

7. Supplementary

7.1. Overview

This document provides more detailed quantitative and qualitative results highlighting the strengths and limitations of AtlasNet.

Detailed results, per category, for the autoencoder

These tables report the metro reconstruction error and the chamfer distance error. It surprisingly shows that our method with 25 learned parameterizations outperforms our method with 125 learned parameterizations in 7 categories out of 13 for the metro distance, but is significantly worse on the cellphone category, resulting in the 125 learned parameterizations approach being better on average. This is not mirrored in the Chamfer distance.

Regularisation In the autoencoder experiment, we tried using weight decay with different weight. The best results were obtained without any regularization.

Limitations We describe two limitations with our approach. First, when a small number of learned parameterizations are used, the network has to distort them too much to recreate the object. This leads, when we try to recreate a mesh, to small triangles in the learned parameterization space being distorted and become large triangles in 3D covering undesired regions. On the other hand, as the number of learned parameterization increases, errors in the topology of the reconstructed mesh can be sometimes observed. In practice, it means that the reconstructed patches overlap, or are not stitched together.

Additional Single View Reconstruction qualitative results In this figure, we show one example of single-view reconstruction per category and compare with the state of the art, PointSetGen and 3D-R2N2. We consistently show that our method produces a better reconstruction.

Additional Autoencoder qualitative results In this figure, we show one example per category of autoencoder reconstruction for the baseline and our various approaches to reconstruct meshes, detailed in the main paper. We show how we are able to recreate fine surfaces.

Additional Shape Correspondences qualitative results

We color each vertex of the reference object by its distance to the gravity center of the object, and transfer these colors to the inferred atlas. We then propagate them to other objects of the same category, showing semantically meaningful correspondences between them. Results for the plane

and watercraft categories are shown and generalize to all categories.

Deformable shapes. We ran an experiment on human shape to show that our method is also suitable for reconstructing deformable shapes. The FAUST dataset [4] is a collection of meshes representing several humans in different poses. We used 250 shapes for training, and 50 for validation (without using the ground truth correspondences in any way). In table 5, we report the reconstruction error in term of Chamfer distance and Metro distance for our method with 25 squared parameterizations, our methods with a sphere parametrization, and for the baseline. We found results to be consistent with the analysis on ShapeNet. Qualitative results are shown in figure 14, revealing that our method leads to qualitatively good reconstructions.

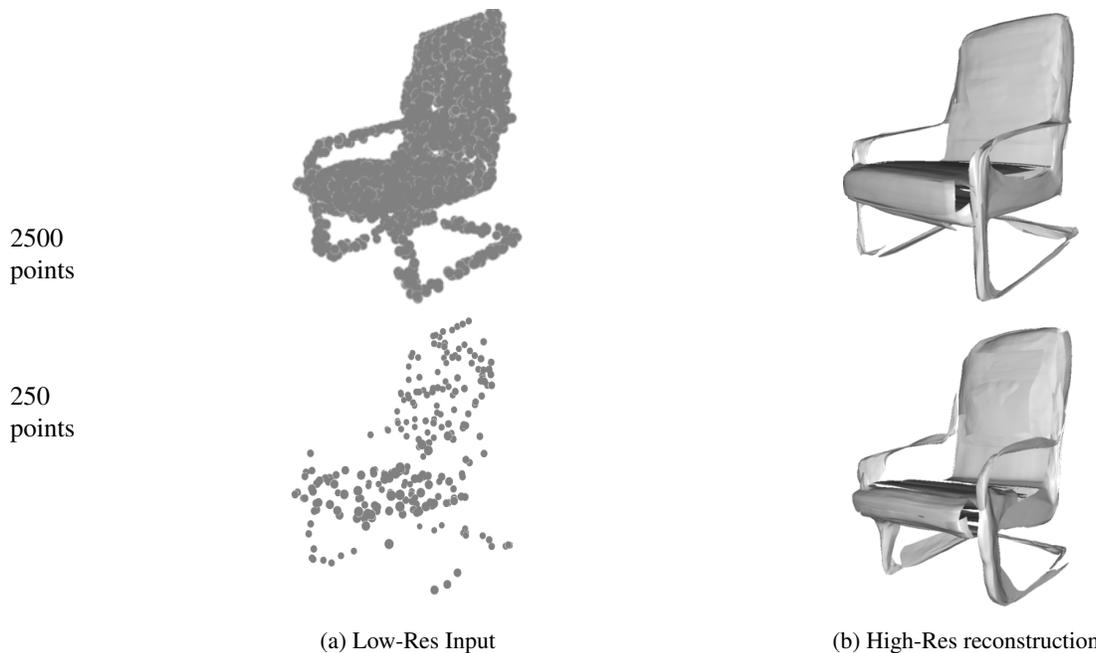
| | Chamfer | Metro |
|--------------|---------|-------|
| 25 patches | 15.47 | 11.62 |
| 1 Sphere | 15.78 | 15.22 |
| 1 Ref. Human | 16.39 | 13.46 |

