Co-Locating Style-Defining Elements on 3D Shapes

Ruizhen Hu[1] Wenchao Li[2] Oliver van Kaick[3] Hui Huang[1,2] Melinos Averkiou[4] Daniel Cohen-Or[5] Hao Zhang[6]
[1]Shenzhen University [2]SIAT [3]Carleton University [4]University of Cyprus [5]Tel Aviv University [6]Simon Fraser University

1 Supplementary material

In this supplementary material, we give more details on our evaluation and show additional results.

Evaluation of algorithm components and settings. We give more details and show experiment plots for the evaluation of the components and settings of our algorithm.

Geodesic radius of elements. We evaluate the effect that changing the main parameter used in the element construction, the geodesic radius τ of the geometric patches, has in the accuracy of style classification. We test values $\tau = [0.03, 0.06, 0.09, 0.12, 0.15]$ which are multiplied by the diagonal of the shape's axis-aligned bounding box. Figure 1 presents this evaluation. We observe that a radius of 0.09 leads to the overall best results.

Shape representations. We test two different representations to encode shapes in our method. We compare our bag-of-words encoding based on element frequency with the encoding used by Arietta et al. [2014]. In their encoding, each entry is the top score returned by applying the SVM classifier for its corresponding candidate element on all shape elements. More precisely, as in the work of Arietta et al., we retain the top 3 scores of the SVM, creating a vector of dimension $3m$, where $m = |\mathcal{C}|$. We observe in Figure 5 that, although the two approaches are comparable for three sets, our representation leads to the best overall results.

Dataset size. We study how the size of the input dataset affects the classification accuracy, since the learning of the element similarity changes based on the input data provided. We divide each dataset into training and test sets, where we run our method on the train set and evaluate the classification accuracy on the test set. We take 10% of the shapes in a dataset as the test set, and then run the experiment with training sets that contain different fractions of the total number of shapes. We average the results for 10 random test tests and for all styles. In Figure 2, we show the result of this test for each individual set, along with a summary for all datasets. In the graph, we see how the classification accuracy is already over 85% for training sets containing 50% of the shapes in the dataset. This corresponds to approximately 300 shapes for the larger set of furniture and 40 shapes for the smaller set of drinking vessels.

Feature weights. Regarding the learning aspects of the method, we also investigate how the different features used to represent the elements are weighted by the similarity measures learned for each element. Figure 3 shows the average weights collected for the measures of all the elements in all the datasets. We note that all the weights are non-zero, implying that each feature is relevant for at least one of the datasets. Moreover, we note that the descriptors with the highest weights range from the more complex point feature histogram, to the simpler height of the element along the upright direction of the shape.

Preliminary experiments on content selection. To show the flexibility of our method regarding the input, we also apply our method to shape-content tables (SC-tables). In an SC-table, shapes are grouped into rows and columns, so that the rows represent different content categories (e.g., chairs and tables), while the columns reflect style groupings. By restricting our method to select elements related to a single cell of the SC-table, we discover elements that define a specific style and content. In Figure 4, we present the selection of elements for content-specific style extraction, applied on our set of furniture grouped by style and content. The examples we show are part of the set of elements that define the style of Ming chairs, but do not define shapes of the same style, nor same content.

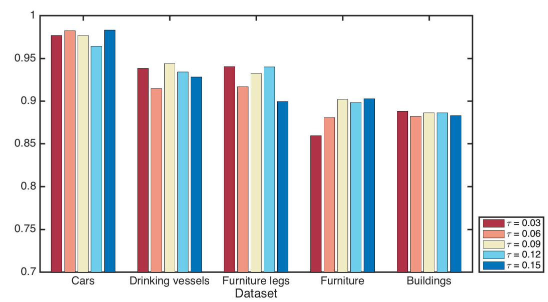


Figure 1: *Evaluation of different geodesic radii used in the construction of the elements. By averaging the accuracies across all sets, we conclude that the best overall performance occurs when the radius is 0.09.*

Given that our method is able to select defining elements relevant to specific types of content, we also study the selection of content-dependent elements independently of their style. In Figure 6, we observe that our method is general enough to also select elements that are content-specific and not linked to specific styles, being able to capture the generic characteristics of a type of content.

In Figure 7, we generalize our application for deriving stylistic scores of shapes to the case of content. We show the scores of all the shapes in our dataset that indicate how well the shapes fit a specific style and content. The scores for content are computed similarly as for styles by counting the number of content-defining elements. We observe that we also obtain a meaningful grouping of shapes for content, along with the grouping for style.

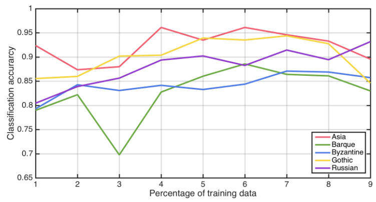
We also investigate how the feature weights used by the similarity measures learned for each element change in the context of content analysis. In Figure 8, we show the feature weights for an example element when it is used to define the style or content of a shape. We see how the weights change considerably to adapt to each setting.

User studies. Figure 9 shows the interface that human subjects used to construct the ground-truth for evaluating our method. Moreover, Figure 10 shows the interface used to evaluate the similarity measures learned per element.

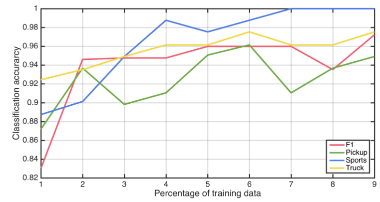
Datasets. Figures 11–18 show all the shapes that compose our dataset, where the shapes are organized by category and styles.

