Supplemental Material
DifferSketching: How Differently Do People Sketch 3D Objects?

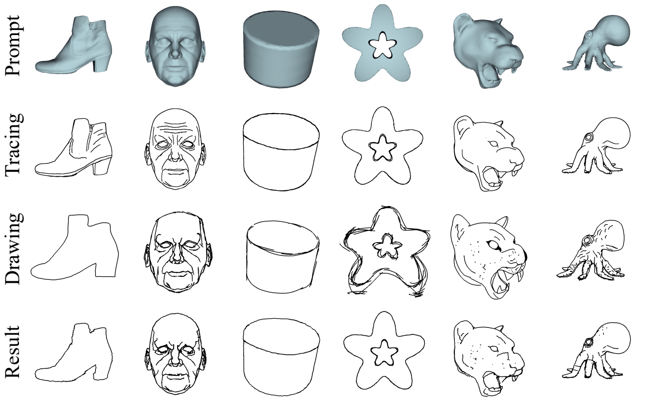
CHUFENG XIAO*, School of Creative Media, City University of Hong Kong, China
WANCHAO SU*, School of Creative Media & Department of Computer Science, City University of Hong Kong, China
JING LIAO, Department of Computer Science, City University of Hong Kong, China
ZHOUHUI LIAN, Wangxuan Institute of Computer Technology, Peking University, China
YI-ZHE SONG, SketchX, CVSSP, University of Surrey, UK
HONGBO FU†, School of Creative Media, City University of Hong Kong, China

1 DATASET OVERVIEW

Please find the thumbnails of all the collected sketches as well as the corresponding prompts and the multi-level reigister results for the 9 categories in the accompanying file folder. For a thumbnail of each category, except for the first column (rendered image prompts and tracings), the left groups of five sketches are from the novices, while the right groups of five drawings are from the professionals, including the drawings and their corresponding registered results on sketch-level, stroke-level, pixel-level (from top to bottom) on the right side.

2 MULTI-LEVEL REGISTRATION

We implemented our pixel-level registration based on the point-to-point registration using SimpleITK [Yaniv et al. 2018] following some settings of Wang et al. [Wang et al. 2021]. In the initialization stage for a displacement field, we applied global affine transformations and B-Spline transformations at four cascaded scales of the images. B-Spline transformation is a deformable transform over a bounded spatial domain using a B-Spline representation for 2D coordinate space, of which the parameters are default values in SimpleITKv4 [Yaniv et al. 2018]. We applied the automatic-initialized displacement field to the sketches one by one and filtered out the failure cases, each of which is far away from the target center or with messy lines. Similar to Wang et al. [2021], we recruited another group of four users to manually assign the corresponding points between the freehand sketches and the corresponding prompt images for the sketches where the automatic initialization failed. We set 4-39 corresponding points for each of 745 freehand sketches and fitted them to thin plate spline models for the semi-automated initialization of the badly initialized sketches. At the optimization stage (`ImageRegistrationMethod()` in SimpleITK), we performed the optimization to get the optimal displacement maps by maximizing the correlation between the sketches and the tracings using stochastic gradient descent optimizer.



(a) Under-sketching (b) Over-sketching (c) Semantically-distinct

Fig. 1. Examples of our proposed pixel-level registration for the three types of special cases. Please find the corresponding progressive results in Figure 2. The sketches (from left to right) were drawn respectively by *Novice 49, Professional 38, Novice 3, Professional 6, Professional 18, and Professional 34.*

We implemented pixel-level registration by incorporating the point-to-point registration and iterative rasterize-and-optimize process, As shown in Figure 1, there exist three special cases in our collected drawings and targeted tracings: a) Under-sketching: a drawing could not cover all of the shape information as the corresponding tracing does; b) Over-sketching: a drawing repeats several strokes to depict the same shape region; c) Semantically-distinct: a drawing contains strokes without semantic correspondence with any line of the corresponding tracing. The "Result" row of Figure 1 shows that our proposed pixel-level registration method performs well for these special cases. Please find the registration progress in Figure 2.

3 DATA ANALYSIS

3.1 Basic Statistics

We conducted the statistical analysis on the collected freehand sketches in both spatial and temporal perspectives, within and across the two user groups. We first report the basic features of our analysis. The average numbers of strokes for novices and professionals in our dataset were 70.45 and 105.00, respectively. The average time spent in drawing a sketch was 3.40 minutes for novices and 4.48 minutes for professionals. Professionals used the undo function more often

*Authors contributed equally.
†Corresponding author.