Table 5. **3D Reconstruction on FAUST [4]**. We trained the baseline and our method sampling the points according from 25 square patches, and from a sphere on the human shapes from the FAUST dataset. We report Chamfer distance ($\times 10^4$) on the points and Metro distance ($\times 10$) on the meshes.

Point cloud super-resolution AtlasNet can generate pointclouds or meshes of arbitrary resolution simply by sampling more points. Figure 8 shows qualitative results of our approach with 25 patches generating high resolution meshes with 122500 points. Moreover, PointNet is able to take an arbitrary number of points as input and encodes a minimal shape based on a subset of the input points. This is a double-edged sword : while it allows the autoencoder to work with varying number of input points, it also prevent it from reconstructing very fine details, as they are not used by PointNet and thus not present in the latent code. We show good results using only 250 input points, despite the fact that we train using 2500 input points which shows the capacity of our decoder to interpolate a surface from a small number of input points, and the flexibility of our pipeline.

Details on the comparison against HSP [13] We perform a quantitative comparison against an octree-based state of the art method. AtlasNet is trained with 25 learned parameterizations on the same data as their publicly available trained model³. 100 random samples are drawn from each category from the test split. We evaluated the the quality of the reconstruction using the Chamfer distance on the normalized meshes, and the metro distance. In table 8, we report

³<https://github.com/chaene/hsp>.



(a) Low-Res Input

(b) High-Res reconstruction

Figure 8. **Super resolution.** Our approach can generate meshes at arbitrary resolutions, and the pointnet encoder [25] can take pointclouds of varying resolution as input. Given the same shape sampled at the training resolution of 2500, or 10 times less points, we generate high resolution meshes with 122500 vertices. This can be viewed as the 3D equivalent of super-resolution on 2D pixels.

per category results. As AtlasNet was specifically trained to optimise the chamfer distance, we outperform HSP in every category. AtlasNet outperforms HSP in metro distance for 10 categories out of 13. List of sampled used, ans trained model for AtlasNet are available in the github repository.

Limitations and future work Our results have limitations that lead to many open question and perspective for future work. First, the patches for our generated shapes are not guaranteed to be connected (except if the surface the input points are sampled from is already closed, as in the sphere experiment). An open question is how to effectively stitch the patches together to form a closed shape. Second, we have demonstrated results on synthetic object shapes. Ideally, we would like to extend to entire real scenes. Third, we have optimized the parameterization of the generated meshes post-hoc. It would be good to directly learn to generate the surfaces with low distortion parameterizations. Fourth, this work generates surfaces by minimizing an energy computed from point clouds. An open question is how to define a loss on meshes that is easy to optimize? Finally, as the atlases provide promising correspondences across different shapes, an interesting future direction is to leverage them for shape recognition and segmentation.