References

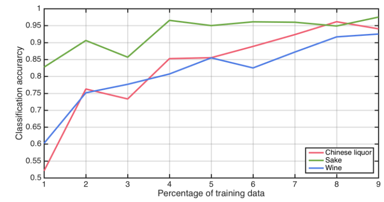
ARIETTA, S., EFROS, A. A., RAMAMOORTHI, R., AND AGRAWALA, M. 2014. City forensics: Using visual elements to predict non-visual city attributes. *IEEE TVCG 20*, 12, 2624–2633.



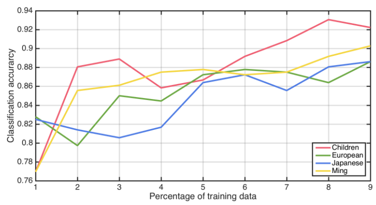
(a) Buildings



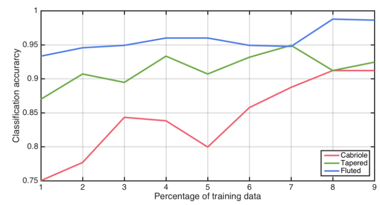
(b) Cars



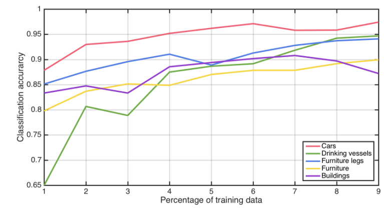
(c) Drinking vessels



(d) Furniture



(e) Furniture legs



(f) Average for all sets

Figure 2: *Effect of the size of the input on the style classification accuracy for different datasets.*

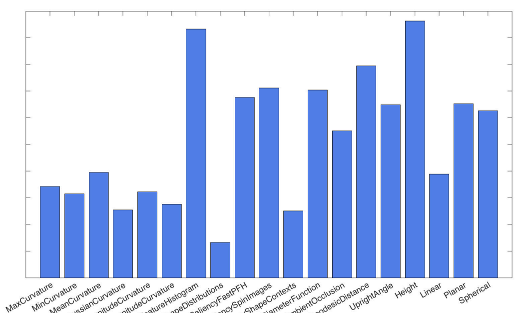


Figure 3: *Weight of each feature in the similarity measure of the elements. Note how all the features have a non-zero weight.*

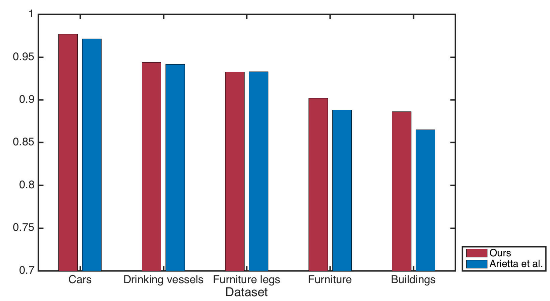


Figure 5: *Evaluation of different shape representations used in style classification. Note how our frequency-based representation leads to overall better results.*

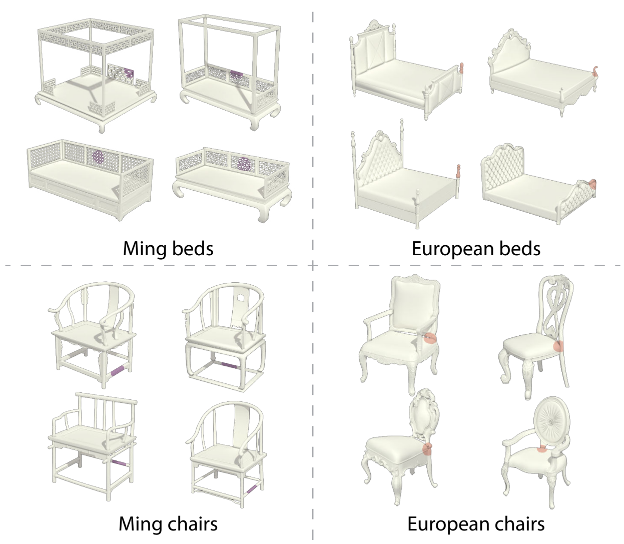


Figure 4: *Examples of content-dependent elements selected by our method. The defining elements selected for beds in Ming style do not define different content of the same style, e.g., Ming chairs, nor define content of different styles, e.g., European beds, nor define shapes with different content and style, e.g., European chairs.*

Figure 6: *Gallery of content-defining elements selected by our method, where we show examples of positive elements selected for each type of content.*