Authors' addresses: Chufeng Xiao, School of Creative Media, City University of Hong Kong, China, chufeng.xiao@my.cityu.edu.hk; Wanchao Su, School of Creative Media & Department of Computer Science, City University of Hong Kong, China, wanchao.su@cityu.edu.hk; Jing Liao, Department of Computer Science, City University of Hong Kong, China, jingliao@cityu.edu.hk; Zhouhui Lian, Wangxuan Institute of Computer Technology, Peking University, China, lianzhouhui@pku.edu.cn; Yi-Zhe Song, SketchX, CVSSP, University of Surrey, UK, y.song@surrey.ac.uk; Hongbo Fu, School of Creative Media, City University of Hong Kong, China, hongbofu@cityu.edu.hk.

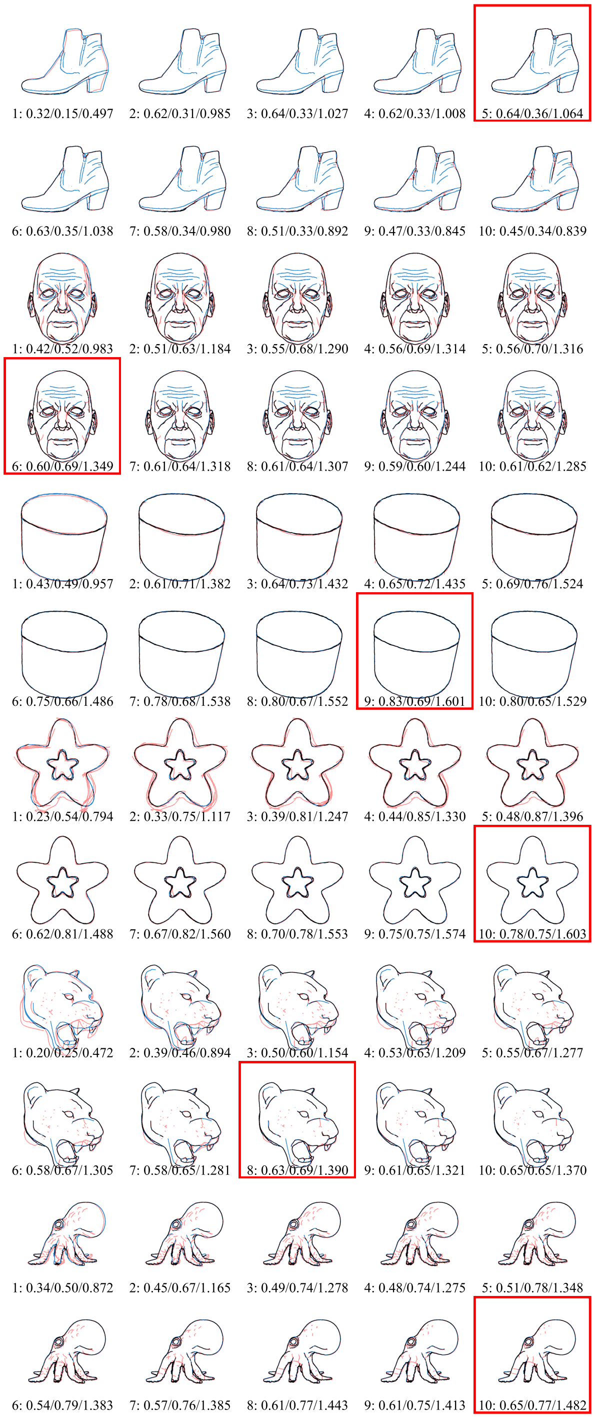
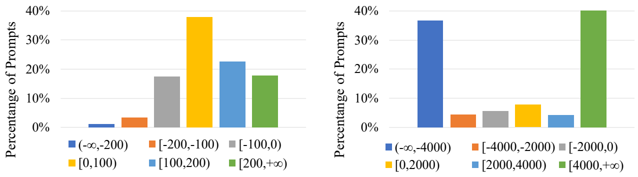


Fig. 2. The progressive results by our proposed pixel-level registration approach for Figure 1. Each image visualizes the registered results of each iteration, where the drawing pixels that have not been registered correctly are in red and the tracing pixels to be registered by the drawings are in blue, while the pixels overlapped by them are in black. The bottom text of each image indicates the performance evaluation ("i-th iteration: E_i / P_i / R_i"), by which we picked the optimal iteration as the final result (red bounding box).



(a) Difference in stroke number

(b) Difference in drawing length

Fig. 3. Two histograms of the differences between the drawings by professional and novice users across all the categories on stroke number and drawing length (accumulated stroke length). The differences are computed by subtracting the total value of five novice users from that of five professional users for the same prompt (P-N).

than novices (17.75 times versus 8.79 times on average). By calculating the pause duration proportions for each drawing, we found that during sketching, all the users spent more time pausing than drawing. Specifically, novice users spent 64%-82% of the time pausing when creating sketches for different categories of objects; and professional users's pausing time ranged from 70% to 83%. Professional users spent slightly more time on pausing than novice users on average (76% vs. 70%). All the users drew with similar strength, with a slightly milder average pressure of 0.40 for the professional drawings (novice as 0.47, range: 0 (no pressure) - 1 (hardest)). Please find more categorical statistics in Table 1.

