Supplemental: 3D VR Sketch Guided 3D Shape Prototyping and Exploration

Ling Luo[1,2] Pinaki Nath Chowdhury[1,2] Tao Xiang[1,2] Yi-Zhe Song[1,2]
Yulia Gryaditskaya[1,3]

[1]SketchX, CVSSP, University of Surrey, United Kingdom
[2]iFlyTek-Surrey Joint Research Center on Artificial Intelligence
[3]CCM, Surrey Institute for People-Centered AI and CVSSP, University of Surrey, United Kingdom

S1. Additional results

More generation results of the proposed method are shown in Fig. S2. Please note that the scales of sketches and 3D shapes differ in these visual results due to the fact that we use different visualization tools. To demonstrate that sketches and reconstructed shapes generally are well-aligned, we provide Fig. S1, created using MeshLab's snapshot.

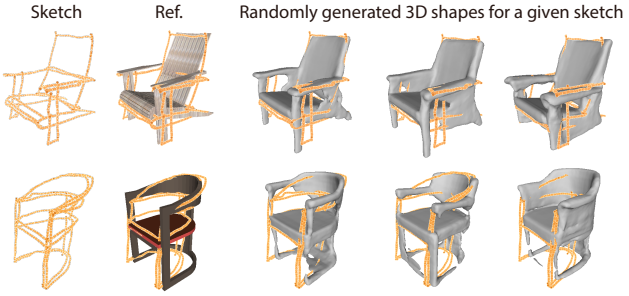

Figure S1. Alignment between sketches and shapes visualized using MeshLab's snapshot.

S1.1. Performance analysis as a function of sketch accuracy

We use a small VR sketch dataset proposed in [4] as an additional test set to explore how our method generalizes to sketches of varying quality. This dataset contains Freehand Sketches (FS) of 139 chairs and 28 bathtub shapes from the ModelNet10 test set. It also contains corresponding curve networks (CNs), which are the minimal set of curves required to accurately represent a 3D shape [1]. Therefore, our method is expected to produce 3D reconstructions from these networks with lower diversity and better sketch fidelity than from freehand sketches.

Abstract sketches vs. clean curve networks We show a qualitative comparison of reconstruction results from abstract sketches and matching detailed curve networks in Fig. S3. It can be observed that, indeed, the sampled shapes conditioned on the curve networks contain less diversity.

Data	$\mathcal{F}_{shape}(G^{e^c}) \downarrow$	$\mathcal{F}^{avg}_{\text{sketch}}(G^{d^c}) \downarrow$	$\mathcal{F}^{\text{avg}}_{\text{shape}}(G^{d^c}) \downarrow$	$\mathcal{D}_{\text{gnrtns}} \uparrow$
[5]	0.368	0.017±0.034	0.431	0.161
[4] FS	0.366	0.028±0.036	0.412	0.161
[4] CN	0.237	0.017±0.032	0.298	0.142

Table S1. Numerical comparison between generations results from freehand sketches (FS) and curve networks (CN). The first line shows the results from the main paper where the test set of freehand sketches represents shapes from the ShapeNetCore dataset. Note that the training data comes from the same dataset, but a different set of shapes. The second line shows the results from the freehand sketches from [4], which represent shapes from the ModelNet dataset. Finally, the third line shows the generation results on the curve networks for the shapes from the ModelNet dataset.

We provide a numerical comparison between generations conditioned on sketches and curve networks of the 139 chair instances in Tab. S1.

If we compare the metric on the curved networks and freehand sketches from [4] for 3D shapes from the ModelNet dataset, then we can see that indeed the reconstructions from clean curve networks are much closer to the reference 3D shape ($\mathcal{F}_{\text{shape}}$), the input ($\mathcal{F}_{\text{sketch}}$) (in the case of a curve network "sketch" = "curve network"), and the diversity of generated shapes is smaller ($\mathcal{D}_{\text{gnrtns}}$).

We can also see that for the two test sets of freehand sketches: the one used in the main paper and the one from [4], the diversity and similarity to a reference 3D shape are both similar. However, the fidelity to the sketch is larger, which we attribute to a large number of sparse sketches in that test set, as can be observed by comparing sketch samples in Fig. S3 and Fig. S2.

Overall, this experiment shows good generalization properties to sketch styles and shapes from a different dataset of the same category.

Figure S2. Additional generation results using chair sketches in [5]. 'Ref.' shows the reference 3D shape. 'AE' shows the deterministic prediction by our AE from the first stage of our method. 'Mean' denotes the shape reconstructed from the sample corresponding to the mean of the conditional distribution. And finally, we show 5 randomly generated shapes conditioned on the input sketch, sorted in the order of fidelity to a reference shape.