| | pla. | ben. | cab. | car | cha. | mon. | lam. | spe. | fir. | cou. | tab. | cel. | wat. | mean |
|-----------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| Baseline PSR | 2.71 | 2.12 | 1.98 | 2.24 | 2.68 | 1.78 | 2.58 | 2.29 | 1.03 | 1.90 | 2.66 | 1.15 | 2.46 | 2.12 |
| Baseline PSR PA | 1.38 | 1.97 | 1.75 | 2.04 | 2.08 | 1.53 | 2.51 | 2.25 | 1.46 | 1.57 | 2.06 | 1.15 | 1.80 | 1.82 |
| Ours 1 patch | 1.11 | 1.41 | 1.70 | 1.93 | 1.76 | 1.35 | 2.01 | 2.30 | 1.01 | 1.46 | 1.46 | 0.87 | 1.46 | 1.53 |
| Ours 1 sphere | 1.03 | 1.33 | 1.64 | 1.99 | 1.76 | 1.30 | 2.06 | 2.33 | 0.93 | 1.41 | 1.59 | 0.79 | 1.54 | 1.52 |
| Ours 5 patch | 0.99 | 1.36 | 1.65 | 1.90 | 1.79 | 1.28 | 2.00 | 2.27 | 0.92 | 1.37 | 1.57 | 0.76 | 1.40 | 1.48 |
| Ours 25 patch | 0.96 | 1.35 | 1.63 | 1.96 | 1.49 | 1.22 | 1.86 | 2.22 | 0.93 | 1.36 | 1.31 | 1.41 | 1.35 | 1.47 |
| Ours 125 patch | 1.01 | 1.30 | 1.58 | 1.90 | 1.36 | 1.29 | 1.95 | 2.29 | 0.85 | 1.38 | 1.34 | 0.76 | 1.37 | 1.41 |

Table 6. **Auto-Encoder (per category)**. The mean is taken category-wise. The Metro Distance is reported, multiplied by 10. The meshes were constructed by propagating the patch grid edges.

| | pla. | ben. | cab. | car | cha. | mon. | lam. | spe. | fir. | cou. | tab. | cel. | wat. | mean |
|-------------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| Baseline | 1.11 | 1.46 | 1.91 | 1.59 | 1.90 | 2.20 | 3.59 | 3.07 | 0.94 | 1.83 | 1.83 | 1.71 | 1.69 | 1.91 |
| Baseline + normal | 1.25 | 1.73 | 2.19 | 1.74 | 2.19 | 2.52 | 3.89 | 3.51 | 0.98 | 2.13 | 2.17 | 1.87 | 1.88 | 2.15 |
| Ours 1 patch | 1.04 | 1.43 | 1.79 | 2.28 | 1.97 | 1.83 | 3.06 | 2.95 | 0.76 | 1.90 | 1.95 | 1.29 | 1.69 | 1.84 |
| Ours 1 sphere | 0.98 | 1.31 | 2.02 | 1.75 | 1.81 | 1.83 | 2.59 | 2.94 | 0.69 | 1.73 | 1.88 | 1.30 | 1.51 | 1.72 |
| Ours 5 patch | 0.96 | 1.21 | 1.64 | 1.76 | 1.60 | 1.66 | 2.51 | 2.55 | 0.68 | 1.64 | 1.52 | 1.25 | 1.46 | 1.57 |
| Ours 25 patch | 0.87 | 1.25 | 1.78 | 1.58 | 1.56 | 1.72 | 2.30 | 2.61 | 0.68 | 1.83 | 1.52 | 1.27 | 1.33 | 1.56 |
| Ours 125 patch | 0.86 | 1.15 | 1.76 | 1.56 | 1.55 | 1.69 | 2.26 | 2.55 | 0.59 | 1.69 | 1.47 | 1.31 | 1.23 | 1.51 |

Table 7. **Auto-Encoder (per category)**. The mean is taken category-wise. The Chamfer Distance is reported, multiplied by 10^3 .

| | | pla. | ben. | cab. | car | cha. | mon. | lam. | spe. | fir. | cou. | tab. | cel. | wat. | mean |
|---------|---------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| metro | HSP | 2.82 | 2.65 | 2.18 | 2.41 | 2.88 | 2.47 | 3.30 | 3.39 | 1.86 | 2.84 | 2.83 | 1.71 | 2.52 | 2.61 |
| | Ours 25 patch | 1.99 | 2.15 | 2.24 | 2.02 | 2.38 | 2.44 | 3.05 | 3.18 | 2.19 | 2.59 | 2.40 | 1.62 | 2.64 | 2.38 |
| chamfer | HSP | 10.1 | 11.8 | 10.6 | 4.08 | 12.4 | 21.2 | 37.7 | 20.4 | 7.32 | 18.4 | 13.9 | 19.3 | 15.4 | 15.6 |
| | Ours 25 patch | 2.24 | 3.23 | 6.44 | 2.00 | 4.77 | 7.78 | 9.05 | 9.86 | 3.02 | 4.91 | 4.27 | 6.01 | 4.57 | 5.24 |

Table 8. **Single-view reconstruction**. Quantitative comparison against HSP [13], a state of the art octree-based method. The average error is reported, on 100 shapes from each category. The Chamfer Distance reported is computed on 10^4 points, and multiplied by 10^3 . The Metro distance is multiplied by 10.

| Weight Decay | Ours : 25 patches |
|--------------|-------------------|
| 10^{-3} | 8.57 |
| 10^{-4} | 4.84 |
| 10^{-5} | 3.42 |
| 0 | 1.56 |

Table 9. **Regularization on Auto-Encoder (per category)**. The mean is taken category-wise. The Chamfer Distance is reported, multiplied by 10^3 .