By comparing the drawings in terms of groups using the category-level statistics, we found that the professional users drew more strokes for the prompts in most categories (7 out of 9). For categories with sharp shapes, i.e., *Chair*, *Industrial Component*, *Lamp*, *Primitive*, the novice users drew longer paths for the majority of the prompts. For the prompts with complex geometries, the professionals drew paths longer than the novices. We computed the difference of accumulated numbers of strokes between the professional and the novice groups for all the drawings. The results are shown in Figure 3 (a). Generally, the professionals drew more strokes for over 78% of the prompts. As to the accumulated stroke length, it shows a bimodal trend (see Figure 3 (b)): for 41% of the prompts, the novices drew more pixels (>4000 pixels) than the professionals, yet for other 41% prompts, the professional users drew more pixels (>4000 pixels). Compared to the novice group, professional users preferred to draw longer paths for complicated models, e.g., *Animal*, *Human Face*, *Vehicle*, etc.

Conclusion: The professional users tended to draw more strokes and spend longer time than the novice users given the same prompts and the professionals modified the drawings more frequently than the novices, especially for the models with richer details.

3.2 Where Does Key Difference Occur When Sketching?

From a sketch-level perspective, over 55% of the drawings from the two groups require small rotation ($R_G \leq 4°$) to match the prompts globally. It shows that people tended to sketch 3D objects globally similar to the prompt orientation, and the tendency is stronger in the professional group. There are more professional drawings with the translation distance under 100 pixels (69% drawings) and the

Table 1. Extended categorical statics for our dataset. We report the comparisons between the drawings by professional (P) and novice (N) users in terms of the number of strokes and the path length. We compute the ratio of the prompts in each category whose corresponding professional sketches contain more strokes than the novice sketches. Similarly, we also report the ratios for stroke path length (in pixels), the average time used for drawing sketches within each category, the ratio of pause duration over the whole drawing time, and the average times of using the undo function. For the last three statistics, we report the values in $novice/professional$ pairs.

Category	Animal	Animal Head	Chair	Human Face	Industrial Cpnt	Lamp	Primitive	Shoe	Vehicle
Stroke Number (P>N)	0.30	0.75	0.75	0.61	0.8	0.76	0.30	0.78	1.00
Path Length (P>N)	0.77	0.53	0.45	0.67	0.06	0	0.45	0.73	0.98
Time (minutes)	3.10/4.91	3.09/3.49	3.45/4.37	8.25/11.02	3.06/3.82	3.59/2.83	1.97/2.16	2.97/4.23	3.32/6.18
Pause Proportion	0.69/0.75	0.74/0.74	0.68/0.78	0.82/0.83	0.69/0.70	0.70/0.72	0.75/0.79	0.64/0.76	0.71/0.77
Undo Count	7.38/20.59	6.77/12.28	11.46/24.87	22.67/42.04	7.61/21.96	7.23/10.58	14.48/12.87	5.51/12.50	7.28/15.24
Pressure Average	0.46/0.44	0.49/0.41	0.45/0.43	0.46/0.43	0.54/0.41	0.49/0.41	0.44/0.30	0.46/0.36	0.45/0.43
Pressure Std	0.10/0.10	0.11/0.10	0.10/0.10	0.11/0.11	0.11/0.09	0.11/0.10	0.09/0.08	0.10/0.09	0.09/0.10

Table 2. The statistics for three levels of registration differences to the corresponding tracings in terms of Rotation (R), Translation (T) and Scale (S). We show $mean/standard\ deviation$ pairs on the two groups for each metric in three analysis levels. Note that we only evaluated the valid drawings / strokes that are registered correctly. The valid counts are also presented here. We have '-' for the pixel-level rotation and scale, since these values cannot be evaluated at the pixel level.

Level	Skill	Count	R (°)	T (pixel)	S
Sketch-level	N	1680	5.06 / 5.52	121.03 / 89.76	0.89 / 0.20
	P	1778	3.57 / 4.58	86.68 / 70.11	0.95 / 0.14
Stroke-level	N	~59.9K	12.83 / 14.54	222.58 / 198.23	1.02 / 0.39
	P	~89.4K	10.60 / 11.98	184.53 / 172.22	1.01 / 0.32
Pixel-level	N	~11.2M	-	8.67 / 14.14	-
	P	~12.6M	-	5.17 / 8.94	-

scaling factor falling into $[0.9, 1.1]$ (75% drawings), compared to the novice drawings (48% and 21%, respectively). The difference of scaling factor indicates that the novice users had diverse scaling settings (a more gentle distribution) for drawing, compared to the professionals who tended to have their drawings with similar scales to the prompt targets.