In Fig. S4, we show what happens if the test set comes from a completely different distribution than the training set. Namely, we use as conditioning freehand sketches and curve networks of bathtubs. *It can be observed, that our approach produces valid or physically plausible chairs which follow the overall shape of the conditioning, in turn demonstrating the robustness of our method, and its ability to generate new shapes.*

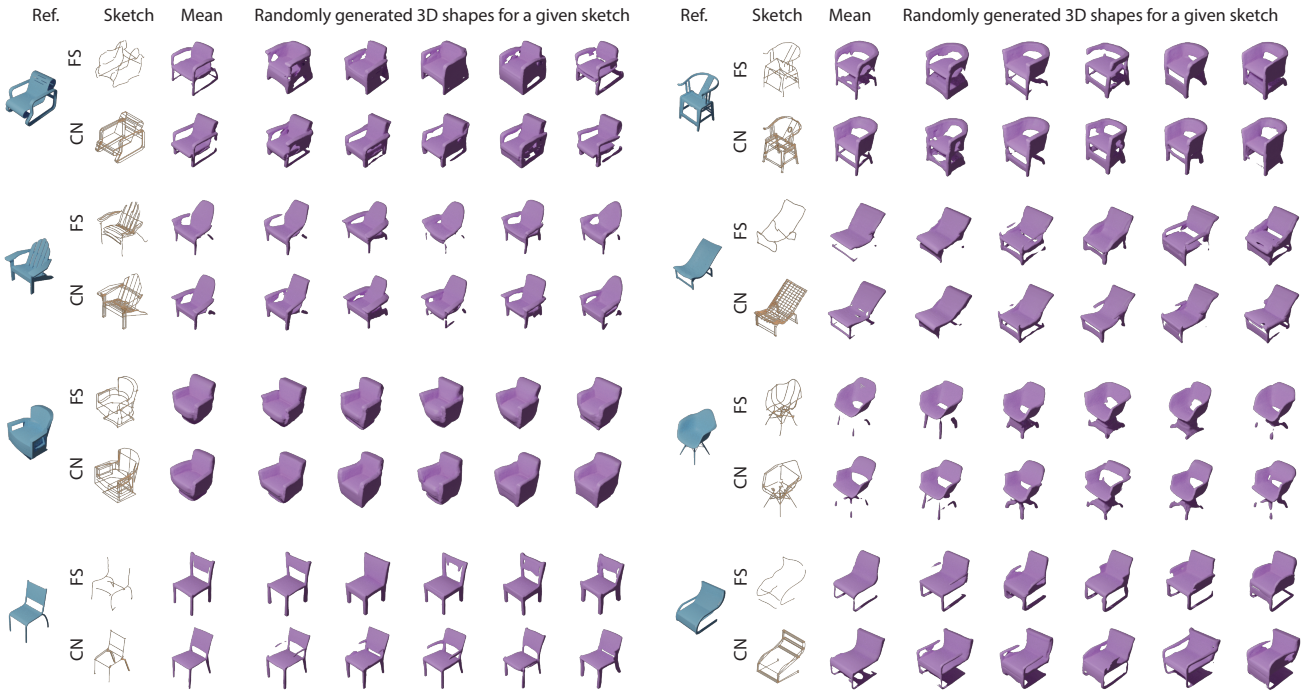

Figure S3. Generation results using chair sketches from [4]. 'Ref.' shows the reference 3D shape. 'Mean' denotes the shape reconstructed from the sample corresponding to the mean of the conditional distribution. And finally, we show 5 randomly generated shapes conditioned on the input sketch.

