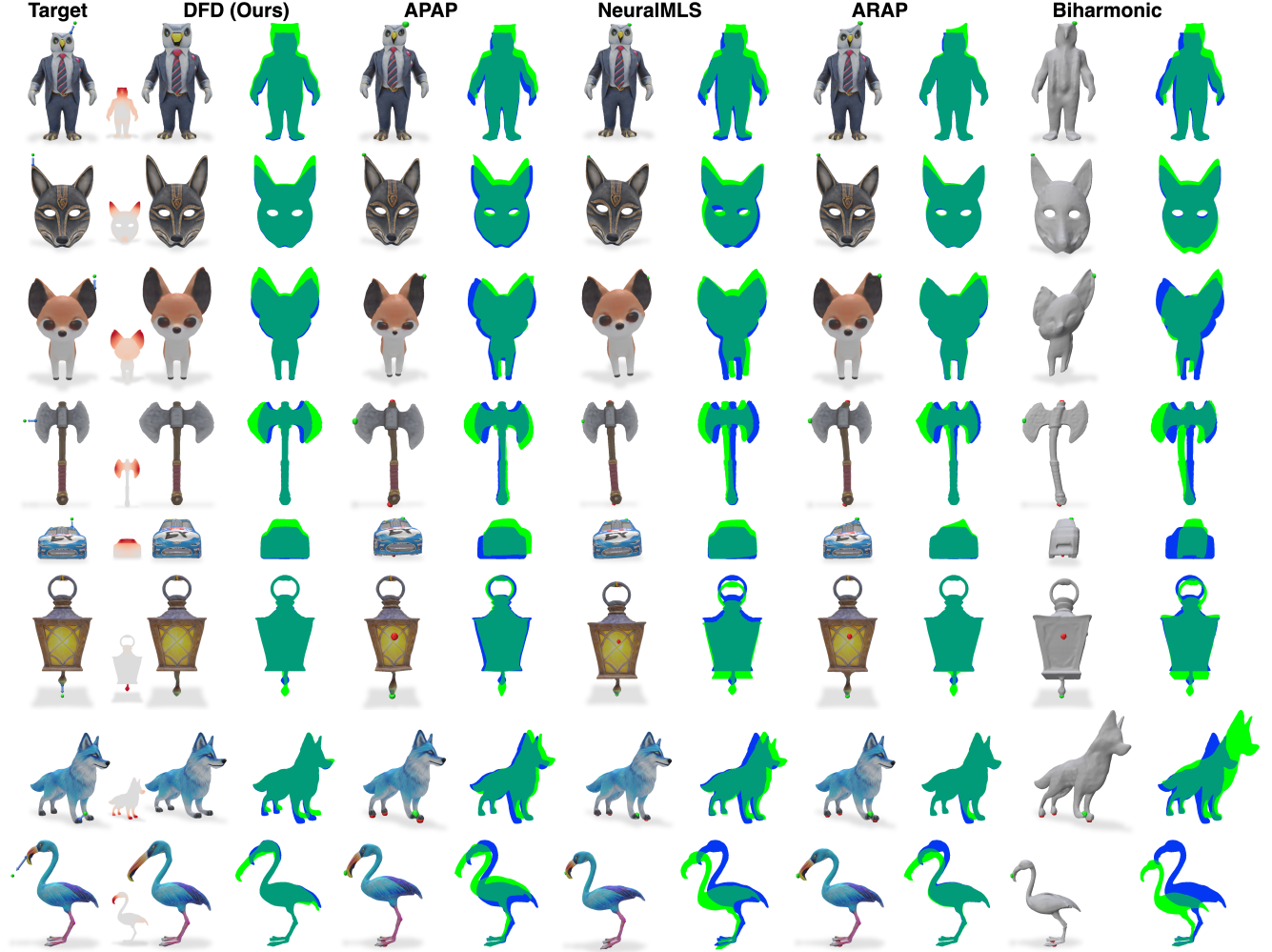


Figure 14. **Full Qualitative Comparison with APAP Results.** We compare against the full array of deformation results shown in the APAP [44] paper. Similar to the results in the main paper, in all examples our method produces visual and symmetry-aware deformations, whereas baselines produce undesirable global rigid transformations and general asymmetries.