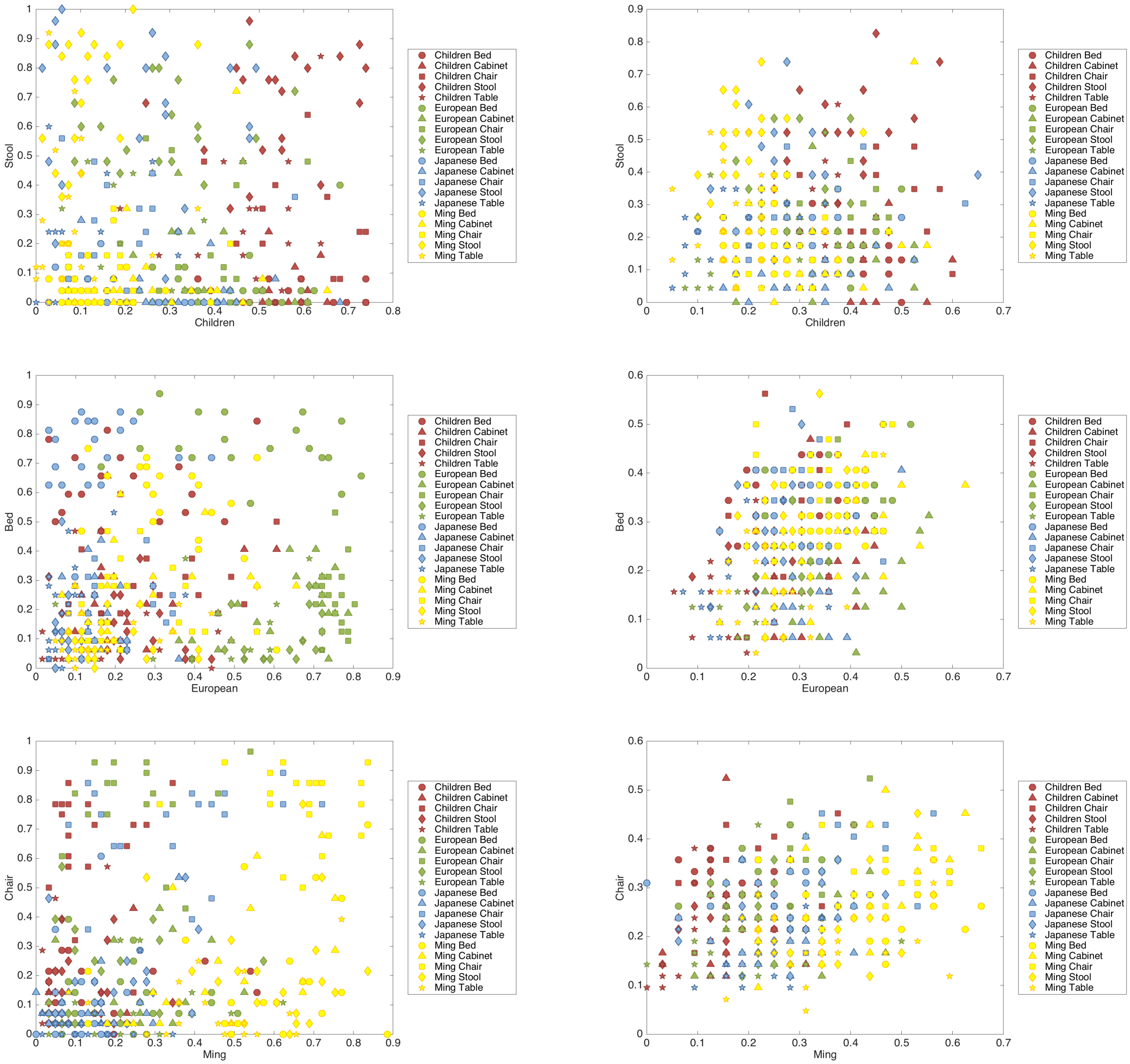
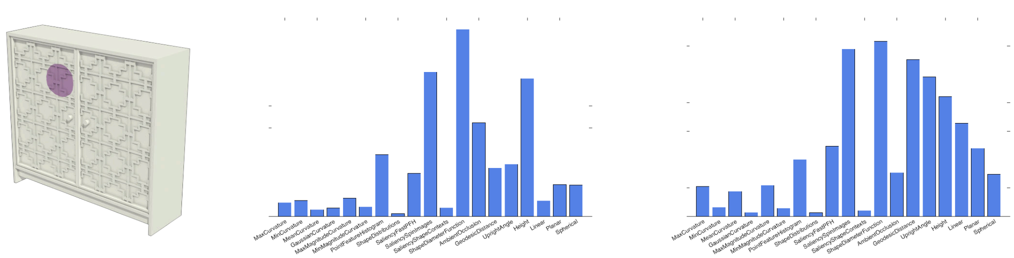


Figure 7: *Graphs reflecting how strongly each individual shape fits within a style and content. We compare our method (left column) with L1-reg (right column). Each marker is a shape, and the larger the values along an axis, the stronger the shape's relationship to the corresponding style or type of content. We observe how the grouping given by our method leads to a clearer separation between the different groups.*



(a) Defining element (b) Style feature weight (c) Content feature weight

Figure 8: *Weight of each feature in the similarity measure learned for the element in (a). We show the feature weights when the element is used to define the style of the shape, in (b), and the content of the shape, in (c).*