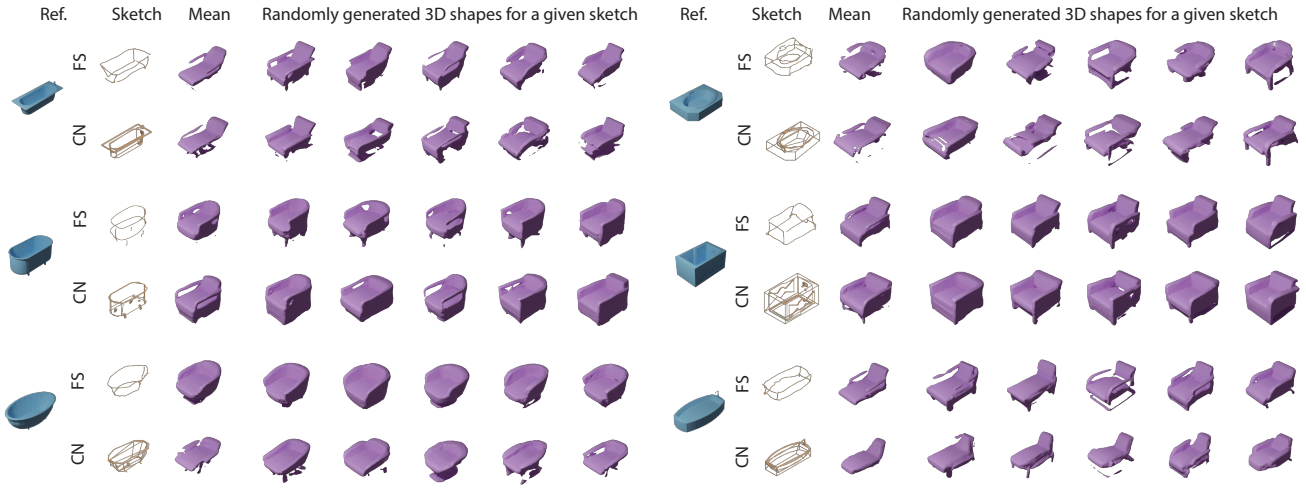

Figure S4. Generation results using bathtub sketches from [4]. 'Ref.' shows the reference 3D shape. 'Mean' denotes the shape reconstructed from the sample corresponding to the mean of the conditional distribution. And finally, we show 5 randomly generated shapes conditioned on the input sketch.

S1.2. Comparison with retrieval

We compute our fidelity and diversity metrics on the top 5 retrieval results of [3] with the ground-truth shape in the gallery. Tab. S2 shows that retrieval results have much lower fidelity to sketches and ground-truth shapes. Additionally, note that if there is no shape close to the sketch in the 3D gallery, the retrieval has no chance to succeed.

S1.3. Evaluation on sketches with style not seen during training

Following [4], our test set includes 5 participants those sketches are not used for training. Tab. S2 shows generalization to lower quality 202 VR chair sketches from memory *FVRS-M* [3].

Experiment	$\mathcal{F}^{\text{avg}}_{\text{sketch}}(\cdot) \downarrow$	$\mathcal{F}^{\text{avg}}_{\text{shape}}(\cdot) \downarrow$	$\mathcal{D}_{\text{gnrtns}} \uparrow$
Unseen 5	0.019±0.035	0.460	0.171
FVRS-M	0.030±0.041	0.431	0.161
Retrieval [3]	0.061±0.056	1.132	1.138

Table S2. The first two rows show the performance of our method on (1) the set of sketches by 5 participants those sketches were not observed during training – they represent new shapes and new styles; (2) sketches from the FVRS-M dataset [3]. The last row provides numerical evaluation on the quality of the top-5 3D shapes returned by the retrieval method by Luo et al. [3].

S1.4. Comparison with 2D sketch-based modeling

We compare with the performance of a sketch-based modeling method, following closely ours, but conditioned on 2D sketches by novices [7]. These sketches match 3D shapes in our training and test sets of VR sketches. We feed all 3 available views to the [2] encoder, which we train with $\mathcal{L}_{\text{align}}$ (Eq.12) and $\mathcal{L}_{\text{SDF}}$ (Eq.5) and leverage our pre-trained decoder, similar to the setting in Tab. 1, lines 3-4, in the main paper. We obtain $\mathcal{F}^{2D}_{\text{shape}} = 0.833$ which is much worse than on 3D sketches: 0.437 (Tab. 1, line 3).

S1.5. Interpolation in sampling space

To verify the continuity of our sampling space, we perform interpolation between two generated shapes in the latent space of CNF. Given the same sketch condition, we first generate two random samples. We then linearly interpolate their latent codes to create 3 additional samples. We then pass those new codes through the reverse path of the CNF with the same sketch condition to map back to the embedding space of the autoencoder, which allows us to compute their SDF representation. The visual interpolation results are shown in Fig. S5.

The figure shows that the interpolation is quite smooth. Therefore, the user can choose any two generated results and explore the shapes in between. *This further supports shape exploration enabled by our method.*

S2. Training details

S2.1. Sampling SDF for training

As we mention in Sec. 4.1, to prepare the training data, we compute SDF values on two sets of points: close to the shape surface and uniformly sampled in the unit box, following [6].

Close to the shape surface First, we randomly sample 250k points on the surface of the mesh. During sampling, the probability of a triangle to be sampled is weighted according to its area. We then generate two sets of points near the shape surface by perturbing each surface point coordinates with random noise vectors sampled from the Gaussian distribution with zero mean and variance values of 0.012 or 0.035.

To obtain two spatial samples per surface point, we perturbed each surface point along all three axes (X, Y, and Z) using zero mean Gaussian noise with variance values of 0.012 and 0.035.

Uniformly sampled in the unit box Second, we uniformly sample 25k points within the unit box. This results in a total of 525K points being sampled.

Loss evaluation When computing the SDF L_1 loss (Eq. (5)) during training, we sample a subset of 8,192 points from the pre-computed set of points for each SDF sample. We then compute the L_1 loss between the predicted SDF values and the ground truth SDF values at the sampled 3D points coordinates.