F. Weight Generalizability

Novel Shape Instances/Remeshings. Thanks to barycentric feature distillation, our weights generalize well from coarse shapes to higher-resolution remeshings. Furthermore, our distilled feature field can even be used to deform novel shapes within the same shape class, as shown in Fig. 19. The neural field representation allows for any point in the ambient space to receive a visual feature, and thanks to the smooth nature of the field, novel shapes with similar visual parts in similar spatial regions can share the same field. Reusing these weights allows for similar visual-aware deformations, such as the co-deformation of the cow legs.

Consistent Shape Understanding. In Fig. 20, we show that even across widely varying geometries, shapes within the same class (e.g. chairs), will induce DFD weights which

identify similar visual relationships, such as the legs, arm-rests, backrest, and seat of the chair models.

G. Dense Handle Results

We evaluate our method’s performance on dense handle configurations with the same handles used by the baselines in Fig. 9. Observe that even though partition of unity is no longer guaranteed under dense handles, our method’s results are still reasonable and symmetry preserving.

H. Surface Metrics

We report the same surface metrics from Implicit-ARAP [1] for two example deformations in Fig. 22. Though our method produces greater distortion along all four metrics, we observe that visually our results maintain greater realism

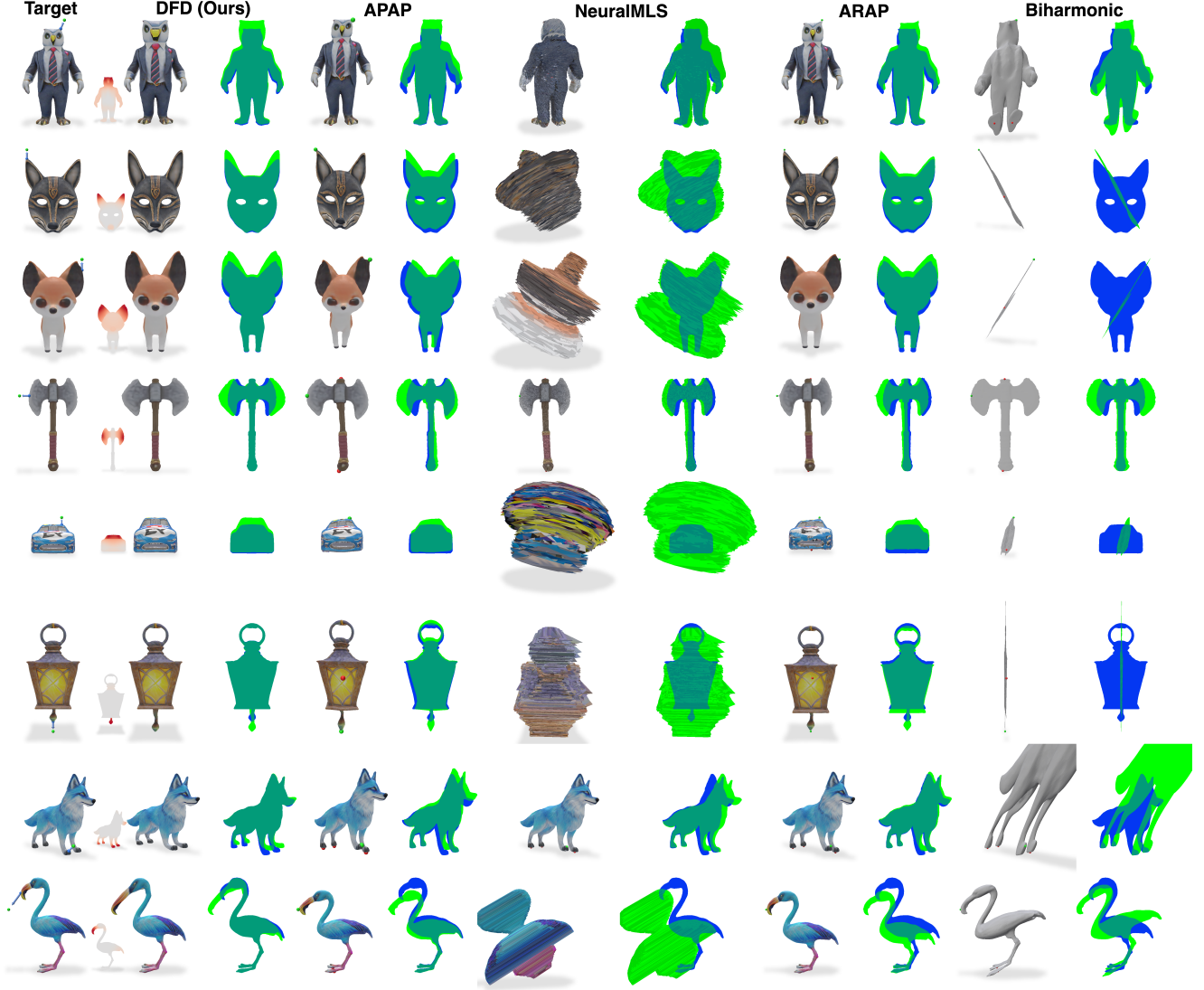