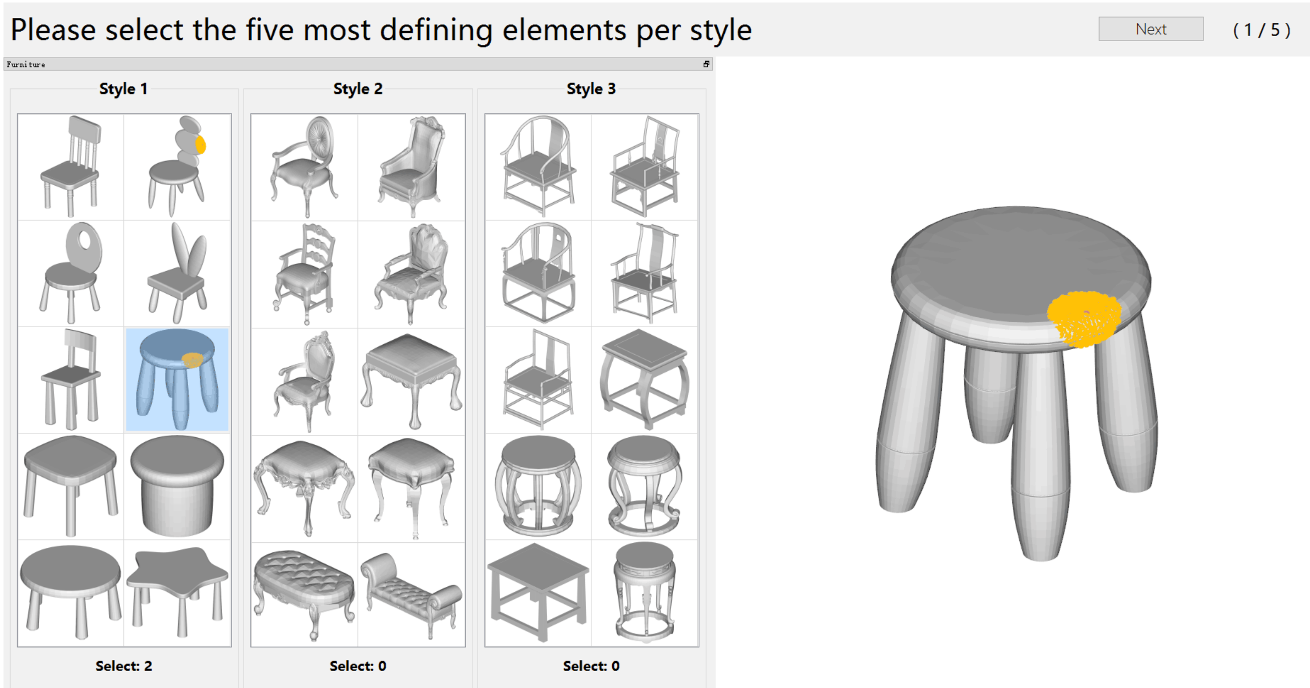


Figure 9: *Interface used for building the ground-truth. The full dataset is shown on the left. When the user clicks on a model, it is shown on the right, where the user can select the defining elements. By pressing 'Next', the next dataset is shown.*

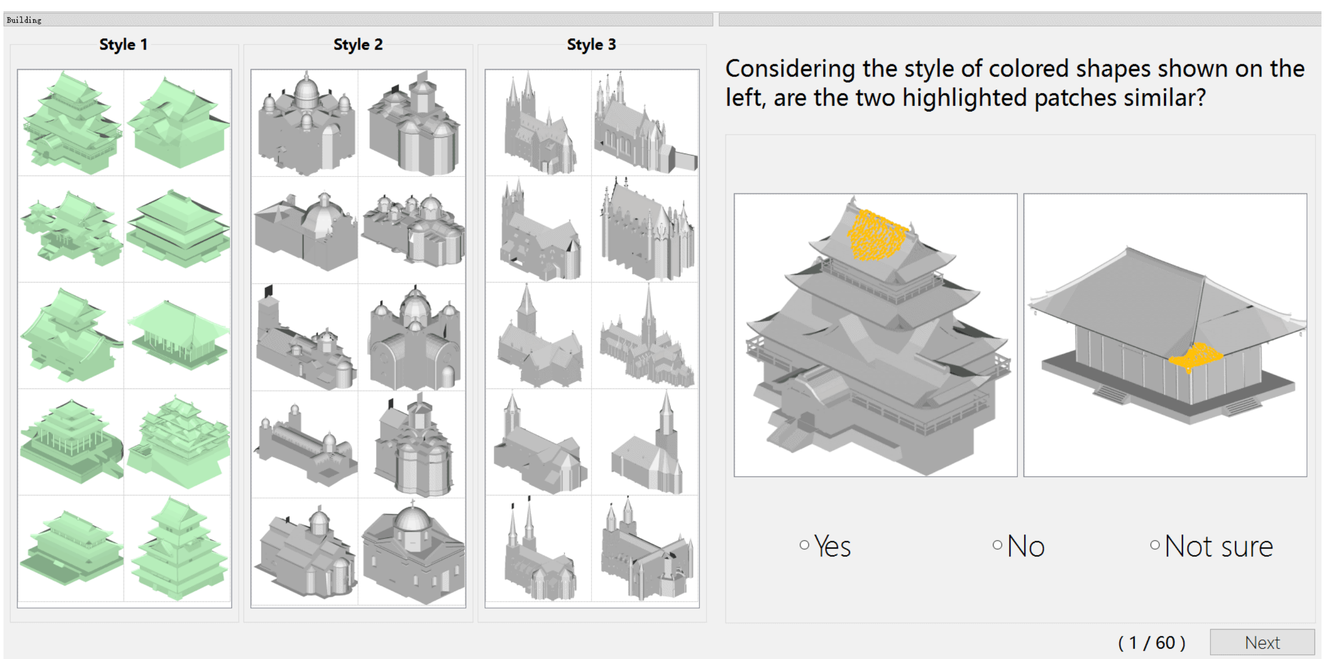


Figure 10: *Interface used for evaluating the stylistic similarity measures learned per element. The user has to assess whether the two patches shown on the right are similar or dissimilar, according to the context style highlighted in green on the left. Other styles are shown for contrast. The user can also select whether he or she is unsure about the similarity. By pressing 'Next', the next pair of patches is shown.*