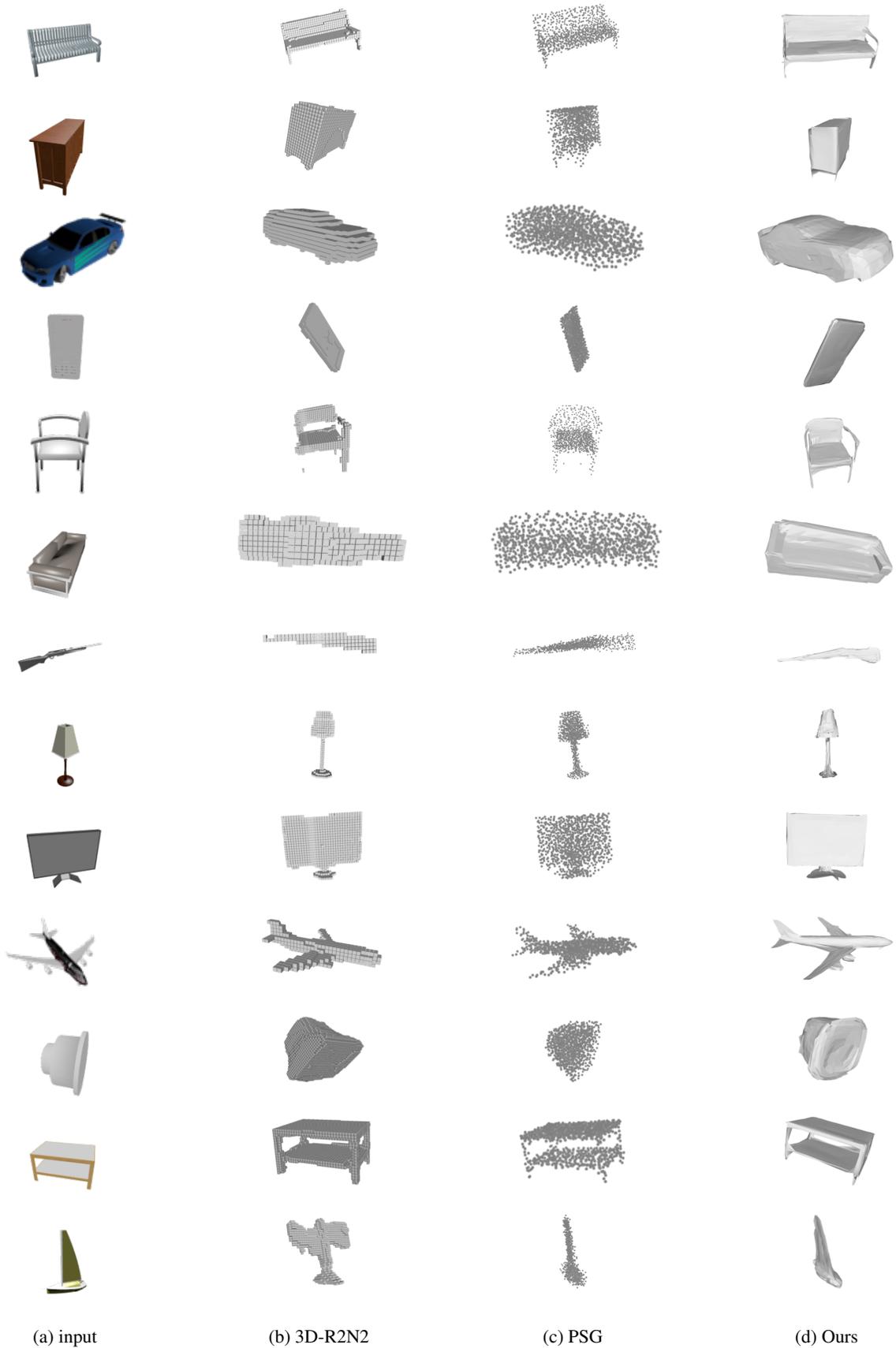