Figure 15. **Single-Handle APAP Comparison.** We compare against the baseline methods *without* adding the 0.01-radius neighbors of each prescribed fixed point, which is a smoothing trick employed by APAP. Note that without the smoothing trick, some baselines fail completely (neuralmls, biharmonic) whereas other methods experience slightly worse artifacts (APAP, ARAP). Our method does not use fixed point constraints and does not rely on such smoothing hacks.

based on the shape’s semantics and part being deformed.

I. Interactive GUI

We provide video examples in the `examples` folder of our interactive GUI demonstrating the semantic understanding and interactivity enabled by DFD weights.

J. User Study Screenshots.

We show screenshots from our user study in Fig. 23. We instruct users to select the top 2 deformation results for 15 different shape deformations from our datasets. We anonymize the 6 different methods (which include both the translation

and affine variants of our results) and randomly shuffle their presentation order. We collect responses from 37 users. Sec. 4.2 shows that both variants of our method are highly preferred relative to the other baselines.

K. Realism User Study.

In order to focus on realism evaluation, we conduct a second user study asking users to select the “more realistic deformation” among the methods (N=23). We show screenshots in Fig. 24. Our 025 method is chosen by users 64% of the time, ARAP 17.7%, NeuralMLS 15.2%, biharmonic 2.2%, and APAP 0.93%, demonstrating our results are more re-