Children

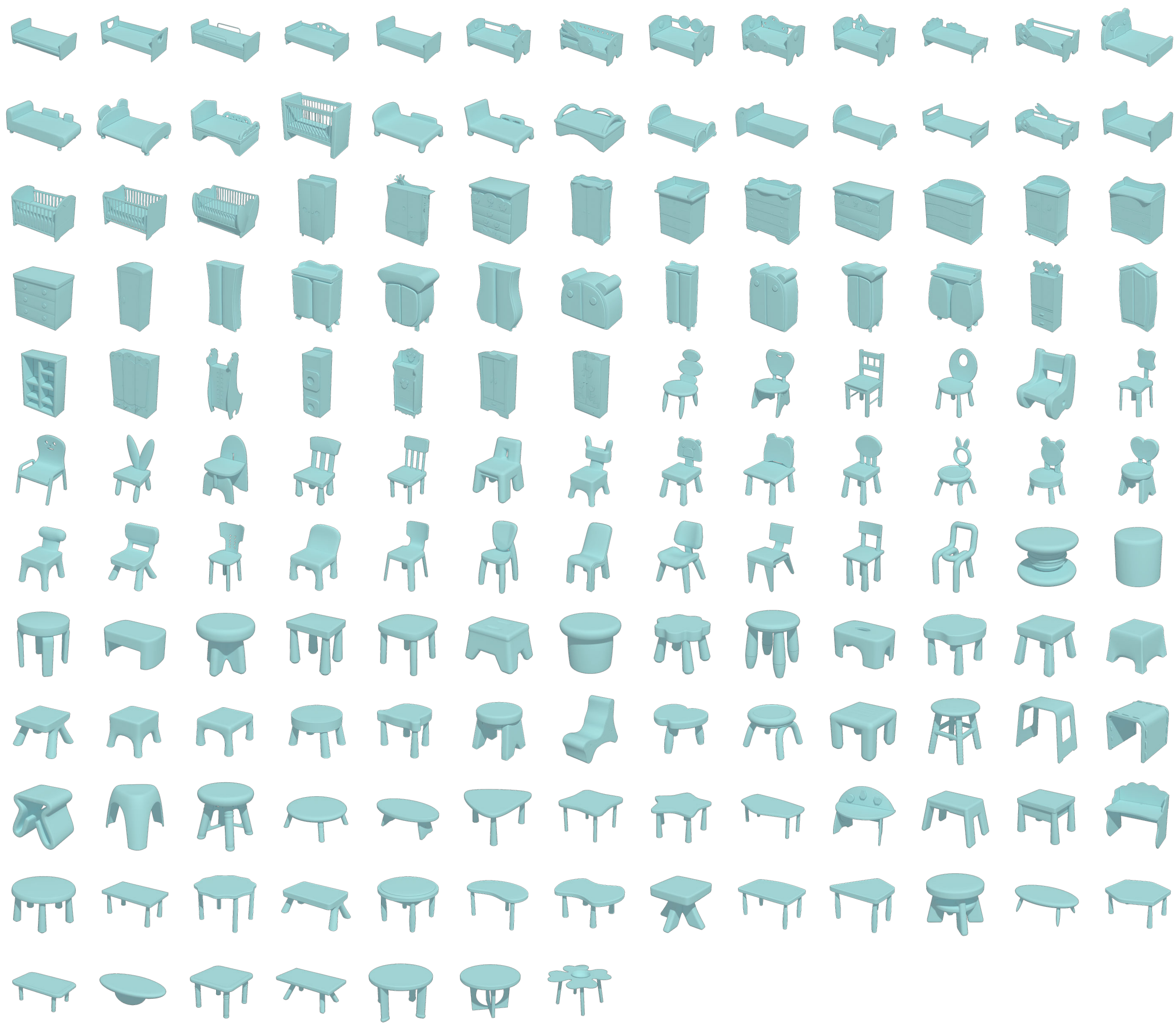


Figure 11: *All the shapes in the set of furniture for the Children style.*

European

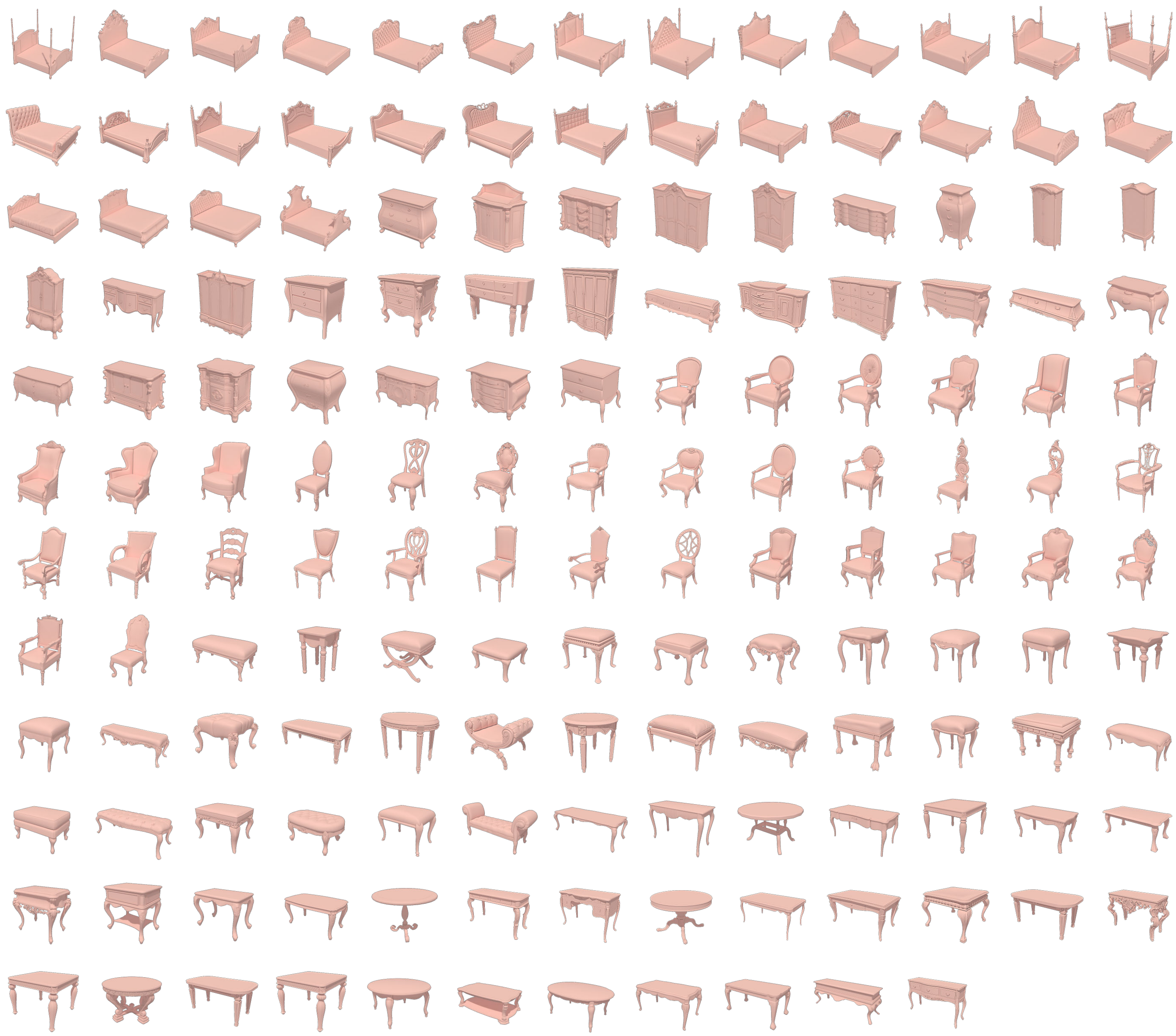