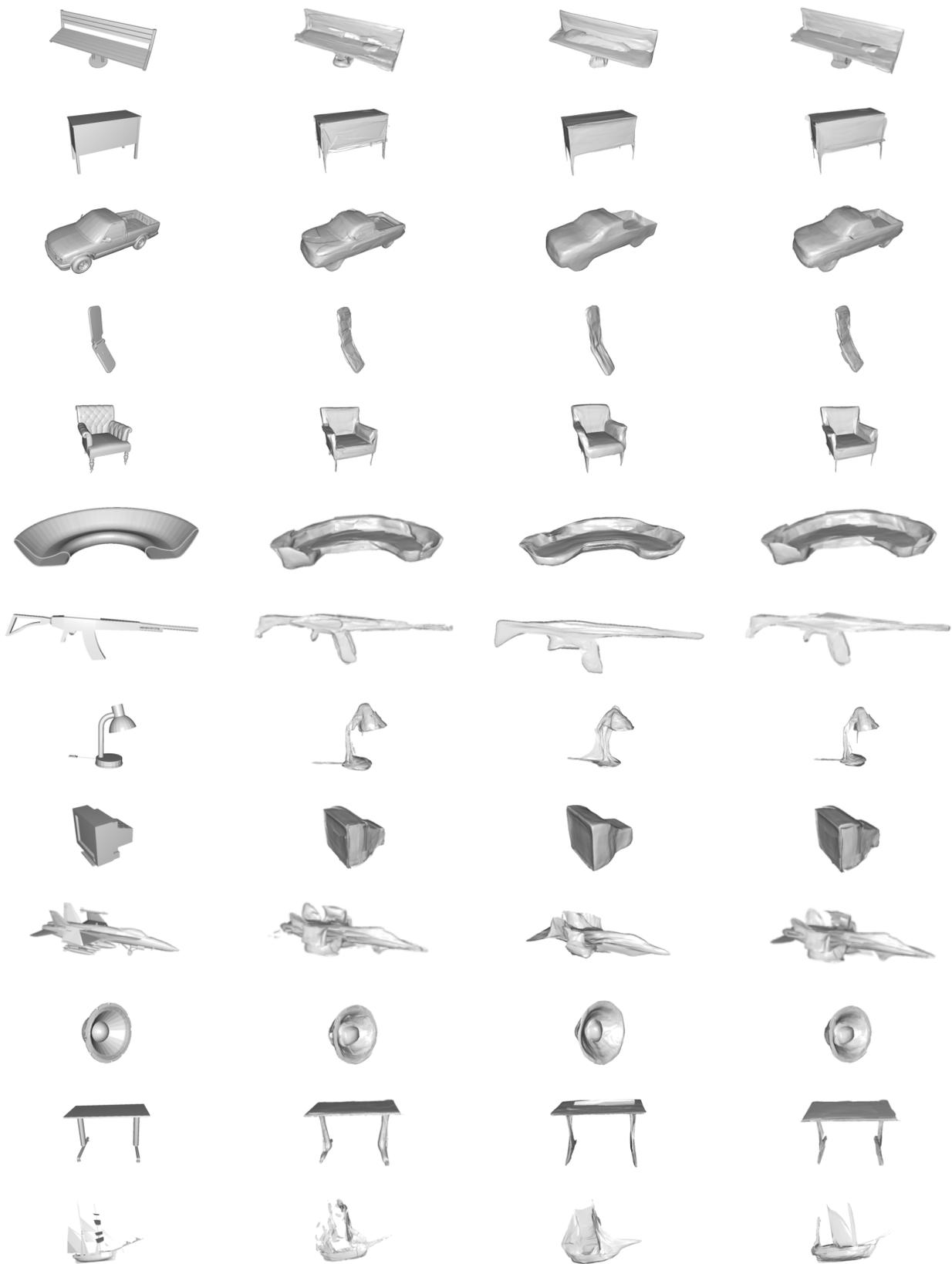