From a stroke-level perspective, in the professional group, there are 64% strokes with $R_L \leq 10°$, 35% strokes with the translation distance error with $T_L \leq 100$, and 45% strokes with the scaling error in $S_L \in [0.9, 1.1]$. In the novice group, the numbers of strokes falling into the correct ranges of the three metrics are all less than those in the professional group, i.e., 57% for R_L, 26% for T_L, and 23% for S_L. It reflects that the professional users were more capable of placing strokes with the proper position, orientation, and especially scale than the novice users. Thus, it is challenging for novice users to make part-by-part ratios stay consistent with each other.

For the pixel level, it is obvious that the professional users drew more pixels (63%) near the correct position (≤4 pixels) and fewer outlier pixels (17%) (≥8 pixels), compared to the novice users who drew 49% and 29% pixels, respectively. We also computed the mean and variance for R, T, and S between the two groups on the three levels, as shown in Table 2. We conducted the statistical tests using a linear mixed model (LMM) with the R-style formula, e.g., "$T \sim$ Group + (1 | PromptID) + (1 | UserID)" for T, showing again the significant differences between the two groups (with the significance level p<0.001) on R, T, and S over the three levels. It proves again that the novice users could not perform well to organize strokes from global to local or draw a correct path of each stroke to depict a target shape as the professional users did.

3.3 Could Scaffold Lines Help Improve Sketches?

Table 4 of the main paper shows that scaffold lines could significantly reduce global errors on scaling E_{GS} and pixel-level errors E_P, for both the novice and professional users. On the stroke level, scaffolding made a significant difference on locating stroke positions (see E_{LT}) for the novice drawings, but not for the professional drawings. In addition, it shows that the scaffold lines did not successfully guide the orientation of neither the whole sketch nor each stroke for the two groups. We further tested Spearman correlation coefficients r with significant level p between the scaffold-line numbers and the pixel-level error E_P for the two groups using scaffold lines. We found that there was a significantly negative correlation for the novice drawings (r=-0.18, p<0.001) with scaffold lines, meaning that more scaffold lines resulted in less E_P error, while no significant correlation existed in the professional drawings (r=-0.05, p=0.11).

3.4 Do People Sketch 3D Shape Differently over Time?

We also computed Spearman correlation coefficient between different drawing features and time for each drawing, including stroke length, the speed of drawing each stroke, stroke duration, and the distance between the two end points of each stroke. Figure 4 shows most of the drawings have no significant correlation between the features and time, both for the novice and professional groups.

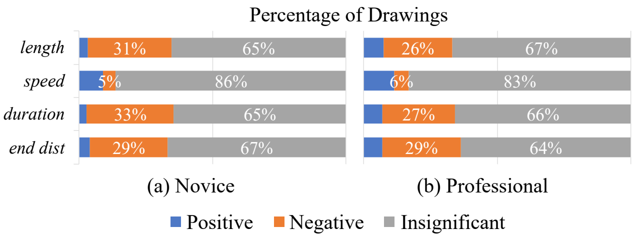


Fig. 4. Histograms of the correlations between different drawing features and time, including stroke length, the speed of drawing each stroke, stroke duration, and the distance between the two end points of each stroke.

We computed the energy costs of the guidelines for each drawing, following Fu et al. [2011]:

$$E_{sim} = \frac{1}{n}\sum_{i}^{n} c_{ind}(l_i)\theta(i), \tag{1}$$

$$E_{pro} = \frac{1}{n-1}\sum_{i}^{n-1} c_{pro}(l_i, l_{i+1}), \tag{2}$$

$$E_{col} = \frac{1}{n-1}\sum_{i}^{n-1} c_{col}(l_i, l_{i+1}), \tag{3}$$

$$E_{anc} = \frac{|\{(l_c, l_s) \in T | c > s\}|}{|T|}, \tag{4}$$

where l_i indicates the i-th drawn stroke and n is the stroke count of a drawing, while T is a set of the detected T-junctions (l_c,l_s) between the crossbar stroke l_c and the stem stroke l_s. For the anchoring cost E_{anc}, we computed the proportion of the T-junctions that were drawn without following the anchoring guideline (i.e., the crossbar l_c is assumed to be drawn earlier than the stem l_s, that is, c<s). Please refer to [Fu et al. 2011] for the detailed definitions for $c_{ind}(\cdot)$, $\theta(\cdot)$, $c_{pro}(\cdot)$, $c_{col}(\cdot)$, etc.