Figure 12: *All the shapes in the set of furniture for the European style.*

Japanese

Figure 13: *All the shapes in the set of furniture for the Japanese style.*

Ming

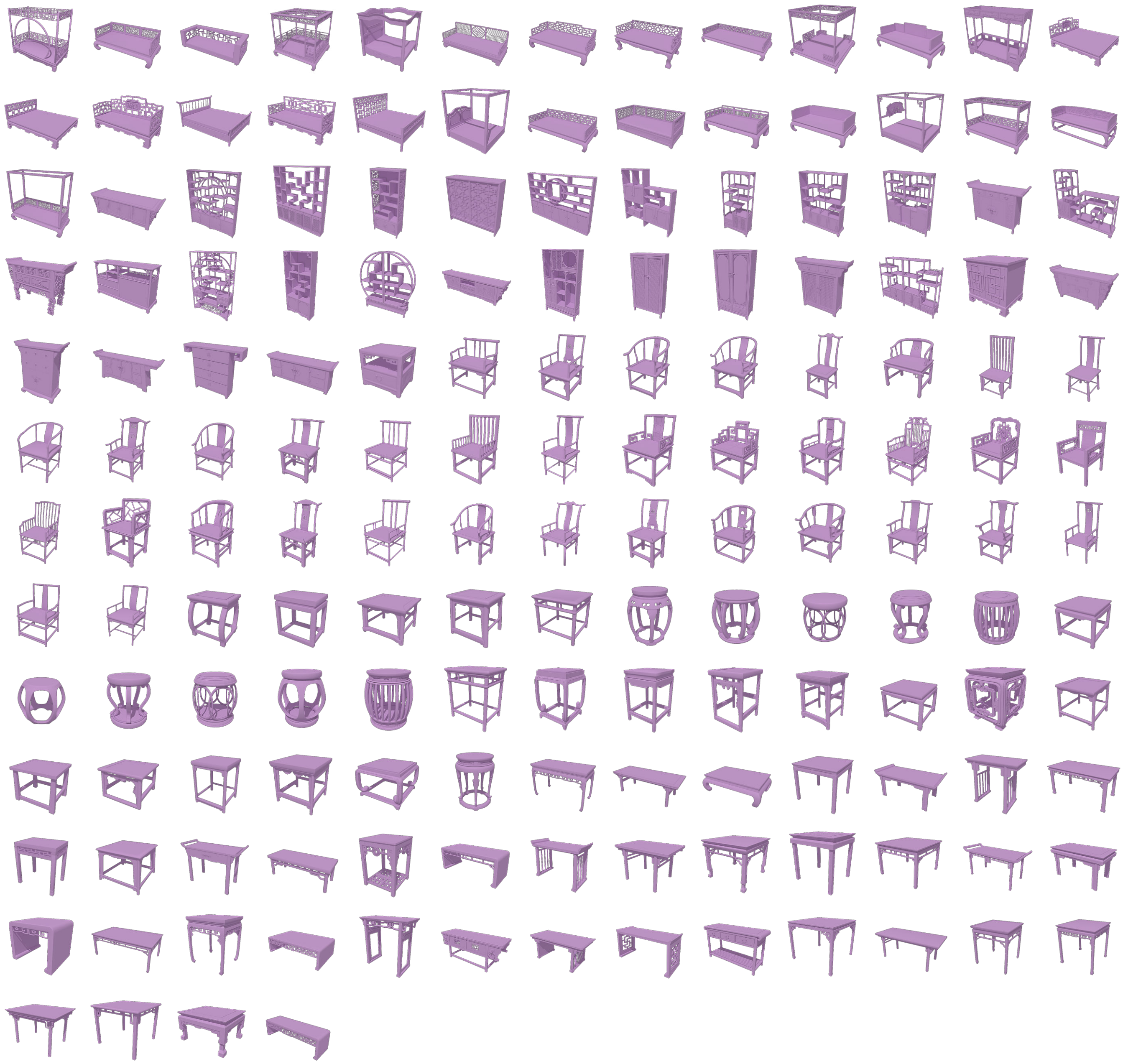