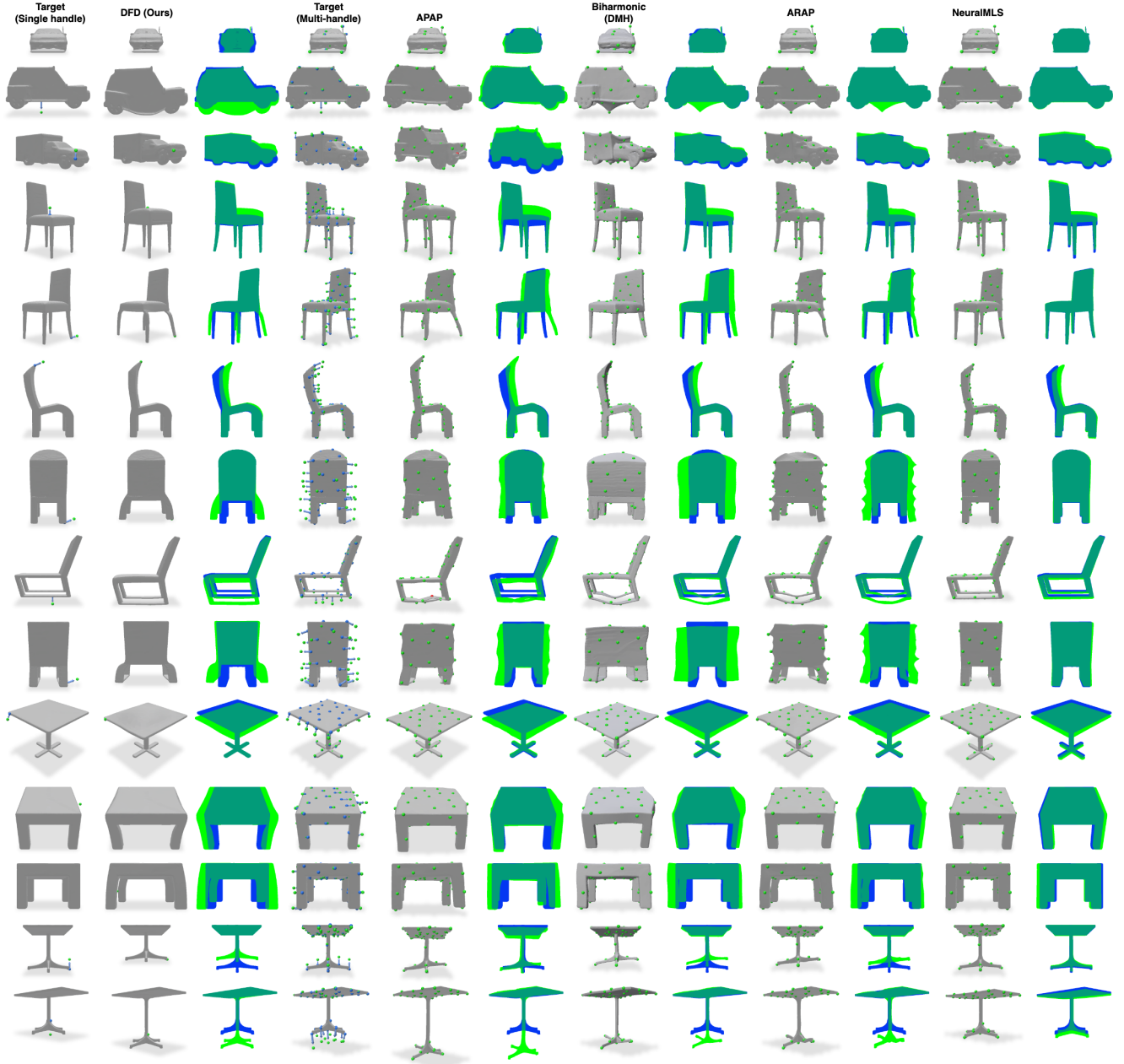


Figure 16. **Full DeepMetaHandles Comparison.** We show more examples comparing methods on the DMH dataset shapes and prescribed handle targets. As in the main paper, DFD weights are consistently just as smooth (or smoother) than DMH, while better-preserving shape semantics (e.g. part proportions, symmetries, etc).

liastic and shape-preserving from a user perspective.

L. Technical Details

Our feature field is parameterized by a 4-layer MLP with ReLU non-linearities and a LayerNorm after each hidden layer. The output is normalized to unit norm.

For all experiments, we sample 24 views using Fibonacci sampling. We optimize our feature field for 15 iterations,

and render at 512×512 . We use the Fast-Quadric-Mesh-Simplification [11, 36] wrapper from PyVista to perform our decimation. Our qualitative results use DINO features, though all image models we tried gave reasonable results (see supplemental). All experiments are run on a single A40 GPU with 48GB RAM.

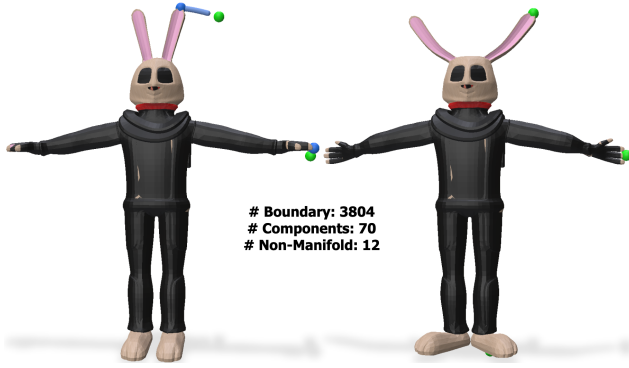


Figure 17. **Topology Robustness.** DFD weights are highly robust to topological defects. The example shown has 3,804 boundary edges, 70 disconnected components, and 12 non-manifold edges, but our weights still generate high quality deformations.