3.5 Stroke-level Comparison

We defined stroke precision as the case that over half of pixels of one stroke are aligned with targets. To observe the difference of the sketching performance on strokes between the novice users and professional users, we computed stroke precision on the stroke-level registered results by setting tracings (we also collected for our dataset) as targets. Figure 5 shows users' performance on each category. Professional users outperformed at all the categories, of which *Human Face*, *Vehicle*, *Animal* have larger disparity between the two groups than *Primitive* and *Lamp* show. It reflects the professional users have better command of sketching complicated 3D objects than novice users do.

4 APPLICATION

4.1 Freehand-style Sketch Synthesis

We trained the three disturbers using three MLPs (100 epochs), each of which consists of two hidden layers (each one following a ReLU activation layer) respectively with 100 and 50 neurons, using the Adam solver with the fixed learning rate 1e-3. We trained each MLP separately on the novice and professional drawings and the corresponding registered results. The dataset for the novice style consists ~70k paired strokes while that for the professional style has ~102k pairs we set a training/testing ratio of the dataset to 20:1 in our experiments.

Stroke Extrinsic Disturbing. The input S_e of the extrinsic disturber is a fitted Bézier curve with six control points $t \in \mathbb{R}^{12}$ for each stroke of tracings with the noise level n_1, i.e., $S_e \in \mathbb{R}^{13}$, while the output O_e is the transformation parameters for rotation, translation and scaling, i.e., $O_e \in \mathbb{R}^4$. When training, the noise level n_1 is computed using the stroke-level errors (mentioned in Section 5.3 of the main paper) by E_{LR}+E_{LT}+E_{LS}. Note that each error is normalized before training. When testing, the curve t is applied by the predicted transformation O_e and then is fed into the next intrinsic disturber.

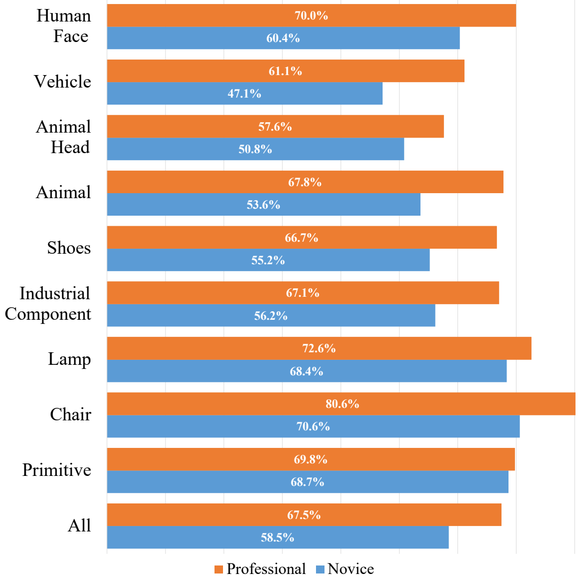


Fig. 5. Stroke precision of the stroke-level registered results for each category.

Stroke Intrinsic Disturbing. The input $S_i \in \mathbb{R}^{13}$ and output $O_i \in \mathbb{R}^{12}$ of the intrinsic disturber are both a Bézier curve, while the input is also incorporated with the noise level n_2. We compute n_2 by measuring the Euclidean distance between S_i and O_i for the training dataset. When testing, the predicted curve O_i is set as an input of the point disturber.

Point Disturbing. For the point disturber, the input $S_P \in \mathbb{R}^{12}$ is a Bézier curve and the output $S_P \in \mathbb{R}^2$ is a normal distribution $N(\mu, \sigma)$. When training, we measure the offsets between all the points of the Bézier curve and the original stroke (we use the sketch-level registered results for training) and compute the mean μ and standard deviation σ of the offsets. When testing, the input curve S_P is added with the predicted normal noise before the next step (layout refinement).

We provide more examples for each step (Figure 6) and with different noise levels (Figure 7) to show the plausibility and diversity of our freehand-style synthesis method.

4.2 Preliminary Evaluation of Sketch-based 3D Reconstruction

We chose four popular image-based shape reconstruction methods that can take single-view sketches as input. Each of the four chosen methods uses one of the commonly used shape representations: meshes, voxels, implicit functions, and point clouds. First, we chose Pixel2Mesh [Wang et al. 2018] as a representative mesh-based approach, which deforms a template ellipsoid to a target shape. Pixel2Mesh is often adopted as a baseline due to its generality and flexibility on shape reconstruction. The well-known 3D-R2N2 [Choy et al. 2016] network was chosen as a representative volumetric approach for shape reconstruction. We chose Occupancy Network