Figure 9. **Single-view reconstruction comparison:** From a 2D RGB image (a), 3D-R2N2 reconstructs a voxel-based 3D model (b), PointSetGen a point cloud based 3D model (c), and our AtlasNet a triangular mesh (d).



(a) Ground truth

(b) PSR on ours

(c) Ours sphere

(d) Ours 25

Figure 10. **Autoencoder comparison:** We compare the original meshes (a) to meshes obtained by running PSR (b) on the dense point cloud sampled from our generated mesh, and to our method generating a surface from a sphere (c), and 25 (d) learnable parameterizations.

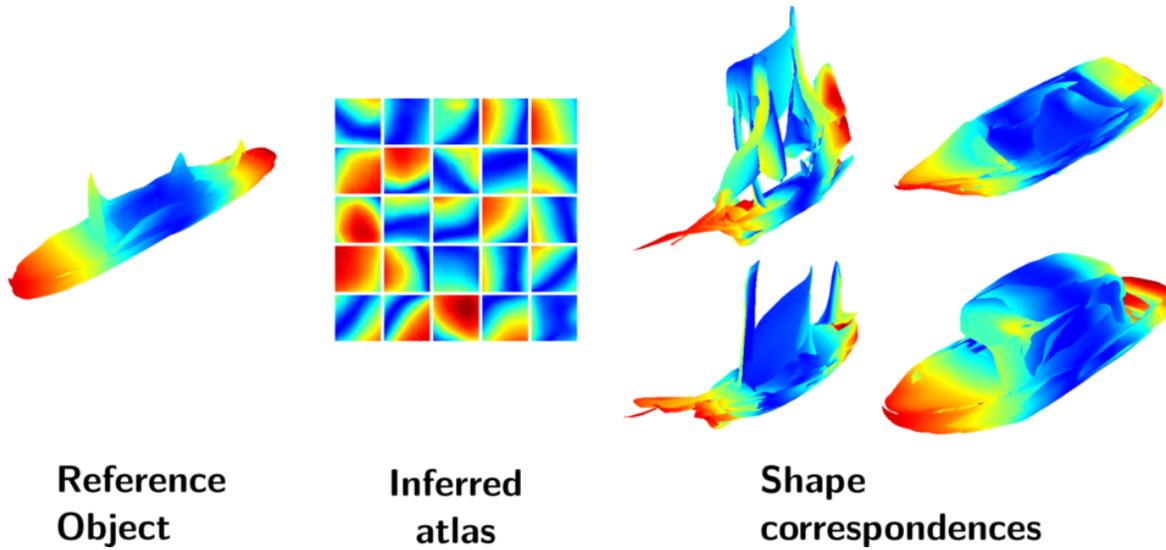


Figure 11. **Shape correspondences:** a reference watercraft (left) is colored by distance to the center, with the jet colormap. We transfer the surface colors to the inferred atlas for the reference shape (middle). Finally, we transfer the atlas colors to other shapes (right). Notice that we get semantically meaningful correspondences, without any supervision from the dataset on semantic information. All objects are generated by the autoencoder, with 25 learned parametrizations.