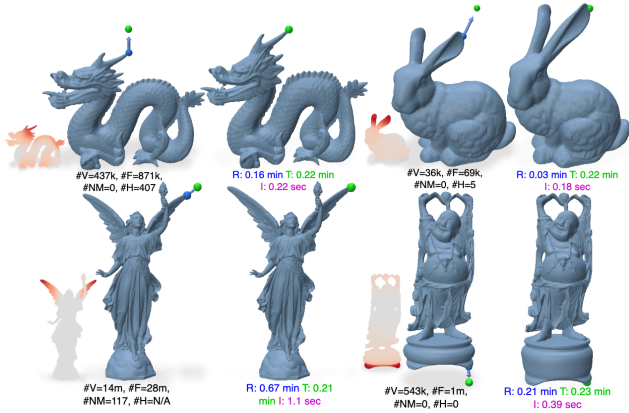


Figure 18. **High Resolution Deformation.** We test our distillation approach on very large meshes from the Stanford 3D Scanning Repository. #V is the number of vertices, #F is the number of faces, #NM is total number of non-manifold elements, #H is the number of holes. **R** reports the decimation/rendering time during distillation. **T** reports the remainder of the time taken for distillation. **I** reports the inference time. The mesh resolution influences the rendering stage (**R**), but otherwise distillation time (**T**) is *independent* of the mesh resolution.

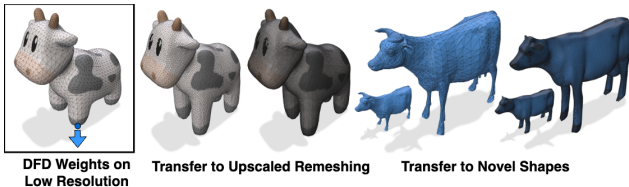


Figure 19. **Weight Generalization.** Distilling features into a neural field allows the same DFD weights to be applied to arbitrary resolution remeshings or even novel instances of the shape class.



Figure 20. **Consistent Weights.** Our DFD weights identify consistent visual relationships across shapes within the same class.

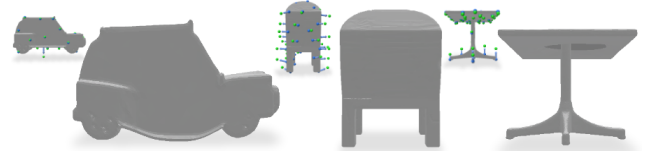


Figure 21. **Dense Handle Results.** We show deformations using our method using the same dense handle configurations as the baselines in Fig. 9.

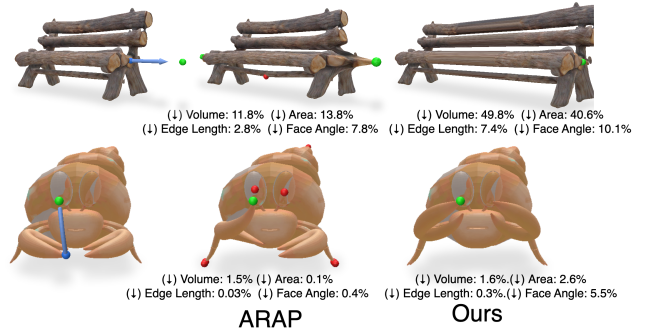


Figure 22. **Surface Metrics.** Though our method produces more surface distortion in terms of traditional metrics, the visual results demonstrate low surface distortion does not necessarily translate to a more realistic or natural deformation.