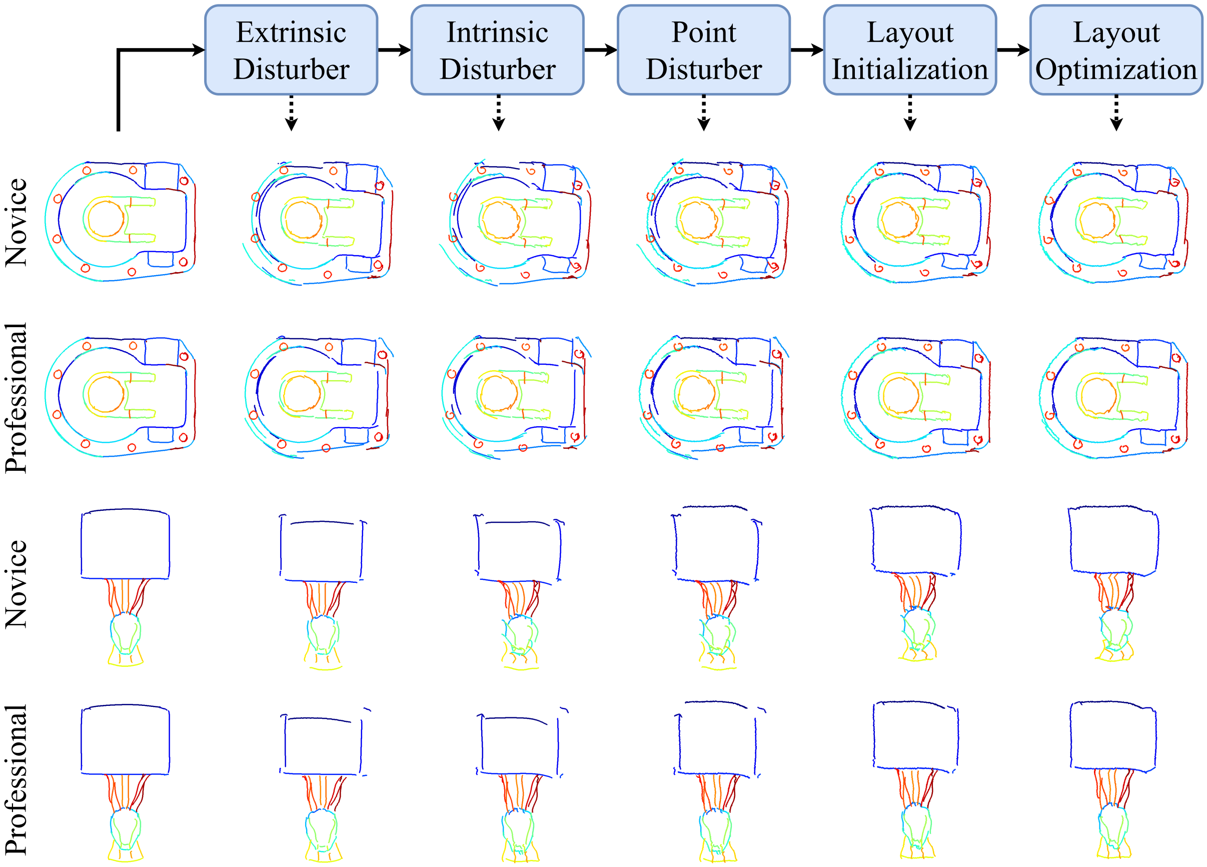


Fig. 6. The pipeline of our freehand-style sketch synthesis method. Each two rows of images show novice-style and professional-style generation, both of which are fed with the same noise levels (n_1=n_2). Each stroke of the sketches is color-coded to highlight their changes.

[Mescheder et al. 2019] and PSGN [Fan et al. 2017] as the representative methods based on implicit functions and point clouds, respectively. For a fair comparison, we trained all the four models with synthetic sketches from [Zhong et al. 2020]. Different from the other three methods, which require only a single sketch as input, Pixel2Mesh requires an input sketch and its view angle as input for the subsequent reconstruction. We thus provided the corresponding view angle information paired with the synthetic sketch in training and applied the view angles associated with the reference image as the corresponding view information of freehand sketches in test.

We provide more visual comparisons of the reconstructed shapes given different input freehand sketches in Figure 8. Since Pixel2Mesh reconstructs shapes by deforming an ellipsoid, the resulting models often present excessive adhesion in the leg areas. Due to the limited sampling resolution of voxels, R2N2 presents results with disconnected parts (e.g., legs and back area) in some cases. In addition, due to the adoption of a volumetric representation, R2N2 produces shapes with surfaces expanding the stacks of individual cubic voxels, which also explains its relatively low NC values in the quantitative evaluation. Since the sketch inputs contain only binary shape information, such scarce information would add more difficulties for reconstruction methods to identify the within/outside shape relation, resulting in messy generation for complex parts (e.g., hollow backs and swivel bases).