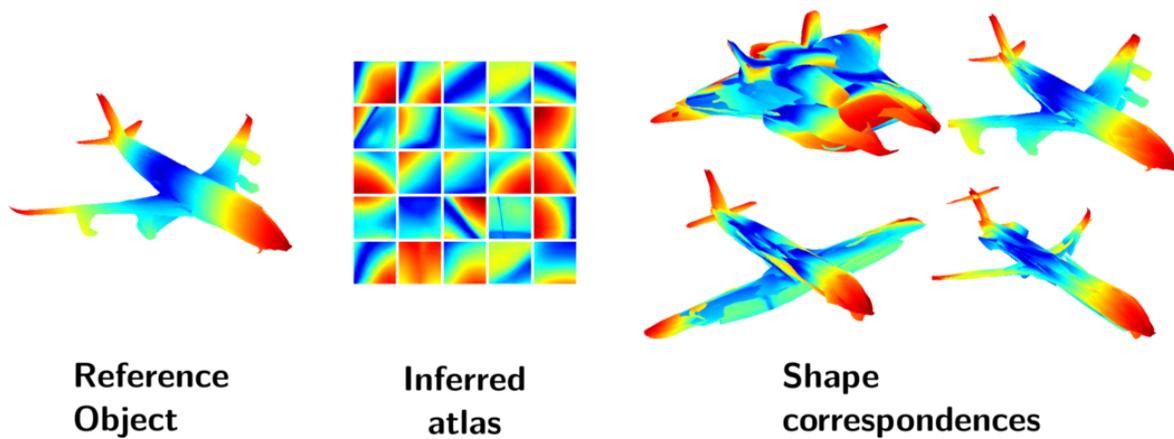
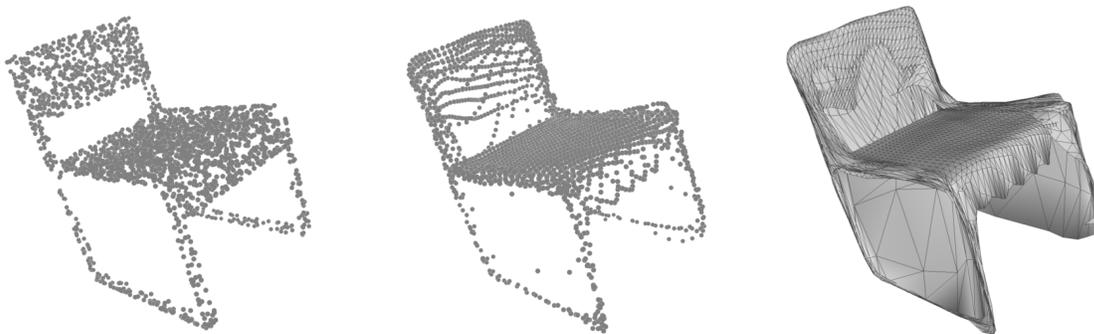
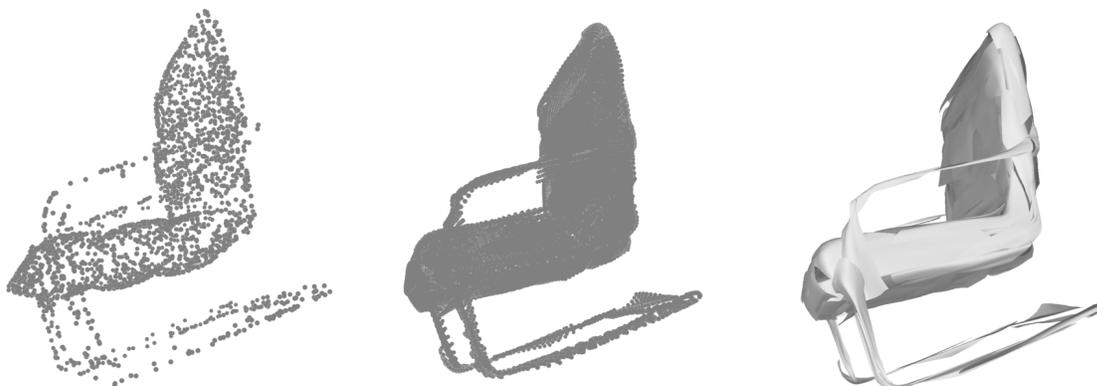


Figure 12. **Shape correspondences:** a reference plane (left) is colored by distance to the center, with the jet colormap. We transfer the surface colors to the inferred atlas for the reference shape (middle). Finally, we transfer the atlas colors to other shapes (right). Notice that we get semantically meaningful correspondences, such as the nose and tail of the plane, and the tip of the wings, without any supervision from the dataset on semantic information. All objects are generated by the autoencoder, with 25 learned parametrizations.



(a) **Excess of distortion.** Notice how, compared to the original point cloud (left), the generated pointcloud (middle) with 1 learned parameterization is valid, but the mapping from squares to surfaces enforces too much distortion leading to error when propagating the grid edges in 3D (right).



(b) **Topological issues.** Notice how, compared to the original point cloud (left), the generated pointcloud (middle) with 125 learned parameterizations is valid, but the 125 generated surfaces overlap and are not stitched together (right).

Figure 13. **Limitations.** Two main artifacts are highlighted : (a) Excess of distortion when too small a number of learned parameterizations is used, and (b) growing errors in the topology of the reconstructed mesh as the number of learned parameterization increases.



Figure 14. **Deformable shapes.** Our method learned on 250 shapes from the FAUST dataset to reconstructs a human in different poses. Each color represent one of the 25 parametrizations.