Figure 14: *All the shapes in the set of furniture for the Ming style.*

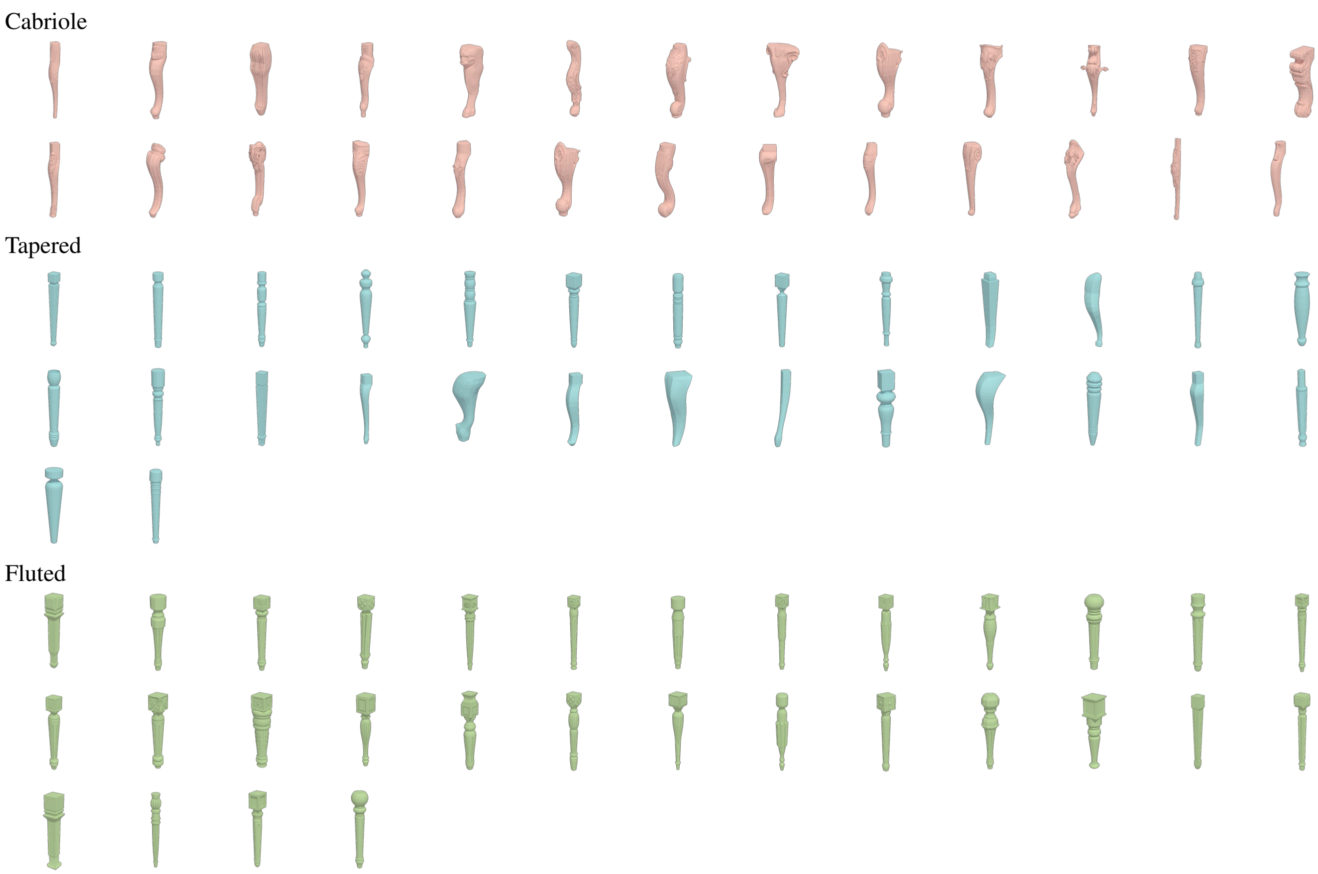


Figure 15: *All the shapes in the set of furniture legs, grouped by their style.*