REFERENCES

Christopher B Choy, Danfei Xu, JunYoung Gwak, Kevin Chen, and Silvio Savarese. 2016. 3d-r2n2: A unified approach for single and multi-view 3d object reconstruction. In *European conference on computer vision*. Springer, 628–644.

Haoqiang Fan, Hao Su, and Leonidas J Guibas. 2017. A point set generation network for 3d object reconstruction from a single image. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 605–613.

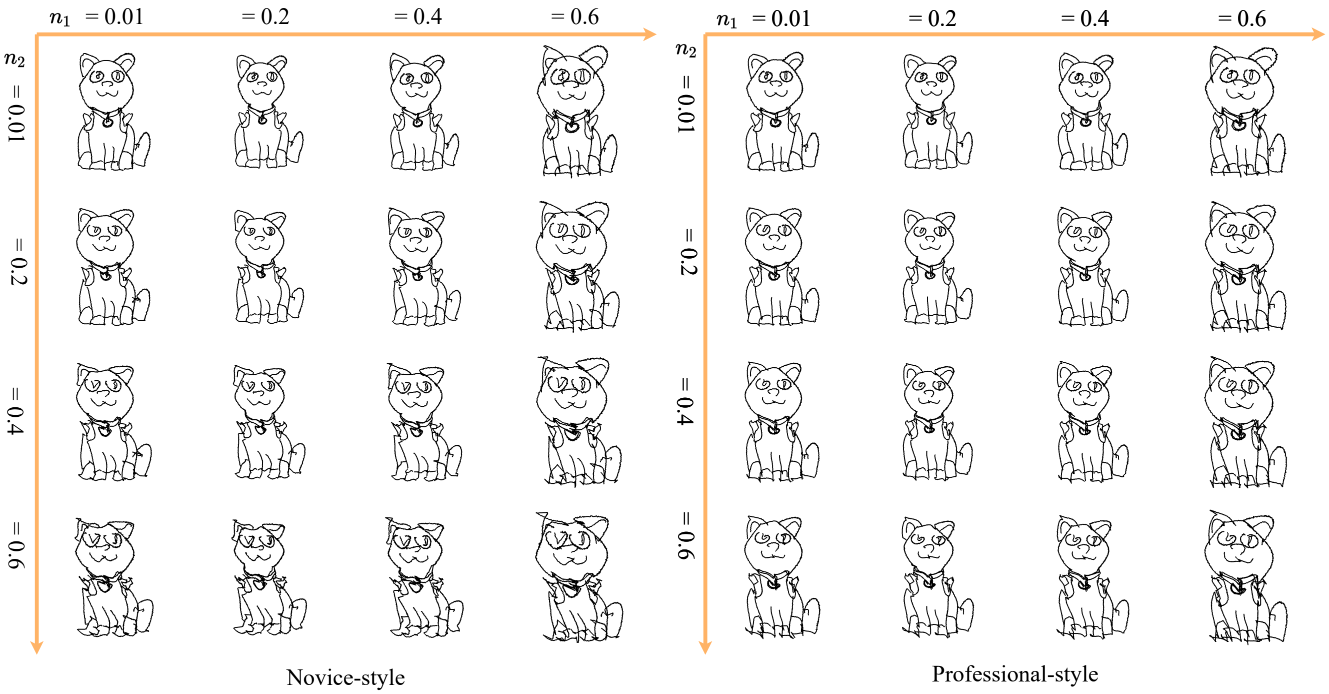
Hongbo Fu, Shizhe Zhou, Ligang Liu, and Niloy J Mitra. 2011. Animated construction of line drawings. In *Proceedings of the 2011 SIGGRAPH Asia Conference*. 1–10.

Lars Mescheder, Michael Oechsle, Michael Niemeyer, Sebastian Nowozin, and Andreas Geiger. 2019. Occupancy networks: Learning 3d reconstruction in function space. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 4460–4470.

Nanyang Wang, Yinda Zhang, Zhuwen Li, Yanwei Fu, Wei Liu, and Yu-Gang Jiang. 2018. Pixel2mesh: Generating 3d mesh models from single rgb images. In *Proceedings of the European conference on computer vision (ECCV)*. 52–67.

Zeyu Wang, Sherry Qiu, Nicole Feng, Holly Rushmeier, Leonard McMillan, and Julie Dorsey. 2021. Tracing Versus Freehand for Evaluating Computer-Generated Drawings. *ACM Trans. Graph.* 40, 4 (Aug. 2021), 12. https://doi.org/10.1145/3450626.3459819

Ziv Yaniv, Bradley C Lowekamp, Hans J Johnson, and Richard Beare. 2018. SimpleITK image-analysis notebooks: a collaborative environment for education and reproducible research. *Journal of digital imaging* 31, 3 (2018), 290–303.