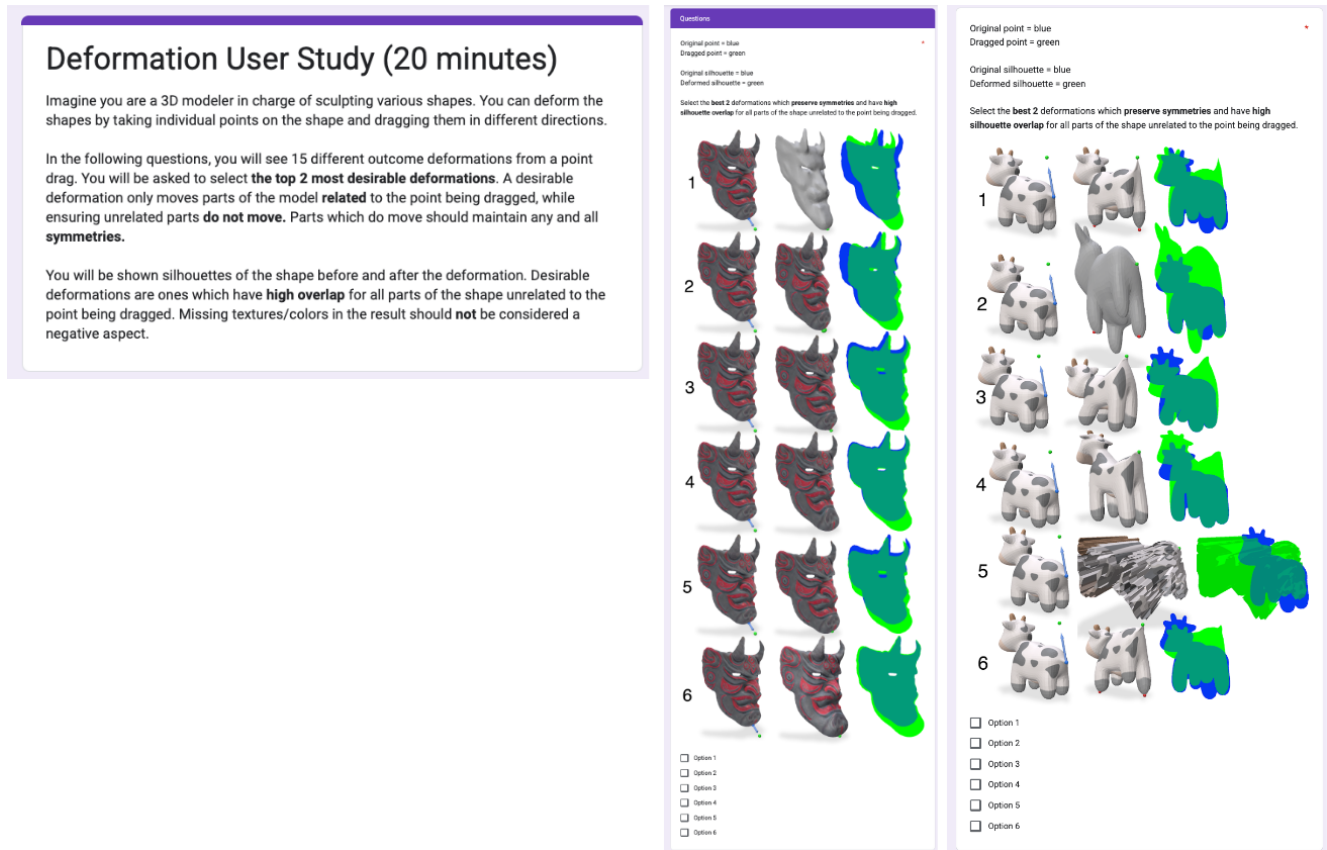


Figure 23. **User Study.** We show screenshots from our user study comparing deformations for various examples shown in the paper across all the methods (including both the affine and translation variants of our method).

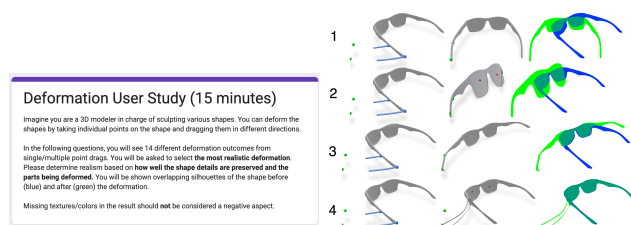


Figure 24. **Realism User Study.** We show screenshots from our second user study, which evaluates the quality of the deformations on the basis of realism.