Learning Elementary Structures for 3D shape generation and matching

We thank the reviewers (Rs) for their comments. We are pleased to receive the positive reviews. If accepted, we will incorporate all feedback in the final version.

Generalization to new categories. (R1, R2, R3) To test the generality of our approach, we followed the reviewers' suggestion and trained on chairs using 10 2D elementary structures and tested on tables. As shown in Figure 1 (this rebuttal), point learning outperforms both transformation learning and AtlasNet trained with 10 patches - all Chamfer results in the rebuttal are multiplied by 10^{-3}. Figure 2 (this rebuttal) shows qualitatively how the elementary structures are positioned on chairs and tables. Notice how the chair and table legs are reconstructed by the same elementary structures.

	Chairs	Table
AtlasNet	1.64	4.70
Transfo.	1.56	4.82
Point.	**1.34**	**4.45**

Figure 1: **Generalization.** Chamfer loss results of the networks trained on chairs and tested on either the chairs or tables test set.

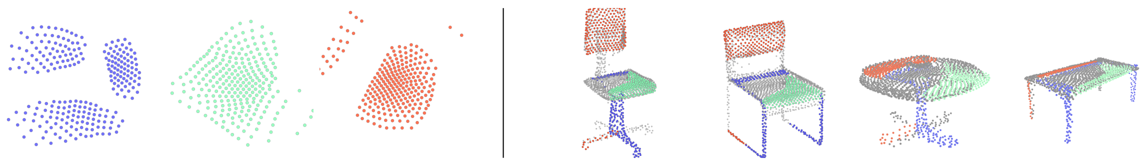

Figure 2: Elementary structures learned on chairs (**left**) used to reconstruct chairs and tables (**right**).

Where does the performance boost come from? (R1-II, R2) In Figure 3 (this rebuttal), we show the number of parameters for AtlasNet and our method. Our method has less than 1% additional parameters to learn the elementary structures – 2.0×10^6 and 2.5×10^3 for transformation and point learning, respectively (notice that they are orders of magnitude smaller than 1.8×10^8). During inference, our approach has the same complexity as AtlasNet as the elementary structures are precomputed and remain fixed for all shapes. As suggested, we also tried training AtlasNet with 6 layers (6-layer AN), which significantly increases the number of parameters. Our approach with points learning outperforms all methods.

	Param.	Chamfer
AtlasNet	1.8×10^8	1.45
6-layer AN	3.9×10^8	1.35
Transfo.	1.8×10^8	1.43
Point.	1.8×10^8	**1.22**

Figure 3: **Impact of number of parameters on reconstruction error.**

Consistency in template elementary structures. (R1-I) We extended the experiment with SURREAL from Figure 5 of our paper to the plane category of ShapeNet using the point learning method. We used a single 3D elementary structure as in the SURREAL experiment. In Figure 4 (this rebuttal), we initialized the elementary structure with either a plane 3D model **(left)** or a set of random 3D points sample uniformly **(right)**. Notice that (1) the learned 3D elementary structures are consistent regardless the template shape and (2) we do not need to input heuristic basis functions since using a set of random 3D points give similar results.

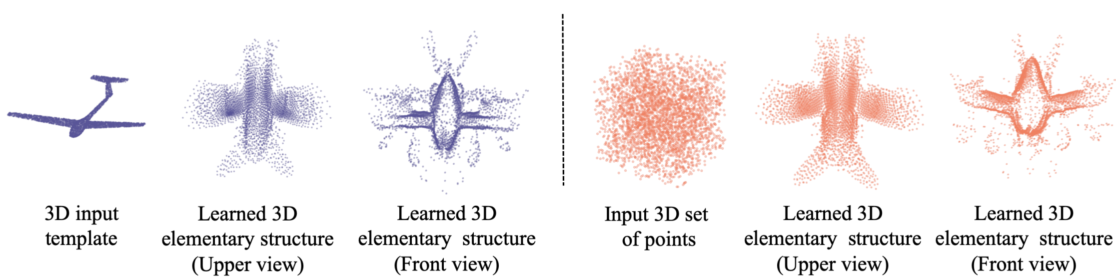

Figure 4: **Robustness of the learned 3D elementary structure.**

Comparison to AltasNet-trained models. (R1-III) Using the trained models from the official implementation on all categories, AtlasNet-25 performance is 1.56 (see also Table 1 in the Atlasnet paper). Using the released code to train AtlasNet-10 yields 1.55 of performance. In our paper, we added a learning rate schedule to the original implementation and got an error of 1.45 (see Table 1 of our paper). Using the same learning rate schedule, PointLearning-10 and TransformationLearning-10 perform, respectively, 1.22 and 1.43. For reference, PointLearning-25 and TransformationLearning-25 perform, respectively, 1.21 and 1.40 – a significant 22% and 9% boost.

Details, References, Writing. (R1, R2, R3) Results in Table 1 (from the paper) are evaluated on the test set (**R1**). We will mitigate the claim that elementary structures correspond to semantic parts (**R1**), add missing discussion on Kanazawa et al. (**R1**) and improve the consistency of the notations (**R3**).