Novice-style

Professional-style

Fig. 7. Examples of our freehand-style synthesis method fed with different noise levels n_1 and n_2.

Yue Zhong, Yulia Gryaditskaya, Honggang Zhang, and Yi-Zhe Song. 2020. Deep sketch-based modeling: Tips and tricks. In *2020 International Conference on 3D Vision (3DV)*. IEEE, 543–552.

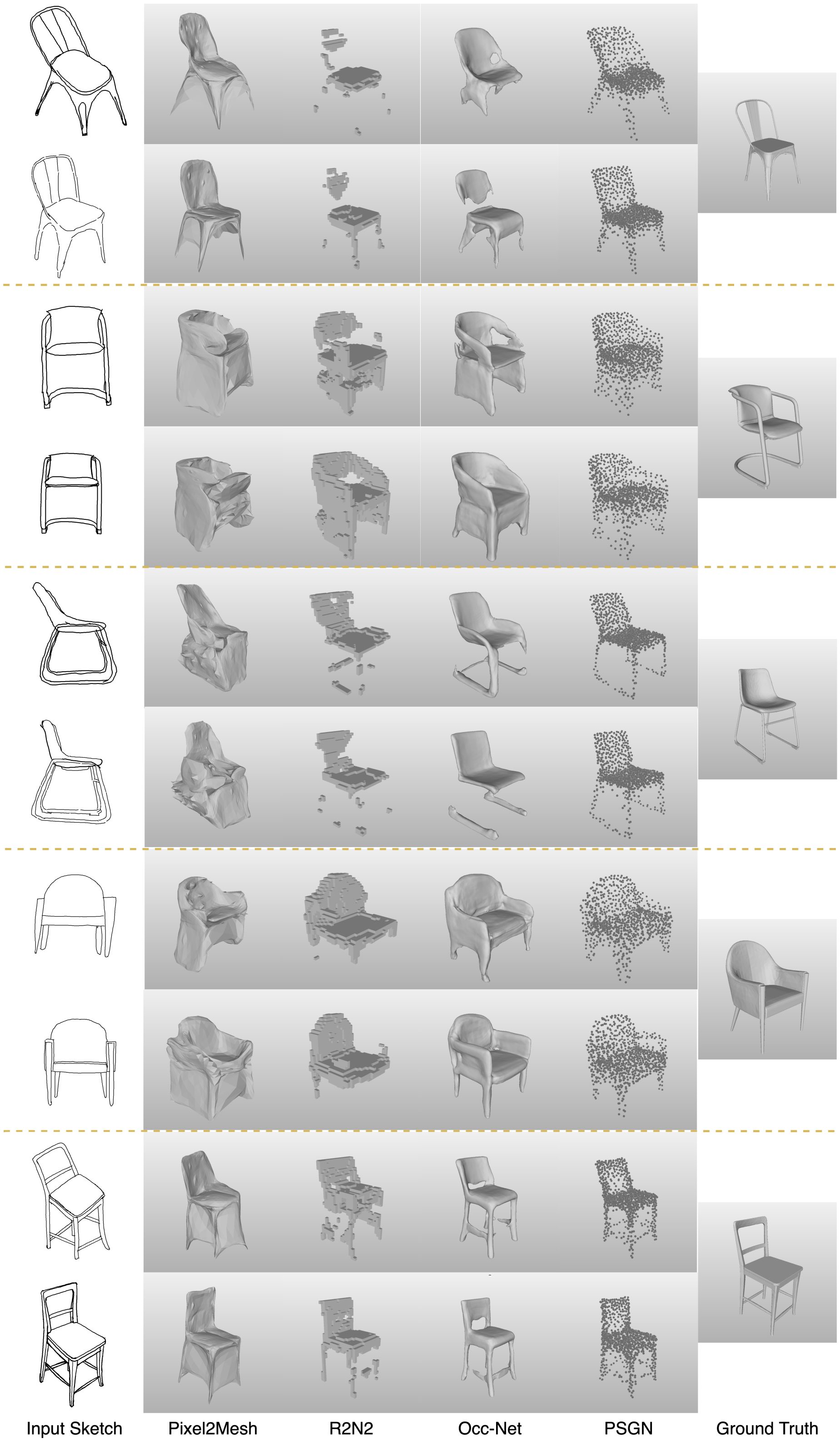


Fig. 8. More results of the 3D reconstruction comparison. In each group, the top presents the results from novice input, the bottom illustrates the corresponding reconstruction results given professional sketches; on the right, we present the ground truth shapes.