S2.2. Sampling point clouds from sketch strokes and 3D shapes

To obtain point cloud representations of shapes and sketches, prior to training, we sample 4,096 points for both shapes and sketches. For shapes, we uniformly sample from the mesh surface[1]. VR sketches are made up of strokes that include a set of vertices and edges. Luo et al. [5] provide sketch point clouds, each containing 15,000 points sampled uniformly from all sketch strokes. We apply Furthest Point Sampling (FPS) to sample 4,096 points from the 15,000 points.

References

[1] Giorgio Gori, Alla Sheffer, Nicholas Vining, Enrique Rosales, Nathan Carr, and Tao Ju. Flowrep: Descriptive curve networks for free-form design shapes. *ACM Transactions on Graphics (TOG)*, 36(4):1–14, 2017. 1

[2] Xinwei He, Tengteng Huang, Song Bai, and Xiang Bai. View n-gram network for 3d object retrieval. In *ICCV*, 2019. 4

[3] Ling Luo, Yulia Gryaditskaya, Tao Xiang, and Yi-Zhe Song. Structure-aware 3D VR sketch to 3D shape retrieval. In *3DV*, 2022. 3, 4

[4] Ling Luo, Yulia Gryaditskaya, Yongxin Yang, Tao Xiang, and Yi-Zhe Song. Towards 3D VR-sketch to 3D shape retrieval. In *3DV*, 2020. 1, 3

[5] Ling Luo, Yulia Gryaditskaya, Yongxin Yang, Tao Xiang, and Yi-Zhe Song. Fine-grained vr sketching: Dataset and insights. In *3DV*, 2021. 1, 2, 4

[6] Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, and Steven Lovegrove. Deepsdf: Learning continuous signed distance functions for shape representation. In *CVPR*, 2019. 4

[7] Anran Qi, Yulia Gryaditskaya, Jifei Song, Yongxin Yang, Yonggang Qi, Timothy M Hospedales, Tao Xiang, and Yi-Zhe

[1] https://github.com/fwilliams/point-cloud-utils

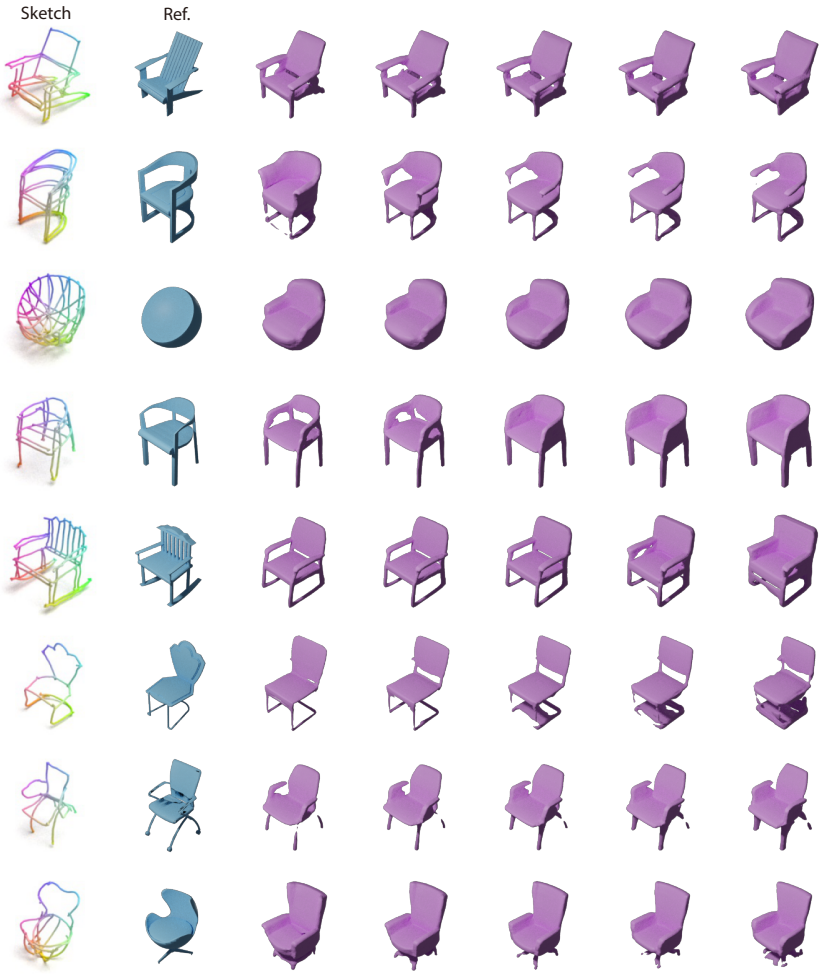

Figure S5. Interpolation results in sampling space. Note the smooth transitions between the first and last purple objects.

Song. Toward fine-grained sketch-based 3D shape retrieval. *TIP*, 2021. 4