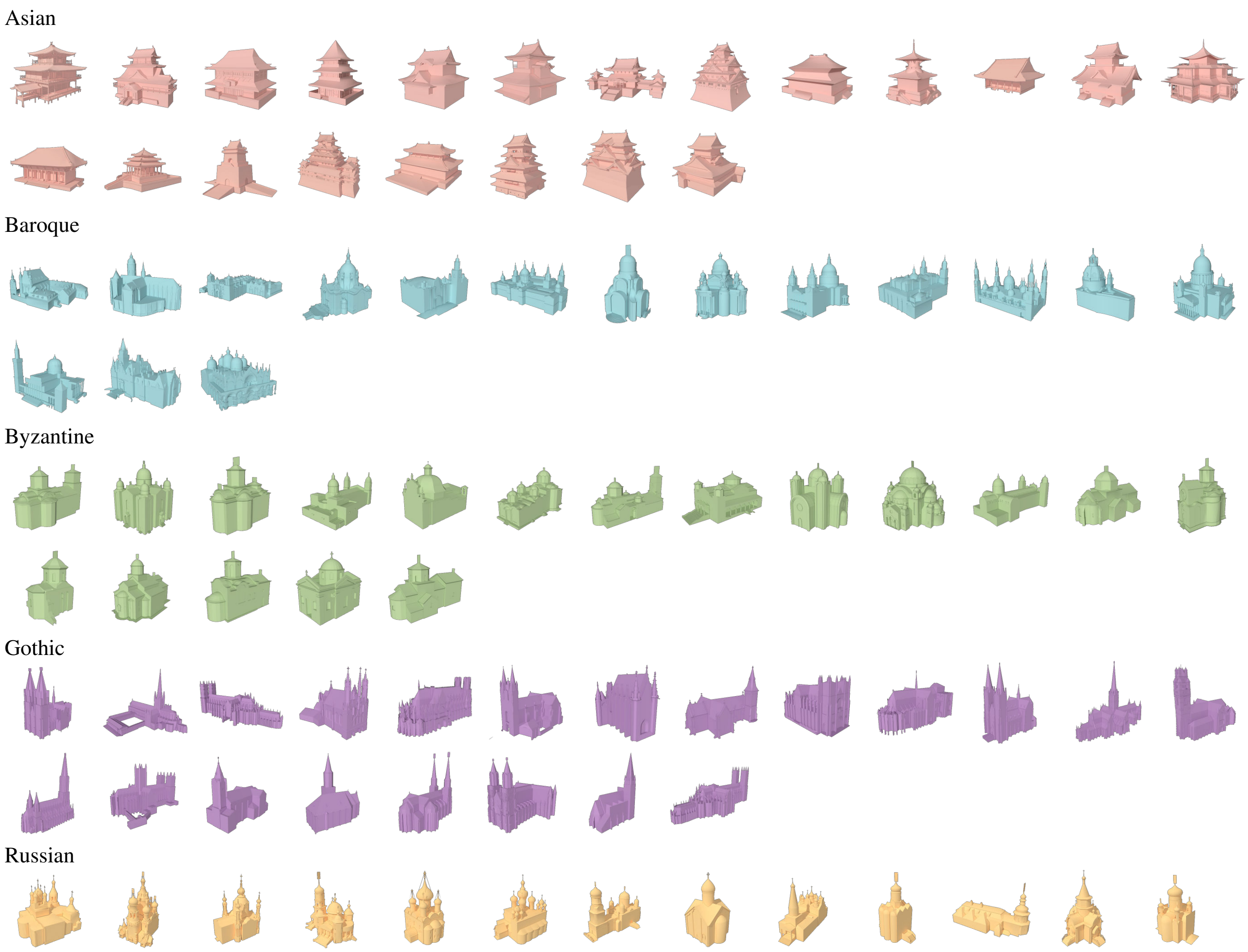


Figure 16: *All the shapes in the set of buildings, grouped by their style.*

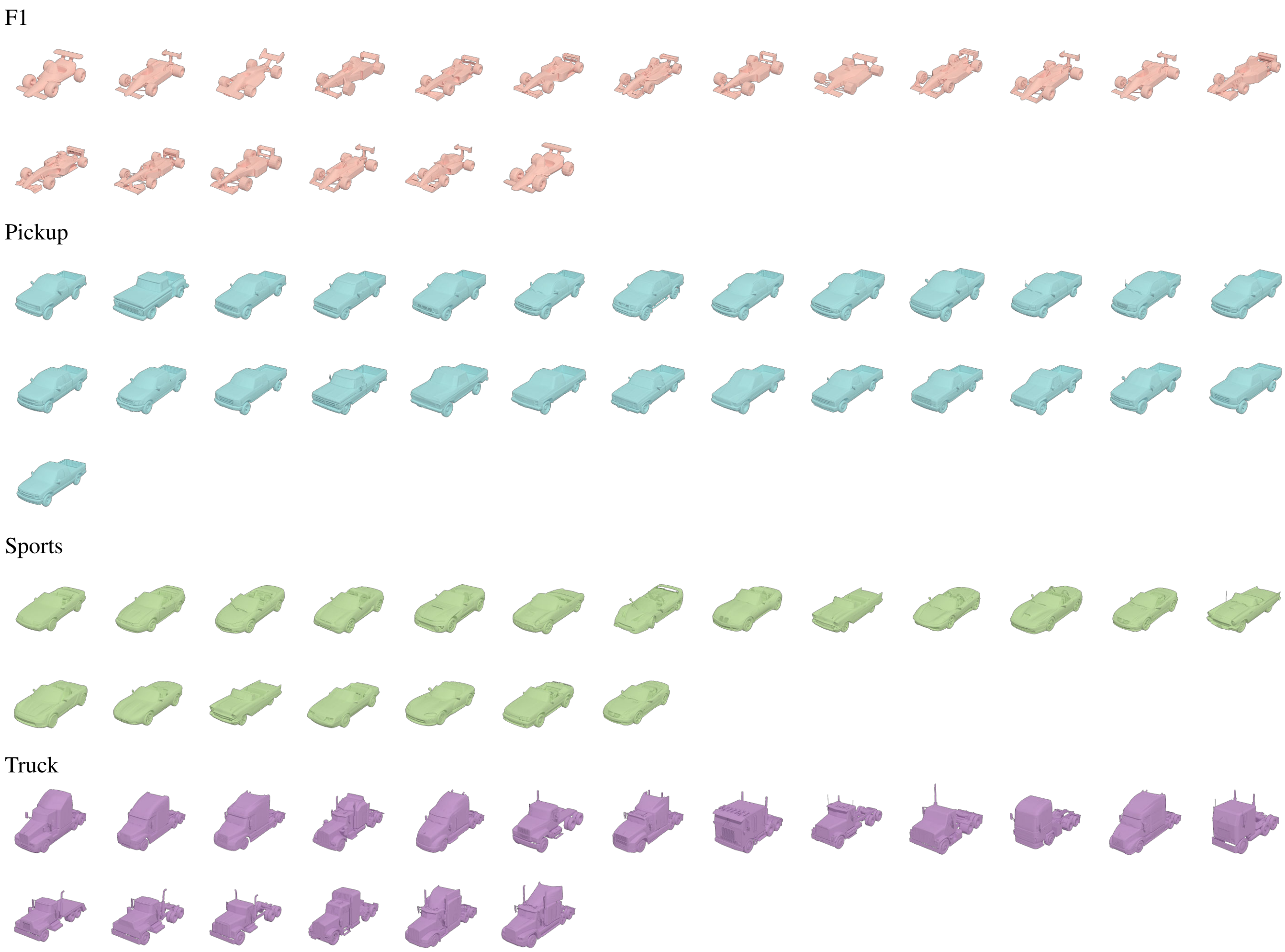


Figure 17: *All the shapes in the set of cars, grouped by their style.*

Figure 18: *All the shapes in the set of drinking vessels, grouped by their style.*